Journal of Molecular Evolution
https://doi.org/10.1007/s00239-018-9837-9

ORIGINAL ARTICLE

Application of Chloroplast Phylogenomics to Resolve Species Relationships Within the Plant Genus *Amaranthus*

Erika Viljoen[1,2] · Damaris A. Odeny[3] · Martin P. A. Coetzee[4] · Dave K. Berger[2] · David J. G. Rees[1,5]

Received: 29 July 2017 / Accepted: 16 March 2018

Abstract

Amaranthus species are an emerging and promising nutritious traditional vegetable food source. Morphological plasticity and poorly resolved dendrograms have led to the need for well resolved species phylogenies. We hypothesized that whole chloroplast phylogenomics would result in more reliable differentiation between closely related amaranth species. The aims of the study were therefore: to construct a fully assembled, annotated chloroplast genome sequence of *Amaranthus tricolor*; to characterize *Amaranthus* accessions phylogenetically by comparing barcoding genes (*mat*K, *rbc*L, ITS) with whole chloroplast sequencing; and to use whole chloroplast phylogenomics to resolve deeper phylogenetic relationships. We generated a complete *A. tricolor* chloroplast sequence of 150,027 bp. The three barcoding genes revealed poor inter- and intra-species resolution with low bootstrap support. Whole chloroplast phylogenomics of 59 *Amaranthus* accessions increased the number of parsimoniously informative sites from 92 to 481 compared to the barcoding genes, allowing improved separation of amaranth species. Our results support previous findings that two geographically independent domestication events of *Amaranthus hybridus* likely gave rise to several species within the Hybridus complex, namely *Amaranthus dubius, Amaranthus quitensis, Amaranthus caudatus, Amaranthus cruentus* and *Amaranthus hypochondriacus*. Poor resolution of species within the Hybridus complex supports the recent and ongoing domestication within the complex, and highlights the limitation of chloroplast data for resolving recent evolution. The weedy *Amaranthus retroflexus* and *Amaranthus powellii* was found to share a common ancestor with the Hybridus complex. Leafy amaranth, *Amaranthus tricolor, Amaranthus blitum, Amaranthus viridis* and *Amaranthus graecizans* formed a stable sister lineage to the aforementioned species across the phylogenetic trees. This study demonstrates the power of next-generation sequencing data and reference-based assemblies to resolve phylogenies, and also facilitated the identification of unknown *Amaranthus* accessions from a local genebank. The informative phylogeny of the *Amaranthus* genus will aid in selecting accessions for breeding advanced genotypes to satisfy global food demand.

Keywords Phylogenomics · Chloroplast · *Amaranthus* · Barcode

Introduction

The plant genus *Amaranthus* comprises approximately 60 species, mostly annuals of naturally open habitats and are distributed throughout the world's tropical and temperate

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00239-018-9837-9) contains supplementary material, which is available to authorized users.

✉ Dave K. Berger
dave.berger@up.ac.za

[1] Biotechnology Platform, Agricultural Research Council, Onderstepoort, Pretoria 0110, South Africa

[2] Department of Plant and Soil Sciences, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Hatfield 0083, South Africa

[3] International Crops Research Institute for the Semi-Arid Tropics, Nairobi, Kenya

[4] Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Hatfield 0083, South Africa

[5] Department of Life and Consumer Sciences, College of Agricultural and Environmental Sciences, University of South Africa, Florida 1710, South Africa

Published online: 19 March 2018

regions (Stetter and Schmid 2017). Although the majority of amaranth species are cosmopolitan weeds, the genus also includes cultivated species used as leafy vegetables (*Amaranthus* subgenus *Albersia*) (Mosyakin and Robertson 1996; van Rensburg et al. 2007), a source of grain (*Amaranthus* subgenus *Amaranthus*) (Mosyakin and Robertson 1996; Maughan et al. 2009) as well as ornamental plants (Sauer 1967). The edible leafy amaranth species are rich in essential micronutrients including β-carotene (Raju et al. 2007; Sangeetha and Baskaran 2010), minerals (Mnkeni et al. 2007) and sulphur-containing amino acids (Mlakar et al. 2010). Grain amaranths, also referred to as "pseudo-cereals" due to their non-grass nature (Das 2011), have a nutritional advantage over more conventional cereal grains due to their lack of gluten and increased starch and lysine content compared to other cereal crops (Rastogi and Shukla 2013). The grain oil is high in squalene (7–11%) and plays an important role in the medicinal and cosmetic fields (Mlakar et al. 2010). Amaranth plants in general are reported to be useful for rehabilitating wastelands (Alamgir et al. 2011) and can be used for biofuel production (Timofte et al. 2009; Akond et al. 2013). Amaranths are therefore an ideal choice for sustainable food production, crop diversification and nutritional security in many nations across the world (Ebert 2014).

Amaranths are cultivated throughout the world, including Central America, Mexico, Eastern Africa and Asia (India, Nepal, China, Indonesia) (Kavita and Gandhi 2017). In Africa and India, *Amaranthus tricolor* is a popular leafy vegetable especially due to the high palatability of the leaves (Srivastava 2017). It has been found that *A. tricolor* has a high tolerance to drought conditions, and especially to high salinity in the soil compared to other amaranth species (Lubbe and Rodda 2016). In addition, leaf extracts of *A. tricolor* show antibacterial activities against plant pathogenic bacteria such as *Xanthomonas, Erwinia* and *Pseudomonas*. In animal studies (as well as a few clinical trials in humans), the anti-cancer, anti-viral, antioxidant and hepatoprotective properties of leaf and root extracts from *A. tricolor* has been demonstrated (Kavita and Gandhi 2017).

Identification of members within *Amaranthus* based solely on morphological characteristics is difficult, because species are mainly separated based on small (sometimes microscopic) morphological characters. The morphology of *Amaranthus* plants are severely influenced by environmental conditions, species types, production techniques and genotypes, which in turn lead to significant differences in phenotypes between and within species groups (Srivastava 2017). Furthermore, several intermediate morphological forms exist in the wild and sporadic cross-species hybridization occurs, which can result in a mixture of morphological characters that differ from the type species (Gudu and Gupta 1988; Brenner et al. 2000; Achigan-Dako et al. 2014). Despite the difficulty associated with morphological identification, *Amaranthus* species are classified into three main sub-genera (*Amaranthus* subspecies *Acnida, Amaranthus* subgenus *Albersia* and *Amaranthus* subgenus *Amaranthus*). However, the exact evolutionary history and domestication events that led to amaranth species today are still under speculation (Stetter and Schmid 2017). To overcome problems associated with morphological classification, recent studies have moved towards molecular analysis based on nuclear and chloroplast sequence information for effective identification and classification of amaranths (Costea et al. 2006; Das 2011).

Earlier studies revealed that sequencing data generated from chloroplast genes could provide sufficient information to delineate plant species and phylogenetic reconstruction of taxa (Bell et al. 2016; Patil et al. 2016; Bezeng et al. 2017; Braukmann et al. 2017). Each chloroplast is uniparentally inherited and non-recombinant; simplifying phylogenetic reconstructions (Zhang et al. 2012). The International Barcoding of Life Initiative (iBOL) has emphasized the taxonomic importance of one or more targeted gene sequences to confidently resolve closely related species within species complexes (Kress and Erickson 2007). Although successful sequence barcodes have been identified for animals (the mitochondrial *cox1* gene—*COI*), bacteria (nuclear 16s rRNA) and fungi (nuclear ITS), no such single universal identifier has yet been developed for plants (Lahaye et al. 2008; Bezeng et al. 2017). It has been proposed that a multi-locus barcode would increase phylogenetic resolution between closely related plant species. Chloroplast gene regions such as *atp*F-H, *mat*K, *psb*K-I, *rbc*L, *rpo*C1, *rpo*B, *trn*H-*psb*A, *trn*L-F and the nuclear intergenic spacer region (ITS) have been evaluated extensively across different species groups (Hollingsworth et al. 2011). Despite receiving considerable criticism due to complications in primer design, PCR amplifications, poor performance in the resolution of closely related species and the complicated occurrence of inversions and insertions, the most valuable gene regions for barcoding are the chloroplast *mat*K, *rbc*L, *trn*H-*psb*A, the nuclear ITS and more recently, the *ycf*1 gene (Dong et al. 2015). However, the cost and time involved with screening large populations with different gene sets makes barcoding largely impractical.

The increasing affordability of Next Generation Sequencing (NGS) technologies have made efficient, rapid and affordable high-quality sequencing of entire genomes or plastomes possible. The high-copy nature, structural simplicity, highly conserved gene content and relatively small size make chloroplast sequences an ideal target for high throughput sequencing (Stull et al. 2013). As a result, a tedious and expensive targeted chloroplast isolation is not necessary and low-coverage sequencing is sufficient to access many phylogenetically informative characters within the chloroplast genome sequence (Chaney et al. 2016). The

multiplex identifier tools employed by NGS technologies further enhance output, as multiple chloroplasts can be sequenced at an adequate depth in a single experiment. As bioinformatic tools become more user-friendly, only moderate computing power and data analysis knowledge will be needed to assemble, annotate and compare full-length chloroplast genomes (Li et al. 2014).

Using the whole chloroplast sequence as a "super-barcode" has gained popularity in recent years (Parks et al. 2009; Yang et al. 2013; Li et al. 2014), as it provides considerably more sequence-based variation leading to greatly increased resolution at lower plant taxonomic levels. This method can also circumvent issues pertaining to low PCR efficiency, missing database information of less popular plant species and limited variation supplied by few gene regions. In addition, an in-depth study of a species subset can lead to the identification of genetic barcodes specific to the genus under scrutiny (Li et al. 2014; Dong et al. 2014). Several studies have employed whole chloroplast sequences with the objective to obtain highly resolved phylogenies. For example, phylogenies were obtained for species in the genera *Oryza* (Hackett et al. 2008; Parks et al. 2009; Nock et al. 2011; Barrett et al. 2013; Ma et al. 2014), *Bamboo* (Poaceae: Bambusoideae) (Zhang et al. 2011), *Lilium* (Kim and Kim 2013), *Camellia* (Huang et al. 2014) and *Acacia* (Williams et al. 2016).

To date, eight full-length chloroplast sequences have been assembled for taxa within *Amaranthus*, including five *A. hypochondriacus* accessions, *A. cruentus, A. caudatus* and *A. hybridus* (Chaney et al. 2016). Sequencing reads for *A. hypochondriacus* (cultivar Plainsman) were produced through long read PacBio technology, which resulted in a high-quality full-length assembly of the chloroplast. Subsequent sequencing and assemblies of related species revealed several INDELs, polymorphic microsatellite markers and informative SNP markers, which can be used in downstream phylogenetic and genetic diversity studies of members of the genus (Chaney et al. 2016).

Due to the mainly maternal inheritance of the chloroplast (and consequent lack of hybridization evidence in the sequence data), several studies are also investigating the use of nuclear SNP based markers to determine plant phylogenies. For *Amaranthus*, a phylogeny for a large set of diverse species was determined using genome wide SNP markers identified through NGS genotyping (Stetter and Schmid 2017). Based on the results from this study, novel conclusions could be reached about the incomplete domestication syndrome of especially grain *A. caudatus*, and the data also shed light on the complex relationship between domesticated and wild species within this genus (Stetter et al. 2017; Stetter and Schmid 2017).

In the current study, we generated and annotated the whole chloroplast sequence of *A. tricolor*, an economically important, but neglected leafy vegetable in South Africa. Phylogenetic analysis of the member species was initially undertaken using the most commonly employed barcoding genes (*mat*K, *rbc*L and ITS). Subsequently, the *A. tricolor* whole chloroplast sequence was used as a reference to assemble representative chloroplast sequences of additional 58 diverse *Amaranthus* accessions from South Africa and elsewhere. Whole chloroplast phylogenomics revealed a highly resolved phylogeny within the genus.

Methods

Growth and Maintenance of Germplasm Accessions

Forty-five accessions representing 13 different *Amaranthus* species were provided by Dr. David Brenner, North Central Regional Plant Introduction Station (NCRPIS), Ames, Iowa, United States (Table 1). Accessions were selected from the Germplasm Resources Information Network (GRIN) computer database of the USDA-ARS National plant germplasm system (hereafter referred to as the known, previously identified germplasm set—GRIN). An additional 14 *Amaranthus* accessions, which had previously been collected from different countries across the world, were obtained from the Agricultural Research Council (ARC)—Vegetable and Ornamental Plant Institute (VOPI), Pretoria, South Africa (hereafter referred to as the South African Germplasm set—SAG) (Table 1). Seeds of the 59 accessions were germinated using potting soil in a glasshouse with natural light intensity during day (25–35 °C) and night (20–25 °C) at ARC-VOPI. Plants were watered every day for the first 3 weeks and three times a week thereafter.

Genomic DNA Isolation

Approximately 5 g of young amaranth leaves were collected from seedlings of each accession and stored at − 80 °C until use. Genomic DNA was isolated using a DNeasy® Plant Mini DNA Isolation kit (Qiagen, Valencia CA, USA) following the protocol provided by the manufacturer. DNA concentrations were determined using the Qubit® 2.0 Fluorometer Broad Range dsDNA quantification assay (Invitrogen, Life Technologies, CA, USA). DNA integrity was evaluated by 1% agarose gel electrophoresis stained with 0.5 µg/ml ethidium bromide.

Illumina Library Preparation and Sequencing

For whole chloroplast sequencing, 5 µg of *Amaranthus tricolor* (SAG29, GenBank accession nr: KX094399) genomic DNA was used to prepare 100 bp paired-end

Table 1 *Amaranthus* germplasm sets

GRIN germplasm				
Code	GRIN accession	Species	Country of Collection	Species based on whole chloroplast phylogeny[b]
GRIN1	Ames 24670	*A. blitum*	Portugal	*A. blitum*
GRIN2	PI 652433	*A. blitum*	Brazil	*A. blitum*
GRIN3	Ames 13890	*A. caudatus*	China	*A. caudatus*
GRIN4	Ames 15179	*A. caudatus*	Argentina	*A. caudatus*
GRIN5	PI 669934	*A. caudatus*	India	*A. caudatus*
GRIN6	PI 481458	*A. caudatus*	Germany	*A. caudatus*
GRIN7	PI 553073	*A. caudatus*	USA	*A. caudatus*
GRIN8	Ames 2056	*A. cruentus*	Nigeria	*A. cruentus*
GRIN9	Ames 5313	*A. cruentus*	USA	*A. cruentus*
GRIN10	PI 566897	*A. cruentus*	India	*A. cruentus*
GRIN11	Ames 1967	*A. dubius*	India	*A. dubius*
GRIN12	PI 482047	*A. dubius*	Zimbabwe	*A. dubius*
GRIN13	PI 612850	*A. dubius*	USA	*A. dubius*
GRIN14	PI 641049	*A. dubius*	Nigeria	*A. dubius*
GRIN15	PI 608661	*A. graecizans*	India	*A. graecizans*
GRIN16	PI 658732	*A. graecizans*	Portugal	*A. graecizans*
GRIN17	Ames 1990	*A. hybridus*	India	*A. hybridus*
GRIN18	Ames 25409	*A. hybridus*	South Africa	*A. hybridus*
GRIN19	PI 604602	*A. hybridus*	Mexico	*A. hybridus*
GRIN20	PI 641051	*A. hybridus*	Nigeria	*A. hybridus*
GRIN21	PI 652416	*A. hybridus*	Brazil	*A. hybridus*
GRIN22	PI 667174	*A. hypochondriacus*	Zimbabwe	*A. hypochondriacus*
GRIN23	Ames 5689	*A. hypochondriacus*	Brazil	*A. hypochondriacus*
GRIN24	PI 337611	*A. hypochondriacus*	Uganda	*A. hypochondriacus*
GRIN25	PI 538322	*A. hypochondriacus*	USA	*A. hypochondriacus*
GRIN26	PI 619247	*A. hypochondriacus*	Mexico	*A. hypochondriacus*
GRIN27	PI 636187	*A. hypochondriacus*	India	*A. hypochondriacus*
GRIN28	Ames 15306	*A. powellii*	Mexico	*A. hypochondriacus*[b]
GRIN29	PI 572260	*A. powellii*	France	*A. powellii*
GRIN30	PI 604671	*A. powellii*	USA	*A. powellii*
GRIN31	AMES 15315	*A. quitensis*	Argentina	*A. quitensis*
GRIN32	PI 652421	*A. quitensis*	Brazil	*A. quitensis*
GRIN33	Ames 21767	*A. retroflexus*	China	*A. retroflexus*
GRIN34	Ames 25428	*A. retroflexus*	Pakistan	*A. retroflexus*
GRIN35	PI 572263	*A. retroflexus*	USA	*A. retroflexus*
GRIN36	Ames 2150	*A. spinosus*	Kenya	*A. spinosus*
GRIN37	PI 482058	*A. spinosus*	Zimbabwe	*A. spinosus*
GRIN38	PI 632248	*A. spinosus*	USA	*A. spinosus*
GRIN39	Ames 5110	*A. tricolor*	West Africa	*A. tricolor*
GRIN40	Ames 5134	*A. tricolor*	USA	*A. tricolor*
GRIN41	Ames 5139	*A. tricolor*	USA	*A. tricolor*
GRIN42	Ames 23271	*A viridis*	India	*A viridis*
GRIN43	Ames 25412	*A. viridis*	South Africa	*A. viridis*
GRIN44	PI 641048	*A. viridis*	Nigeria	*A. viridis*
GRIN45	PI 654388	*A. viridis*	USA	*A. viridis*

Table 1 (continued)

SAG—South African germplasm				
Code	Accession	Species (preliminary)[a]	Country of collection	Species based on whole chloroplast phylogeny[b]
SAG 1	50612	*A. bouchonii*	Unknown	*A. powellii*
SAG 3	50613	*A. caudatus*	Unknown	*A. quitensis*/*A. hybridus*
SAG 4	PI 477913 (Grain)	*A. cruentus*	Mexico	*A. cruentus*/*A. caudatus*
SAG 7	Arusha leaf	sp.	Unknown	*A. cruentus*
SAG 9	Tanzania	sp.	Tanzania	*A. cruentus*/*A. caudatus*
SAG 10	Botswana	sp.	Botswana	*A. praetermissus*[c]
SAG 11	W6927N	sp.	Unknown	*A. tricolor*
SAG 12	Bosbok	sp.	South Africa	*A. praetermissus*[c]
SAG 14	Local 33	sp.	South Africa	*A. hybridus*
SAG 17	Vukani Thepe	sp.	South Africa	*A. praetermissus*[c]
SAG 29	*A. tricolor*	*A. tricolor*	USA	*A. tricolor*
SAG 30	Arusha Grain	sp.	Unknown	Unknown
SAG 34	AM Fune	sp.	Unknown	*A. dubius*
SAG 36	AC7	*A. tricolor*	Unknown	*A. tricolor*

[a]Identified based on preliminary morphological analysis (data not shown)

[b]Based on whole chloroplast sequencing—Fig. 5

[c]Based on *mat*K sequencing (data not shown)

sequencing libraries with the Nextera™ DNA Sample Prep Kit (Illumina, San Diego, USA) according to the manufacturer's protocol. Size selection was performed by excising an approximate 300 bp size fragment from a 1% agarose gel stained with 0.5 µg/ml ethidium bromide using the MinElute Gel Extraction Kit (Qiagen). The sample was sequenced on a HiScanSQ Illumina sequencer (Illumina) using TruSeq SBS v3. The sequencing was performed at the Biotechnology Platform Sequencing Facility, Agricultural Research Council, South Africa. For the remaining 58 *Amaranthus* accessions, 5 µg genomic DNA of each sample was used to prepare sequencing libraries followed by barcode-indexing (Illumina) and sequencing on one lane of a MiSeq Illumina sequencer (Illumina) at the same facility. Approximately 100–200 Mb data were generated for each accession, and the resulting raw reads were de-multiplexed.

Assembly of the *Amaranthus tricolor* Chloroplast Genome Sequence

The paired-end sequencing data (2×100 bp) were imported to CLC Bio Genomics Workbench v8 (CLCBio, CLC Inc., Aarhus, Denmark). 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 sugarbeet (*Beta vulgaris*, GenBank accession nr: EF534108.1), a close relative within the Amaranthaceae family, with default parameters. The consensus *A. tricolor* chloroplast sequence was retrieved and used as a reference for a second round of mapping of *A. tricolor* reads in order to validate the consensus *A. tricolor* chloroplast sequence. All trimmed and quality-filtered sequence reads (including reads that mapped onto the reference chloroplast sequence and that were used to assemble the *A. tricolor* chloroplast sequence) have been deposited in the SRA archive of NCBI (Accession # PRJNA318736). Non-mapped reads, which are assumed to be of non-plastid origin, were excluded from further analysis. To close gaps and resolve the four junction region sequences between the large single-copy, small single-copy and inverted repeat regions (IR_A and IR_B), 22 primer pairs (Online Resource Table S1) were designed and used to amplify *A. tricolor* genomic DNA before sequencing on an ABI 3130XL sequencer (Applied Biosystems, CA, USA) at the Inqaba Biotechnical Industries Pty. Ltd. (Pretoria, South Africa). In addition, an *A. tricolor* chloroplast sequence was assembled based on mapping reads to the recently published *A. hypochondriacus* chloroplast sequence (GenBank accession nr: KX279888.1) (Chaney et al. 2016); however, this was only used for comparison with the final *A. tricolor* chloroplast sequence determined by mapping to the sugarbeet chloroplast, which was used for all subsequent analyses.

Amaranthus tricolor Chloroplast Genome Annotation

The complete *A. tricolor* chloroplast genome sequence was annotated using the Dual Organellar GenoMe Annotator (DOGMA, http://dogma.ccbb.utexas.edu/). DOGMA predicts protein coding genes, ribosomal RNA and transfer RNA genes, together with start and stop codons, as well as the presence of pseudogenes. A circular diagram for the chloroplast was generated using the web-based chloroplast visualization software GenomeVx (http://oldwolfe.gen.tcd.ie/GenomeVx/). The assembled and annotated *A. tricolor* chloroplast genome sequence was deposited at NCBI (Accession KX094399).

Mapping of the Additional *Amaranthus* Germplasm Sets

Trimmed reads from each species sampled (45 accessions of GRIN and 13 accessions of SAG) were independently mapped to the *A. tricolor* reference (SAG29) sequence using CLC Genomics Workbench v8 with default parameters. Due to the identical nature of the inverted repeat regions of the chloroplast (IR_A and IR_B), the second inverted repeat (IR_B) was removed from the reference chloroplast sequence to simplify mapping and subsequent phylogenetic analysis, as well as to avoid inherent redundancy. By using the default mapping parameters, it was possible to assemble individual consensus chloroplast genome sequences using a small sample read set, while retaining minimal gapped regions (mapping statistics in Online Resource Table S2). The raw reads for each individual mapping was deposited to the SRA archive of NCBI (BioProject: PRJNA318736; SRA: SRS1400803-SRS1400864).

Angiosperm Whole Chloroplast Genome Phylogeny

The entire chloroplast genomes of 26 angiosperm species (Online Resource Table S3) representing the main plant family groups were downloaded from the National Centre for Biotechnology Information (NCBI) nucleotide database. Sequences of the 26 species were aligned together with that of *A. tricolor* using MUltiple Sequence Comparison by Log-Expectation (MUSCLE) algorithm in the Molecular Evolutionary Genetics Analysis (MEGA) v6.06 software (Tamura et al. 2011) with default settings. Constant, variable and parsimoniously informative site analysis was performed for all sequences. jModelTest v2.1.5 (Guindon and Gascuel 2003; Darriba et al. 2012) was used to determine the nucleotide substitution model with the best fit for the dataset, which subsequently led to the incorporation of a General Time Reversible (GTR) model into the phylogenomic analysis. A maximum likelihood analysis was performed using RaxmlGUI v1.31 (Stamatakis 2014) and confidence for nodes determined using bootstrap analysis with 1000 replicates without partitioning the data.

Amaranthus Barcoding and Phylogenetic Analysis

Three barcoding gene regions were investigated during the phylogenetic analysis (chloroplast *mat*K, *rbc*L and nuclear ITS) as previously suggested (Dong et al. 2015). The *mat*K gene was amplified using primers *mat*K-F (5′-CACTATGTATCATTTGATAACCCTC-3′) and *mat*K-R (5′-TATTACAATCAACATTTCAGAATAG-3′) (Burgess et al. 2011). Primers *rbc*L-F (5′ATGTCACCACAAACAGAGACTAAAGC-3′) and *rbc*L-R (5′-GTAAAATCAAGGTCCACCRCG-3′) were used to amplify the *rbcL* gene (Burgess et al. 2011). The ITS regions were amplified with primers ITS-F (5′-TCCTCCGCTTATTGATATGC-3′) and ITS-R (5′-GGAAGTAAAAGTCGTAACAAGG-3′) (Xu and Sun 2001). Amplification of chloroplast and nuclear barcoding genes was performed for 45 GRIN samples (Table 1). DreamTaq PCR Mastermix (2x, DreamTaq DNA Polymerase, 2x DreamTaq Buffer, dNTPs and 4 mM $MgCl_2$) (Thermo Fisher Scientific, Waltham, MA, USA) was used to amplify each sample in a final volume of 25 µl. The PCR was performed on a G-Storm Thermal Cycler (BioRad Laboratories, CA, USA) with the following conditions: 95 °C for 1 min, 30 cycles of 94 °C for 30 s, 55 °C for 1 min, 72 °C for 1 min and a final extension of 72 °C for 10 min. PCR products were visually assessed after electrophoresis on a 1% (w/v) agarose gel stained with 0.5 µg/ml ethidium bromide. Single amplicons obtained for each germplasm accession were sequenced using Sanger sequencing at Inqaba Biotec, Pretoria, South Africa.

Phylogenetic trees were constructed for the *mat*K, *rbc*L and ITS datasets. Sequences obtained for the three barcoding genes of the 45 GRIN *Amaranthus* accessions (Table 1) have been deposited at GenBank as follows: chloroplast *mat*K (KX079543-KX079587); chloroplast *rbc*L (KX079588-KX079632); and nuclear ITS (KX079498-KX079542). Sequence data for *mat*K, *rbc*L and ITS regions for *Beta vulgaris* (used as an outgroup) were obtained from GenBank (accession numbers AY514832.1, DQ067450.1 and AY858597.1, respectively). Each individual barcoding gene was aligned for all the *Amaranthus* species using MUSCLE in MEGA v6.06. In total, there were 702 (ITS), 836 (*mat*K) and 606 (*rbc*L) base pair positions in the final dataset after primer sequences were removed. Model testing (jModelTest) was used to determine the nucleotide substitution model that best fit the sequence alignment for each gene region individually. The three regions were concatenated and phylograms were obtained using a partitioned analysis in RAxML by applying models that were unlinked for each of the genomic regions. Confidence for the nodes on the phylogram was determined using bootstrap analysis with 1000 replicates.

Amaranthus Whole Chloroplast Genome Phylogenetic Analysis

The large single-copy region (LSC), small single-copy region (SSC) and IR_A chloroplast sequence regions of 45 GRIN and 14 SAG *Amaranthus* accessions were aligned using Clustal Omega (Sievers et al. 2011). Sequence alignments were adjusted manually where necessary. The majority of the sites (70%) that contained missing data for two or more accessions were consequently removed using a custom python script. MEGA was used to determine whether the sites were variable, constant or parsimoniously informative. Phylogenetic analyses were performed using the consensus chloroplast genome sequences of the 45 GRIN and 14 SAG *Amaranthus* accessions. Two datasets were produced; (A) 45 GRIN accessions to confirm identities and classifications of known *Amaranthus* accessions; and (B) 45 GRIN combined with 14 unknown SAG accessions for identifications based on associations to the GRIN accessions. Analyses were conducted on both datasets by treating them as unpartitioned and partitioned in separate analyses. The *B. vulgaris* chloroplast sequence was included as the outgroup to root the trees. Phylogenetic trees were constructed based on both Maximum Likelihood and Bayesian inference of phylogenies. jModelTest was used to determine the nucleotide substitution model that best fit the dataset. Consensus trees and support for nodes were viewed using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Phylogenies based on unpartitioned datasets incorporated entire chloroplast sequences, including genic and intergenic regions. RaxmlGUI v1.31was used for maximum likelihood estimation, utilizing a GTR + G model of nucleotide substitution and 10,000 bootstrap replicates. Bayesian analysis was performed using MrBayes v3.2.4 (Huelsenbeck and Ronquist 2001). The Markov Chain Monte Carlo (MCMC) algorithm was performed for 10,000,000 generations with trees sampled every 1000 generations. Tree convergence was assessed by evaluating effective sample size (ESS) values in Tracer v1.6 (available from http://beast.bio.ed.ac.uk/Tracer). The first 20% of trees from the runs were discarded as burn-in and the remaining trees used to generate a consensus tree and calculate the posterior probabilities for each node.

For the partitioned analysis, each of the 119 genic regions on the chloroplast genome sequence was extracted and analysed individually with jModelTest to determine nucleotide substitution models. Phylogenetic analyses were performed using RAxML and MrBayes as described above but with data being partitioned and applying a gene-specific substitution model on each partition in the Bayesian analysis and a GTR + G model on each partition in the maximum likelihood analysis. Genic informative site analysis and model parameters are shown in Online Resource Table S4.

Identification of an Alternative Barcoding Region for the *Amaranthus* Genus

The annotated whole chloroplast alignment data were used to identify genic regions demonstrating a high level of nucleotide variation between the different amaranth accessions (Table 2). Sequences from these individual regions were subsequently concatenated to the original barcoding dataset (*mat*K, *rbc*L and ITS). A phylogenetic analysis was performed to evaluate whether their addition would add significant phylogenetic signal, and increase the robustness of the tree generated from the original barcoding gene dataset. Partitioned phylogeny reconstruction was performed for each newly generated alignment (MEGA) using a maximum likelihood analysis with RAxML and 1000 bootstrap analysis. The TOPological Distance/From Multiple To Single (TOPD/FMTS) (Puigbò et al. 2007) software was used to evaluate the differences and similarities of each new barcoding tree topology, compared to the partitioned tree obtained through whole chloroplast analysis. Parameters were set to calculate and evaluate the Split Distance (SD), using a random analysis with 1000 repetitions.

Results

Assembly and Annotation of *A. tricolor* Chloroplast Genome Sequence

The chloroplast genome of *Beta vulgaris* (sugarbeet) was used as a reference to construct a preliminary full-length *A. tricolor* chloroplast genome sequence. We generated approximately 1.6 Gb paired-end sequencing data of *A. tricolor*, 7% of which mapped to the sugarbeet chloroplast genome, with an average coverage of 2236x. The overall distribution of coverage across the *B. vulgaris* genome can be found in Online Resource Fig. S1. Assembly of these reads produced a consensus *A. tricolor* draft chloroplast genome sequence of 149,200 bp. There were 18 gap regions with an average size of 2436 bp. The gapped regions were resolved through Sanger sequencing (Online Resource Table S1), resulting in a total *A. tricolor* complete chloroplast genome sequence of 150,027 bp (Fig. 1).

The new assembly conformed to the expected Angiosperm chloroplast topology, divided into a large single-copy region (LSC), a small single-copy region (SSC) and two identical inverted repeats (IR_A and IR_B) (Sato et al. 1999). Each inverted repeat region had a size of 24,345 bp, separated by 83,735 bp (LSC) and 17,598 bp (SSC). The GC content was 36.6%, similar to *Arabidopsis thaliana* (36.3%) (Sato et al. 1999), *Beta vulgaris* (37%) and *Spinacia oleraceae* (36.8%) (Schmitz-Linneweber et al. 2001). The IR_A and IR_B regions each had a GC content of 42.7%, while the

Table 2 Attributes of potential alternative chloroplast barcoding regions

Barcoding region	Size (bp)	Chloroplast region	Informative sites (%)	TOPD results	
				Disagreement (%)	Split distance
*acc*D	1659	LSC	1.6	54	0.72
*atp*A	1524	LSC	1.0	63	0.67
*atp*B	1495	LSC	1.3	67	0.74
*atp*E	409	LSC	1.4	52	0.65
*atp*F	503	LSC	1.8	54	0.72
*atp*H-*atp*I IGR	679	LSC	4.8	70	0.69
*ccs*A	972	SSC	2.9	69	0.74
*cem*A	702	LSC	1.13	67	0.65
*ndh*D	1503	SSC	1.9	43	0.58
*ndh*F	2316	IR	4.0	76	0.81
*ndh*G	534	SSC	1.9	58	0.69
*ndh*H	1204	IR	3.9	60	0.74
*ndh*I	514	SSC	1.16	63	0.63
*ndh*K	861	LSC	2.0	65	0.70
*pet*A	1005	LSC	14	60	0.81
*pet*B	648	LSC	1.0	67	0.65
*pet*D	544	LSC	7.0	67	0.72
*pet*G-*trn*W IGR	136	LSC	13.2	74	0.72
*psa*C	246	SSC	1.6	60	0.69
*psb*A	1068	LSC	1.2	54	0.70
*psb*C	1476	LSC	11.9	74	0.88
*psb*D	1092	LSC	15.0	71	0.81
*psb*I	156	LSC	17.0	74	0.74
*psb*I-*trn*S IGR	226	LSC	27.2	59	0.74
*psb*M	103	LSC	7.7	67	0.74
*rpl*16	401	LSC	2.0	65	0.67
*rpl*16-*rps*3 IGR	1588	LSC	10.1	85	0.83
*rpl*22	592	LSC	5.6	52	0.65
*rpo*A	1012	LSC	2.4	65	0.72
*rpo*B	3249	LSC	1.2	57	0.67
*rpo*B-*trn*C IGR	1197	LSC	5.4	65	0.72
*rpo*C2	2584	LSC	2.9	50	0.72
*rps*14	311	LSC	3.5	56	0.62
*rps*16-*trn*Q	1555	LSC	3.2	74	0.74
*rps*18	309	LSC	13.9	78	0.79
*rps*2	708	LSC	1.0	67	0.74
*rps*3	656	LSC	2.13	58	0.58
*rps*8	404	LSC	1.5	69	0.72
*rrn*16	1549	IR	18.0	87	1.00
*rrn*16-*trn*I IGR	302	IR	22.5	78	0.86
*rrn*23	2842	IR	6.5	87	0.97
*rrn*23-*rrn*4.5 IGR	98	IR	10.2	74	0.72
*trn*C	72	LSC	6.9	74	0.72
*trn*G	83	LSC	18.0	70	0.72
*trn*N	73	IR	21.9	70	0.72
*trn*Q	91	LSC	18.6	59	0.76
*trn*R-*atp*A	109	LSC	13.7	65	0.69
*trn*S	89	LSC	11.2	67	0.72
*trn*S-*rps*4	340	LSC	12.35	71	0.72

Table 2 (continued)

Barcoding region	Size (bp)	Chloroplast region	Informative sites (%)	TOPD results	
				Disagreement (%)	Split distance
*trn*T-*psb*D IGR	1326	LSC	5.8	65	0.72
*ycf*1	1407	IR	0.6	65	0.72

IGR intergenic region, *LSC* long single-copy region, *IR* inverted repeat region

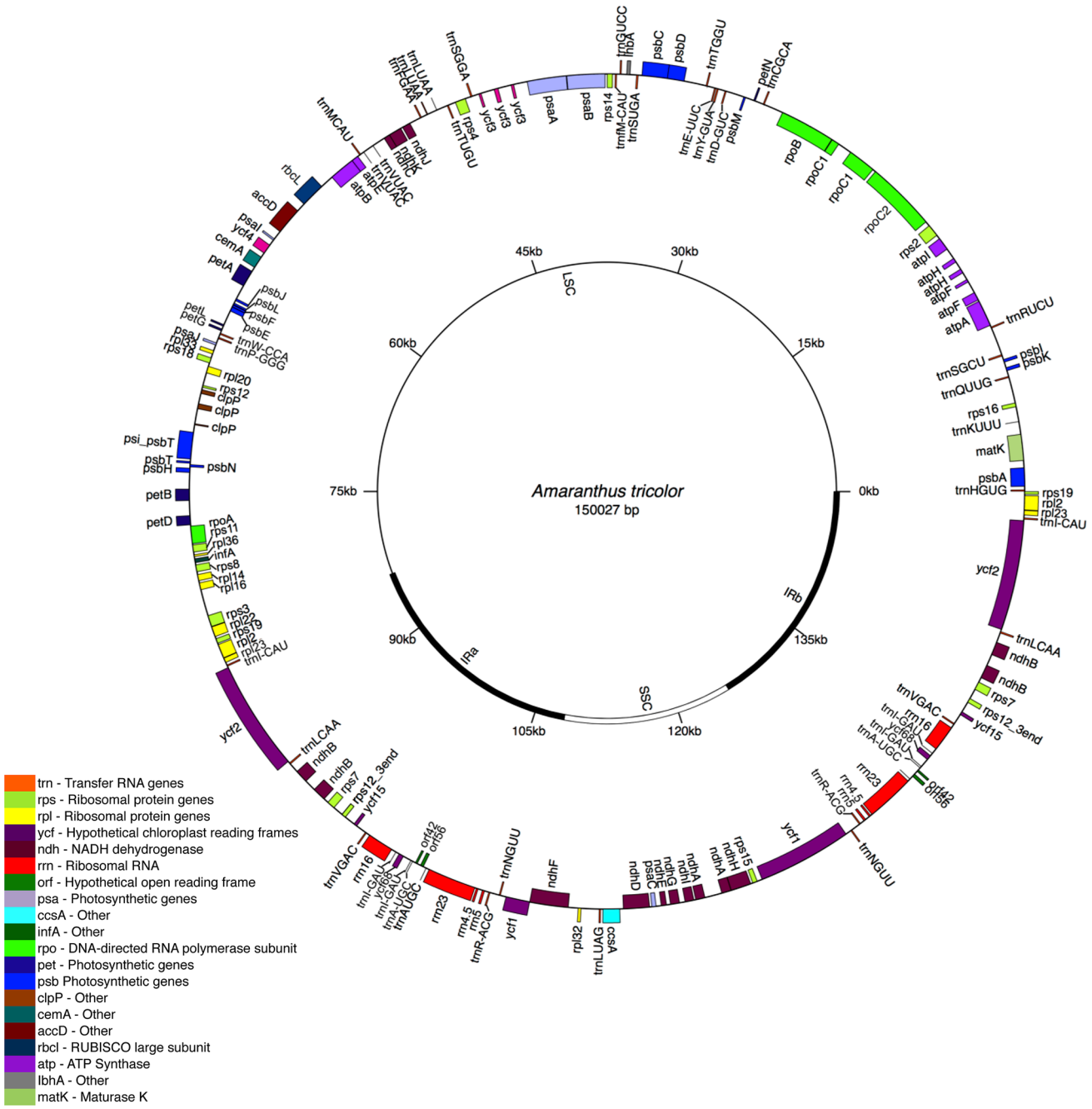


Fig. 1 Circular gene map for *Amaranthus tricolor* chloroplast genome. Genes indicated on the outer region are transcribed clockwise while genes on the inside are transcribed counter-clockwise. Genes with similar functions are grouped together and colour coded. (Color figure online)

Table 3 Classification of gene regions identified from the *Amaranthus tricolor* chloroplast genome

RNA genes						
Ribosomal RNA genes						
*rrn*16[a]	*rrn*23[a]	*rrn*4.5[a]	*rrn*5[a]			
Transfer RNA genes						
*trn*A-UGC[a,b]	*trn*C-GCA	*trn*D-GUC	*trn*E-UUC	*trn*F-GAA	*trn*M-CAU	*trn*G-UCC
*trn*H-GUG	*trn*I-CAU[a]	*trn*I-GAU[a,b]	*trn*K-UUU	*trn*L-CAA[a]	*trn*L-UAA[b]	*trn*L-UAG
*trn*M-CAU	*trn*N-GUU[a]	*trn*P-GGG	*trn*P-UGG	*trn*Q-UUG	*trn*R-ACG[a]	*trn*R-UCU
*trn*S-GCU	*trn*S-GGA	*trn*S-UGA	*trn*T-GGU	*trn*T-UGU	*trn*V-GAC[a]	*trn*V-UAC[b]
*trn*W-CCA	*trn*Y-GUA	*trnf*M-CAU				
Polypeptide genes						
Ribosomal protein genes						
*rpl*14	*rpl*16	*rpl*2[a]	*rpl*20	*rpl*22	*rpl*23[a]	*rpl*32
*rpl*33	*rpl*36					
*rps*11	*rps*12[e]	*rps*12_3end[a]	*rps*14	*rps*15	*rps*16	*rps*18
*rps*19[a]	*rps*2	*rps*3	*rps*4	*rps*7[a]	*rps*8	
*orf*42[a]	*orf*56[a]					
Transcription/translation genes						
*rpo*A	*rpo*B	*rpo*C1[b]	*rpo*C2	*inf*A		
Photosynthetic genes						
*rbc*L						
*psa*A	*psa*B	*psa*C	*psa*I	*psa*J		
*psb*A	*psb*C	*psb*D	*psb*E	*psb*F	*psb*H	*psb*I
*psb*J	*psb*K	*psb*L	*psb*M	*psb*N	*psb*T	*psi_psb*T
*pet*A	*pet*B	*pet*D	*pet*G	*pet*L	*pet*N	
*atp*A	*atp*B	*atp*E	*atp*F[b]	*atp*H[b]	*atp*I	
*ycf*3[c]	*ycf*4					
NAHD dehydrogenase genes						
*ndh*A[b]	*ndh*B[a,b]	*ndh*C	*ndh*D	*ndh*E	*ndh*F	*ndh*G
*ndh*H	*ndh*I	*ndh*J	*ndh*K			
Other Protein genes						
*acc*D	*clp*P[c]	*ccs*A	*mat*K	*cem*A	*ihb*A	
Open reading frames						
*ycf*1[d]	*ycf*15[ad]	*ycf*2[a]	*ycf*68[a]			

[a]Two copies due to inverted repeat

[b]Contains an intron

[c]Contains two introns

[d]Pseudogene

[e]Gene divided into two independent transcription units

GC contents of LSC and SSC were 34.5 and 30.2%, respectively. One hundred and nineteen genes were identified from the *A. tricolor* chloroplast genome (Table 3), 21 of which were duplicated in the inverted repeat regions, taking the total number of genes to 140. Based on the annotations, the highest percentage of genes was related to photosynthesis (28.6%, Table 3).

As a further validation step of the *A. tricolor* chloroplast sequence based on mapping to sugarbeet and subsequent gap-filling by Sanger sequencing, a second *A. tricolor* chloroplast sequence was assembled after mapping all the *A. tricolor* reads to the published *A. hypochondriacus* (KX279888.1) chloroplast sequence. A total of 8.1% *A. tricolor* reads mapped to the *A. hypochondriacus* chloroplast sequence, with an average coverage of 1272x. The two *A. tricolor* consensus chloroplast sequences produced by mapping to *B. vulgaris* or *A. hypochondriacus* chloroplasts were 99.8% identical, and thus the *A. tricolor* chloroplast sequence based on mapping to *B. vulgaris* was used for all further analyses.

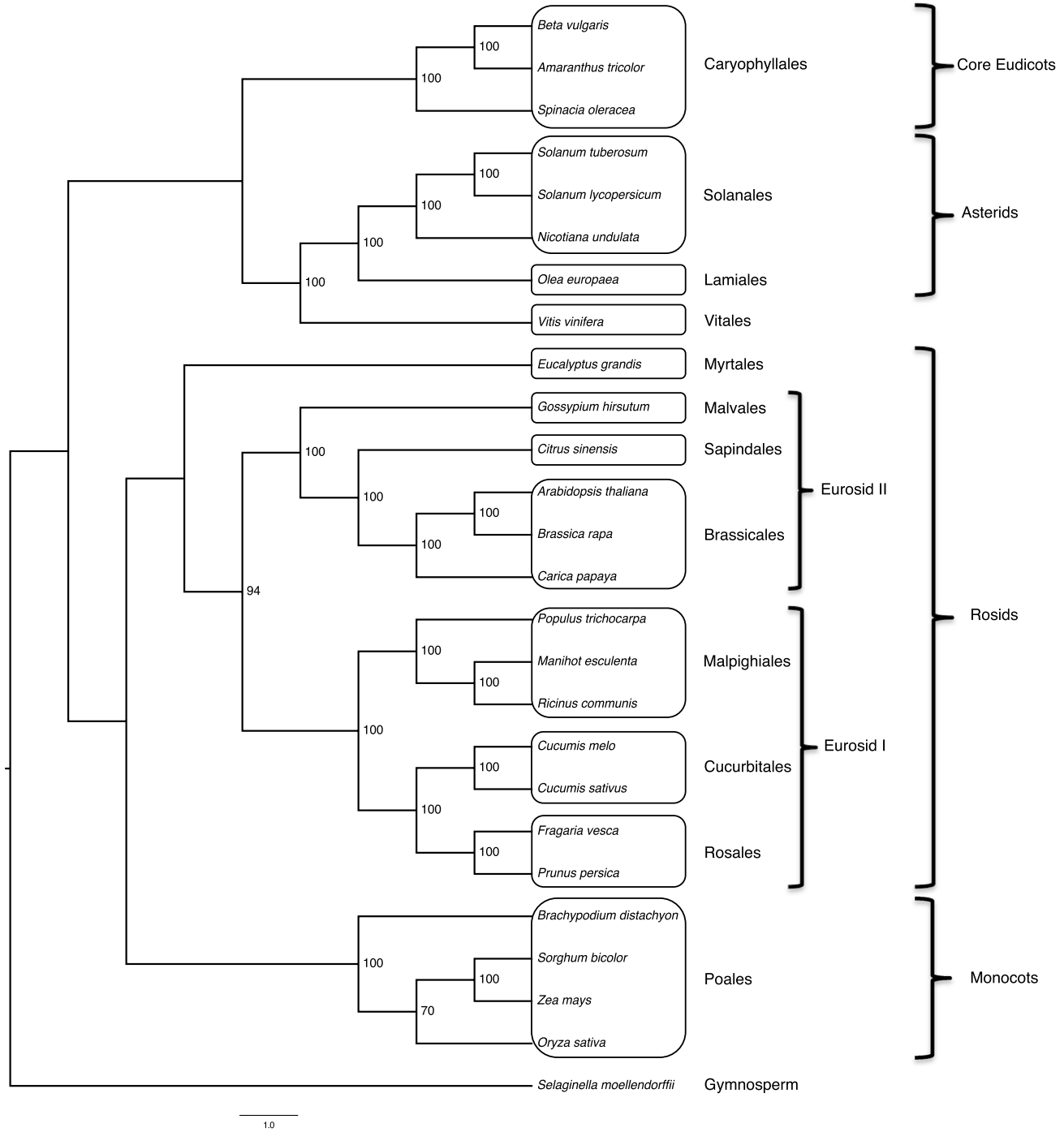


Fig. 2 Classification of species within the angiosperm plant group based on maximum likelihood phylogeny (evolutionary model = GTR, 1000 bootstrap replicates) of whole chloroplast genomes. Bootstrap confidence values (> 60%) are indicated at branch nodes. The addition of the newly assembled *A. tricolor* chloroplast genome reiterated its position within the Caryophyllales flowering plant order, and formed a sister branch to *Beta vulgaris* with 100% bootstrap support

Angiosperm Whole Plastid Genome Phylogeny

To confirm the phylogenetic position of *A. tricolor* alongside other angiosperm species, an alignment of 87,064 sites was obtained of which 72% (63,184 bp) sites were variable and 52% (32,817 bp) were parsimoniously informative. The topology of the phylogenetic tree generated (Fig. 2) was consistent with those from previous

studies of angiosperms using a combination of nuclear and plastid genes (Soltis et al. 1999; Kuzoff and Gasser 2000). The chloroplast sequence of *Amaranthus tricolor* grouped within the Caryophyllales clade [bootstrap (BS) = 100%] that included *Beta vulgaris* and *Spinacea oleraceae*, which represented the Amaranthaceae family.

Amaranthus Phylogeny Using Barcoding Genes

Chloroplast *mat*K, chloroplast *rbc*L and nuclear ITS gene region sequences from 45 GRIN accessions of the *Amaranthus* genus were determined. Alignment of the concatenated sequences (total length 2164 bp per accession) produced 80.8% constant, 16% variable, and 2.8% parsimoniously informative sites (Table 4). Phylogenetic relationships between the 45 *Amaranthus* accessions were inferred from all nucleotide sites using the partitioned maximum likelihood method based on the GTR + G (ITS and *mat*K) and HKY (*rbc*L) evolutionary models (Table 4).

The *Amaranthus* phylogeny determined by the partitioned maximum likelihood analysis was divided into five clades A–E (Fig. 3). Overall, clades A, B and C corresponded to species previously assigned to the *Amaranthus* subgenus *Amaranthus*, while clade D represented species within the *Amaranthus* subgenus *Albersia* (Mosyakin and Robertson 1996). Weedy amaranths (clade A), a mix of grain and leafy amaranths (clade B and C) and leafy amaranths (clade D) formed separate groupings in this phylogeny (Fig. 3).

Although only about one-third of the total nodes were supported with a BS > 60%, certain clades and sub-clades could be identified (Fig. 3). *Amaranthus retroflexus* (A1) shared a monophyletic origin with *A. powellii* (A2), while clade B was composed of one pure *A. hypochondriacus* subclade (B1), as well as *A. spinosus* (B2), *A. dubius* (B3) and a separate, mixed subclade composed of *A. hypochondriacus, A. hybridus* and *A. caudatus* accessions (B4). Clade B represented both leafy and grain amaranth and formed a sister group with clade C, which was composed of a mixture of species (*A. caudatus, A. quitensis* (C1) and *A. caudatus; A. cruentus; A. dubius; A. hybridus* and *A. hypochondriacus* (C2)). The bootstrap support values were lower than 20% within the entire subclade C2 and positive identifications based on nodal support was not possible.

Clade D (a paraphyletic group with the remaining amaranth accessions) indicates a close relationship between *A. tricolor* (D1) and *A. viridis* (D2), but bootstrap support for the monophyletic grouping was not acquired. Clade E: *A. blitum* formed a basal group to clades A, B and C. The remaining accessions (GRIN15, GRIN16—*A. graecizans*; GRIN28—*A. powellii*; GRIN7—*A. caudatus*; GRIN12—*A. dubius* and GRIN18—*A. hybridus*) did not fall within a clade and their identities remained uncertain. Overall, the phylogenetic tree generated from the barcoding genes was not well supported due to a lack of statistical confidence (above 60%) at more than 70% of the terminal nodes (Fig. 3).

A Bayesian analysis of the partitioned barcoding analysis is provided in Online Resource Fig. S2. Overall, the phylogeny was highly comparable to the maximum likelihood analysis, where several nodes were collapsed and had no posterior probability (PP) support for making positive identifications.

Amaranthus Whole Chloroplast Genome Phylogenetic Analysis

The chloroplast genome sequence for leafy amaranth *A. tricolor* was successfully constructed and subsequently served as a mapping reference for the chloroplast genomes of an additional 45 GRIN and 13 unknown SAG *Amaranthus* accessions. After strict mapping, there were 3% gaps on average for most accessions with GRIN14 having the least gapped sites (0.2%) and GRIN7 having the highest number of gapped sites (21%). Gapped regions were excluded from further analysis, regardless of whether other accessions had sequence information for those nucleotide positions.

Partitioned Analysis of 45 GRIN Accessions (Chloroplast Sequences)

The 45 previously identified GRIN accessions were investigated for their genetic relatedness and revealed better resolution of clade separation through whole chloroplast phylogenomics (both partitioned and unpartitioned datasets), than for barcoding analysis. For the unpartitioned analysis, the alignment of 45 GRIN amaranth chloroplast sequences (with IR_B removed) revealed a total alignment length of 246,850 bp. In the partitioned analysis, 119 genic regions

Table 4 Nucleotide site analysis of three barcoding genes from Amaranthus GRIN accessions

	Length (bp)	Constant sites	Variable sites	Parsimoniously informative sites		JModelTest
*mat*K	837	723	96	18	2.5%	GTR + G
*rbc*L	625	592	26	7	1.1%	HKY
ITS	702	434	231	37	5.3%	GTR + G
Total	2164	1749	353	62	2.8%	

on the chloroplast genome were included in the final alignment. The site characteristics before and after gap removal are presented in Table 5. Both partitioned and unpartitioned analysis resulted in highly similar tree topologies (data not shown).

Due to the highly similar tree topology resulting from both the partitioned and unpartitioned analyses, only the partitioned tree was further examined (Fig. 4). Overall, the four main clades (clades A to D) obtained from the barcoding gene tree were retained. These clades reflected weedy amaranths (clade A, BS = 96%, PP = 1), a combination of weedy and grain amaranths (clade B, BS = 89%, PP = 0.97), grain amaranths (clade C, BS = 90%, PP = 0.95) and leafy amaranths (clade D, BS = 100%, PP = 1).

Clade A was composed of subclade A1 (*A. powellii*) and subclade A2 (*A. retroflexus*) (Fig. 4). Each of the subclades was supported with BS = 100% and PP = 1. The grouping of these species mirrors the monophyly obtained from the barcoding analysis (Fig. 3). The grouping of accessions within clades A1 and A2 were supported with BS = 100%, respectively.

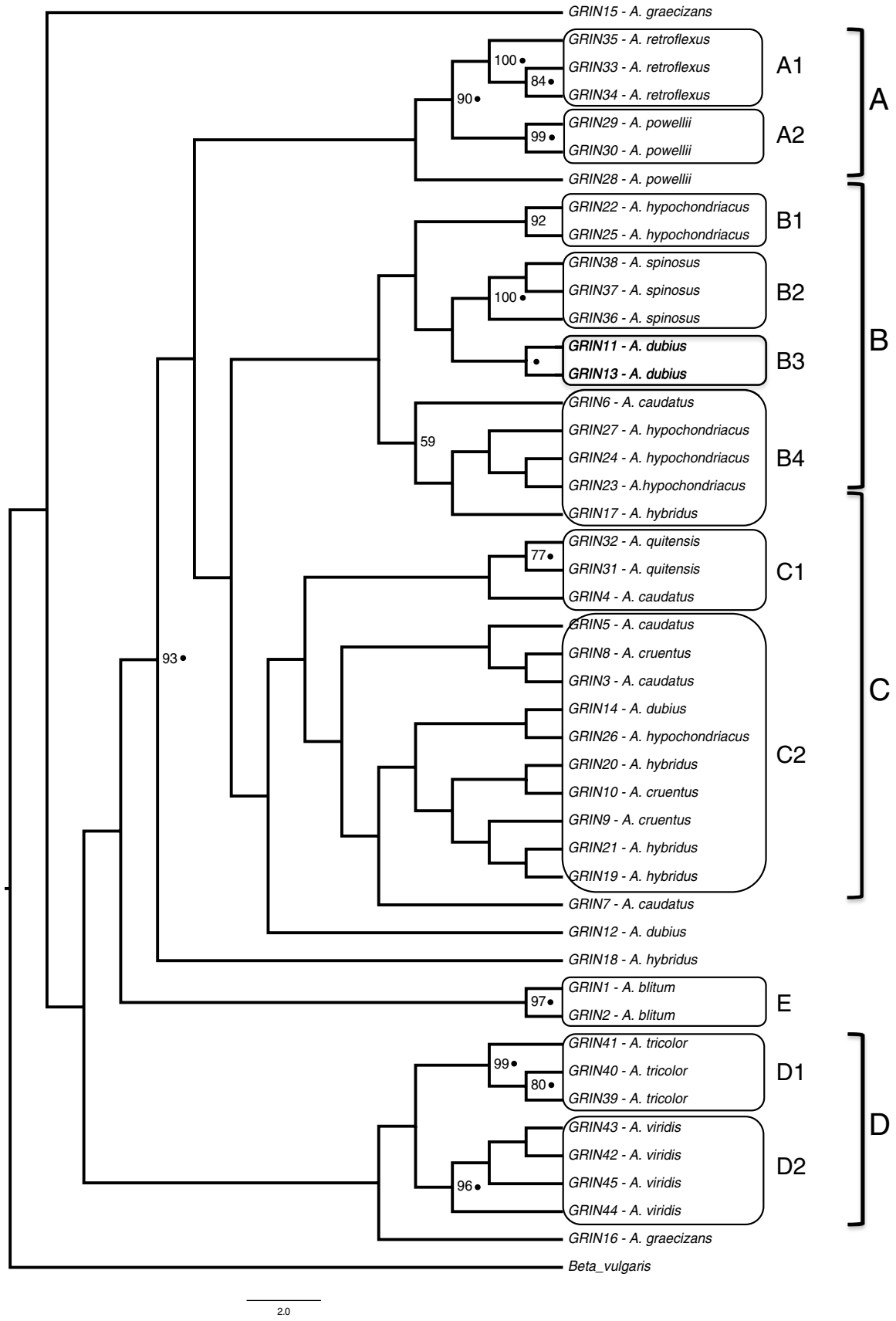
Clade B was comprised of three main subclades (B1 + B2, B3 and B5). Subclades B1 + B2 (BS = 95%, PP = 1) were combined (in contrast to the barcoding analysis, Fig. 3) and included three accessions of *A. spinosus* and two accessions of *A. hypochondriacus*. Subclade B3 (two accessions of *A. dubius*) was separated into a highly supported clade (BS = 100%) that formed a sister group to clade B1 + B2. Accessions residing in sub-clade B4 from the barcoding analysis (Fig. 3) had been redistributed among the remaining clades on the tree generated from the whole chloroplast analysis. Accessions GRIN23, GRIN24 and GRIN27 (*A. hypochondriacus*) now formed part of sub-clade C7 while GRIN 17 (*A. hypochondriacus*) fell within sub-clade C8. Accessions GRIN6 (*A. caudatus*) formed an outlier to subclades C7 and C6 (Fig. 4). Subclade B5 represented a group not observed during the analysis of the barcoding genes, and contained a mixture of *A. quitensis, A caudatus and A. hybridus* accessions.

Four subclades could be separated in clade C based on phylogenetic analysis of the chloroplast genome (Fig. 4). The subclades included in this main clade were: C5 (two accessions of *A. dubius*); C6 (two *A. cruentus*, one *A. hybridus*, one *A. hypochondriacus* and one *A. powellii* accession); C7 (three *A. hypochondriacus* and one *A. cruentus* accessions) and C8 (two accessions of *A. hybridus*). Subclades C1 and C2, which were observed in the barcoding analysis (Fig. 3), were not subsequently found in the tree obtained from the whole chloroplast analysis and their nodes were redistributed across newly formed subclades. Within subclade C6, accession GRIN8 (*A. cruentus*) could possibly have been reclassified as *A. hypochondriacus* due to the close association with three other *A. hypochondriacus* accessions, but there was little support for the node grouping these species. The *A. dubius* subclade (C5) was separated from the remaining subclades with BS = 90% and PP = 0.99, while *A. hybridus* (subclade C8) formed a polytomy at the node that groups clade C. Phylogenetic trees generated from this analysis indicated that subclades C6 and C7 have a monophyletic origin, but identification within each subclade remains difficult due to low internal nodal support of the majority of the branches. A mix of mainly *A. cruentus* and *A. hypochondriacus* accessions appeared in subclades C6 and C7 and conclusive identification of each individual accession remains elusive.

Clade D included four subclades, each supported with BS = 100% and PP = 1 that represented clear species groups (Fig. 4). These groups included subclades D1 (*A. tricolor*), subclade D2 (*A. viridis*), subclade D3 (*A blitum*) and subclade D4 (*A. graecizans*). *Amaranthus tricolor* (subclade D1) and *A. graecizans* (subclade D4) appeared to share a common ancestor BS = 100% and PP = 1. *Amaranthus blitum* (D3, previously clade E from the barcoding analysis (Fig. 3)) had a sister relationship (BS = 100%; PP = 1) with subclades D1 and D4. Clade E formerly only had a basal relationship to the D clade. *Amaranthus viridis* (subclade D2) formed a sister group with the monophyletic clade that included subclades D1, D3 and D4 (BS = 100%; PP = 1). Clade D exclusively represented species belonging to the leafy amaranth group. Whole chloroplast genome alignments were also subjected to Bayesian analysis (Online Resource Fig. S3). For 96% of the accessions, the Bayesian analysis was identical to the RAxML result but higher bootstrap support can be seen for 13 of the terminal nodes. Two accessions of *A. caudatus* (GRIN6 and GRIN8) formed a collapsed side branch to subclade C6, while in the RAxML analysis GRIN8 was part of subclade C7.

Partitioned Analysis of 45 GRIN and 14 SAG Accessions (Chloroplast Sequences)

Thirteen unknown SAG *Amaranthus* accessions were added to the GRIN analysis to investigate the possibility of their identification and classification. The total length of the alignment was 246,850 bp and the phylogenetic analysis was performed with RAxML and MrBayes software. Nucleotide site analysis for the alignments is presented in Table 6. The majority of the main clades and subclades (A, B, C and D) obtained from the GRIN analysis (Fig. 4) was observed for the combined analysis of GRIN and the unknown SAG accessions (Fig. 5). However, the inclusion of additional amaranth individuals resulted in the formation of three new subclades; C9, C10 (sister group to clade C7) and D5 (sister group to clades D1 and D4) (BS = 100%). The Bayesian analysis of this dataset revealed an identical topology and only minor differences in nodal support values for eight of the nodes could be observed (data not shown).


GRIN15 - A. graecizans
GRIN35 - A. retroflexus
GRIN33 - A. retroflexus
GRIN34 - A. retroflexus
A1
GRIN29 - A. powellii
GRIN30 - A. powellii
A2
A
GRIN28 - A. powellii
GRIN22 - A. hypochondriacus
GRIN25 - A. hypochondriacus
B1
GRIN38 - A. spinosus
GRIN37 - A. spinosus
GRIN36 - A. spinosus
B2
GRIN11 - A. dubius
GRIN13 - A. dubius
B3
B
GRIN6 - A. caudatus
GRIN27 - A. hypochondriacus
GRIN24 - A. hypochondriacus
GRIN23 - A.hypochondriacus
GRIN17 - A. hybridus
B4
GRIN32 - A. quitensis
GRIN31 - A. quitensis
GRIN4 - A. caudatus
C1
GRIN5 - A. caudatus
GRIN8 - A. cruentus
GRIN3 - A. caudatus
GRIN14 - A. dubius
GRIN26 - A. hypochondriacus
GRIN20 - A. hybridus
GRIN10 - A. cruentus
GRIN9 - A. cruentus
GRIN21 - A. hybridus
GRIN19 - A. hybridus
C2
C
GRIN7 - A. caudatus
GRIN12 - A. dubius
GRIN18 - A. hybridus
GRIN1 - A. blitum
GRIN2 - A. blitum
E
GRIN41 - A. tricolor
GRIN40 - A. tricolor
GRIN39 - A. tricolor
D1
GRIN43 - A. viridis
GRIN42 - A. viridis
GRIN45 - A. viridis
GRIN44 - A. viridis
D2
D
GRIN16 - A. graecizans
Beta_vulgaris
100
84
90
99
92
100
59
77
93
97
99
80
96
2.0

◂**Fig. 3** Amaranthus phylogeny based on DNA barcoding of chloroplast *mat*K, *rbc*L and nuclear ITS gene regions. The phylogeny was constructed using a partitioned maximum likelihood analysis (evolutionary model GTR + G (ITS, *mat*K) and HKY (*rbc*L), 1000 bootstrap replicates). Bootstrap confidence values (> 60%) were indicated at branch nodes. Bayesian probability values above 0.95 are indicated on nodes by (•). The genus is divided into three main subclades, broadly representing weedy amaranth (clade *A*), mix of grain and leafy amaranth (clade *B* and *C*) and leafy amaranth (clade *D* and *E*) accessions. GRIN—previously identified amaranth accessions (Table 1)

The partitioned chloroplast sequence analysis enabled species identifications for most of the SAG accessions (Fig. 5), and their updated identities have been included in Table 1. Firstly, the accession sequenced in this study, SAG29, together with SAG11 and SAG36 were confirmed to be *A. tricolor*, since they grouped in clade D1 with the three *A. tricolor* GRIN accessions. SAG1 was identified as an accession belonging to the *A. powellii* group with strong bootstrap support (clade A2). SAG14 was identified as *A. hybridus* since it forms a highly supported sub-clade C9 with GRIN18 (*A. hybridus*). SAG4 and SAG9 are members of the Hybridus complex, since they group with GRIN6 (*A. caudatus*) and GRIN8 (*A. cruentus*). SAG3 was identified as an accession of *A. quitensis* or *A. hybridus* based on its grouping with GRIN32 (*A. quitensis*) and GRIN21 (*A. hybridus*), and all three of these accessions originate from Brazil, which adds confidence to the identification. SAG34 was identified as *A. dubius* since it groups with two *A. dubius* accessions in sub-clade B3 with good bootstrap support. SAG7 groups with two accessions of *A. cruentus* (GRIN9 and GRIN10), with low bootstrap support, and is therefore cautiously classified as an accession of *A. cruentus*. SAG10, SAG12 and SAG17 are closely related in a newly formed sub-clade D5 and are separated from the monophyletic *A. tricolor*/*A. graecizans* clades (100% bootstrap). All three of these accessions were collected in southern Africa and BLAST sequence analysis of their *mat*K barcoding gene revealed a potential identification as *A. praetermissus* (data not shown) The accession SAG30 could not be identified since it is included in clade B with strong bootstrap support, but forms an outgroup to sub-clades B1/B2 and B3.

Investigation of Genes to Complement Existing Barcoding Genes for Phylogenetic Analysis of *Amaranthus* Species

A set of genic and intergenic regions of the whole chloroplast genome sequence were investigated for regions which could contribute to an informative phylogeny for *Amaranthus* without using whole chloroplast sequence data. Using the whole chloroplast alignment, 51 potential new barcoding regions were extracted that contained more than 1% parsimoniously informative characters across the total sequence length. Each individual region was joined to the original alignment of ITS, *mat*K and *rbc*L and a new sequence alignment was obtained. In total, 51 maximum likelihood phylogenetic trees were generated based on the newly selected gene together with the original barcoding genes; and were compared to the whole chloroplast genome phylogeny to assess the contribution of the gene to the overall phylogenetic signal by using the TOPD software (Table 2).

Firstly, when the disagreement between the original barcoding tree (ITS, *mat*K and *rbc*L) and the whole chloroplast tree was assessed, it was found that 27 out of 46 (58.7%) terminal node positions varied between the trees (blue bar in Online Resource Fig. S4). The split distance (the number of splits that disagree between the two trees) was calculated as 74%, which indicated that only 26% of the bipartitions were shared between the respective trees. Phylogenies were constructed using the original barcoding genes plus one extra chloroplast gene from the set of 51 additional barcoding genes. The best four genes (when added individually to the original barcoding genes) to produce phylogenies with the least disagreement (therefore highest congruence) to the whole chloroplast phylogenetic tree were found to be *ndh*D, *rpo*C2, *atp*E and *rpl*22 (43, 50, and 52% disagreement, respectively) (Table 2, Online Resource Fig. S4). Previous studies have used *ndh*D due to the number of informative sites available, but it seems to vary between genera (1.9% in *Amaranthus* vs. 19% in *Asteraceae*) (Panero and Funk 2008; Nock et al. 2011; Dong et al. 2013; Shaw et al. 2014). Due to their low level of agreement, *rpo*C2, *rpl*22 and *atp*E were not considered for further analysis.

The new barcoding phylogeny based on *ndhD* as well as ITS, *mat*K and *rbc*L (Online Resource Fig. S5) indicated improved clade placements (and therefore putative identifications) of five accessions compared to the original barcoding tree (GRIN31, GRIN32—*A. quitensis* and GRIN4, GRIN5, GRIN7—*A. caudatus*). However, the addition of the *ndh*D gene region also resulted in the placement of six accessions as outliers to defined clades and therefore hampered their identifications (GRIN8, GRIN9, GRIN10—*A. cruentus*; GRIN19, GRIN20—*A. hybridus* and GRIN26—*A. hypochondriacus*). Overall, the addition of *ndh*D did not significantly improve clade resolution within the *Amaranthus* genus. As a whole, the whole chloroplast phylogeny remained superior in resolving *Amaranthus* species relationships, compared to the barcoding analysis.

Discussion

The plant genus *Amaranthus* contains many economically important species, as well as potentially useful orphan crops, which can contribute to global food security. Several earlier

Table 5 Nucleotide site analysis of chloroplast sequences from *Amaranthus* GRIN accessions

	Length (bp)	Constant sites	Variable sites	Parsimoniously informative sites	
Unpartitioned analysis					
Chloroplast alignment with gaps	246,850	115,389	17,831	4296	1.7%
Chloroplast alignment with gaps removed	76,919	72,175	4744	809	1.0%
Partitioned analysis					
Chloroplast alignment with gaps	58,219	53,345	4338	764	1.3%
Chloroplast alignment with gaps removed	45,504	42,593	2911	481	1.0%

studies showed that the phylogenetic relationships between the *Amaranthus* species are highly intricate and difficult to resolve (Chan and Sun 1997; Xu and Sun 2001; Mandal and Das 2002; Costea et al. 2006; Mallory et al. 2008; Gerrano et al. 2015), especially among the grain and weedy types. Phylogenetic analysis based on datasets containing large numbers of DNA or amino acid characters can reveal increased resolution (and higher support) for clade hypotheses (Straub et al. 2012). In the past, phylogenetic analysis was constrained by the high cost of sequencing and was therefore limited to a few gene loci that were considered highly informative (mainly *mat*K, *rbc*L and ITS). However, with the advent of next-generation sequencing technologies, massive parallel sequencing has become the method of choice for rapid sequencing of plastid genomes (Parks et al. 2009) resulting in 2086 complete eukaryote chloroplast genome sequences being available in GenBank (26 February 2018, http://www.ncbi.nlm.nih.gov/genomes).

The *Amaranthus tricolor* Chloroplast Assembly

This study is the first to report a fully assembled and annotated chloroplast sequence of the leafy vegetable, *A. tricolor*. The chloroplast sequence was assembled by mapping to the chloroplast sequence of *B. vulgaris*, which is from a different genus of the family Amaranthaceae. Comparison to an *A. tricolor* chloroplast sequence based on mapping to the *A. hypochondriacus* chloroplast revealed high identity (99.8%). This indicates that whole chloroplast sequence assembly of a plant species does not require a reference genome from the same genus, since a chloroplast genome from the same family will suffice.

Sequence identities, gene organization and relative positions of the genes in *A. tricolor* and other angiosperm species were highly similar, corroborating the conserved nature of plant chloroplast genomes (Sugiura 1992; Schmitz-Linneweber et al. 2001). 9 of the 119 gene regions identified on the *A. tricolor* chloroplast genome contained a single intron (*trn*A-UGC, *trn*I-GAU, *trn*L-UAA, *trn*V-UAC, *rpo*C1, *atp*F, *atp*H, *ndh*A, *ndh*B), while two genes contained two introns each (*ycf*3, *clp*P). The remaining genes contained no intronic regions. The intron that was found in *trn*K-UUU of spinach (*Spinacea oleraceae*, Amaranthaceae), another leafy vegetable, was not found in *A. tricolor* (Chaney et al. 2016) nor in *Arabidopsis thaliana* (Sato et al. 1999), *Solanum tuberosum* (Chung et al. 2006) and *Nicotiana tabacum* (Shinozaki et al. 1986). The intron sizes for *A. tricolor* were more than 80% similar to the introns found in chloroplast genomes of *S. oleraceae* (Schmitz-Linneweber et al. 2001), *A. thaliana* (Sato et al. 1999), *S. tuberosum* (Chung et al. 2006) and *Artemisia frigida* (Liu et al. 2013). The only gene with a significant difference in intron size was *trn*L-UAA, which was 50% larger in *A. tricolor* than in *S. oleraceae* (Schmitz-Linneweber et al. 2001).

The presence of pseudogenes was observed in the *A. tricolor* chloroplast genome. Analysis of *ycf*1 and *ycf*15 showed that both had premature stop codons within the sequence. However, it is possible for these genes to still produce functional proteins after translation (Poliseno et al. 2010). As in plastids of other higher plants, there were potential open reading frames (ORFs) for which no functions have yet been inferred. Particular ORFs are conserved between different plant species (also known as hypothetical chloroplast reading frames—*ycf*) (Schmitz-Linneweber et al. 2001). The *A. tricolor* chloroplast genome harboured six *ycf* genes (*ycf*1, *ycf*2, *ycf*3, *ycf*4, *ycf*15 and *ycf*68) and the comparable predicted open reading frames suggested they may form similar polypeptides as *ycf* genes in other species. Dong et al. (2015) recently suggested that the *ycf*1 gene could be part of a protein channel present in the membrane of chloroplast cells and can also potentially be used as an additional plant barcode due to its high variability between taxa. In addition, two more open reading frames (*orf*42 and *orf*56) were identified for *A. tricolor*, which had 98% sequence identity between *A. tricolor, S. oleraceae*, and *S. tuberosum*.

Resolving the Position of *A. tricolor* Within the Angiosperm Phylogeny

Utilizing the large collection of angiosperm plastid sequences available on GenBank, it was possible to reconstruct a broad phylogeny of several different plant orders.

The newly assembled chloroplast of *A. tricolor* was placed within the *Caryophyllales* together with *B. vulgaris* and *S. oleraceae*, representing the Amaranthaceae family. The correct placement of *A. tricolor* is supported by previous work on two important morphological traits shared by species within the Amaranthaceae, namely the presence of betalain pigments conferring leaf/stem/flower colours instead of anthocyanins (Cuénoud et al. 2002; Venskutonis and Kraujalis 2013) and the C_4 photosynthetic machinery enabling these plants to thrive in warm, arid areas (Alemayehu et al. 2015), both of which have been reported for *A. tricolor* (Achigan-Dako et al. 2014).

Whole Chloroplast Analysis Provides Better Phylogenetic Resolution Within *Amaranthus* than DNA Barcoding

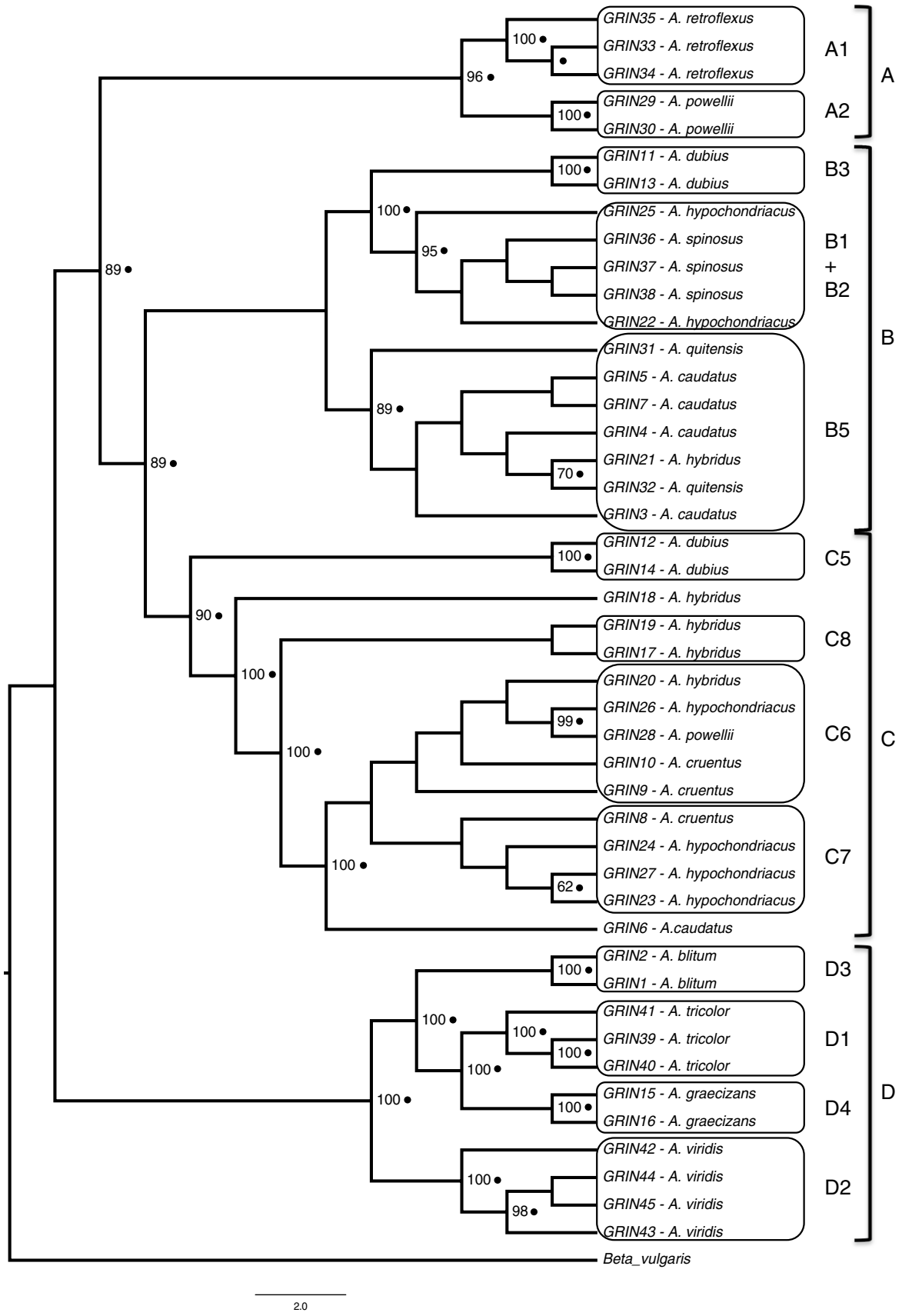
During this study, the complicated nature of *Amaranthus* phylogeny was demonstrated as previously reported when using AFLP's (Xu and Sun 2001), RAPD's (Mandal and Das 2002), combined AFLP and micromorphology (Costea et al. 2006), microsatellite markers (Mallory et al. 2008) and SNP marker analysis (Stetter and Schmid 2017). Chan and Sun (1997) initially revealed a close relationship between the grain *A. cruentus* and *A. hypochondriacus*; as well as the potential ancestor *A. hybridus* using RAPD analysis. In addition, *A. caudatus* formed a close sister clade together with a weedy *A. dubius* relative in their study. This result was highly congruent with the phylogeny obtained when genome-wide SNP data were used, indicating a mostly robust relationship between the grain amaranth species (*A. cruentus, A. caudatus* and *A. hypochondriacus*), their potential progenitors (*A. hybridus* and *A. quitensis*) and the wild *A. dubius* (Stetter and Schmid 2017). Our analysis corresponds in part with what was reported previously, where *A. cruentus* accessions formed a sister clade to *A. hypochondriacus* (Chan and Sun 1997). However, in contrast to the aforementioned studies, the *A. caudatus, A. hybridus* and *A. dubius* accessions did not conform to separate clades and were found scattered between and within the *A. cruentus* and *A. hypochondriacus* clades. In particular, the presence of *A. hybridus* within different clades implies that the complex had a recent split from the remaining species in the *Amaranthus* genus. Our results may in part be explained by the fact that the maternally inherited chloroplast sequences were used, and that better resolution between species that share the common ancestor of *A. hybridus*, which diverged recently, may be obtained from nuclear markers.

When more accessions of leafy and wild amaranth species are included in the phylogenies, a consistent grouping of *A. powellii* together with *A. retroflexus* was observed using microsatellite (Mallory et al. 2008) and SNP marker (Stetter and Schmid 2017) analysis. This was also reflected in the chloroplast analysis conducted in the present study. In all cases, the *A. powellii*/*A. retroflexus* clade form a sister clade to the Hybridus complex (overall consisting of *A. cruentus, A. hypochondriacus, A. caudatus, A. hybridus, A. quitensis* and *A. dubius*). These clades correlate well with members assigned to the *Amaranthus* subgenus *Amaranthus*, although the groupings within this subgenus are still mostly unresolved (Mosyakin and Robertson 1996). The addition of leafy amaranth accessions *A. tricolor*/*A. viridis* (which fall within one sub-clade) and *A. blitum* (forming a close sister clade to *A. tricolor*/*A. viridis*) in the present study resulted in a highly supported separated clade from the species discussed previously, and concurs with the results obtained during RAPD (Chan and Sun 1997) and SNP (Stetter and Schmid 2017) analysis. This clade of leafy amaranth corresponds to the previously defined *Amaranthus* subgenus *Albersia* (Mosyakin and Robertson 1996). The only species that did not show a consistent phylogeny in all the studies mentioned was *A. spinosus*, which grouped together with *A. powellii*/*A. quitensis* using RAPD markers (Chan and Sun 1997) and formed a completely separate clade during SNP analysis (Stetter and Schmid 2017). In the barcoding phylogeny obtained in our study, it grouped strongly with two *A. hypochondriacus* accessions.

Another aim of this study was to identify a universal set of chloroplast barcoding genes for *Amaranthus* species. Previous studies indicated that the *ndh*D gene, in combination with the other chloroplast barcoding genes, might have the potential to correctly delineate species (Panero and Funk 2008; Nock et al. 2011; Dong et al. 2013; Shaw et al. 2014). However, in *Amaranthus*, the *ndh*D gene had the same level of informativeness (1.9% parsimoniously informative sites) as the previously used chloroplast *mat*K (1.9%) and *rbc*L (1.1%) and did not enhance the barcoding phylogeny. In contrast to previous studies, a decreased number of accessions for which positive identifications could be made was observed. It is clear that future studies aiming at the development of genus/taxonomic specific barcodes in *Amaranthus* should focus on chloroplast intergenic regions, nuclear gene regions or other genetic markers.

Implication of the Low Phylogenetic Resolution to *Amaranthus* Phylogeny

The barcoding phylogenetic tree of the *Amaranthus* germplasm set generated from what is considered the 'core plant barcodes' (including *mat*K, *rbc*L and nuclear ITS gene regions) overall suffered from low resolution, making inferences regarding the phylogenetic relationships of the taxa problematic. The main significance of the resultant *Amaranthus* phylogeny is that standard plant barcodes may be best suited to substantiate existing species classifications, rather than having the power to discriminate between unknown


GRIN35 - A. retroflexus
GRIN33 - A. retroflexus
GRIN34 - A. retroflexus
GRIN29 - A. powellii
GRIN30 - A. powellii
GRIN11 - A. dubius
GRIN13 - A. dubius
GRIN25 - A. hypochondriacus
GRIN36 - A. spinosus
GRIN37 - A. spinosus
GRIN38 - A. spinosus
GRIN22 - A. hypochondriacus
GRIN31 - A. quitensis
GRIN5 - A. caudatus
GRIN7 - A. caudatus
GRIN4 - A. caudatus
GRIN21 - A. hybridus
GRIN32 - A. quitensis
GRIN3 - A. caudatus
GRIN12 - A. dubius
GRIN14 - A. dubius
GRIN18 - A. hybridus
GRIN19 - A. hybridus
GRIN17 - A. hybridus
GRIN20 - A. hybridus
GRIN26 - A. hypochondriacus
GRIN28 - A. powellii
GRIN10 - A. cruentus
GRIN9 - A. cruentus
GRIN8 - A. cruentus
GRIN24 - A. hypochondriacus
GRIN27 - A. hypochondriacus
GRIN23 - A. hypochondriacus
GRIN6 - A.caudatus
GRIN2 - A. blitum
GRIN1 - A. blitum
GRIN41 - A. tricolor
GRIN39 - A. tricolor
GRIN40 - A. tricolor
GRIN15 - A. graecizans
GRIN16 - A. graecizans
GRIN42 - A. viridis
GRIN44 - A. viridis
GRIN45 - A. viridis
GRIN43 - A. viridis
Beta_vulgaris
A1
A2
A
B3
B1 + B2
B5
B
C5
C8
C6
C7
C
D3
D1
D4
D2
D
2.0

◂**Fig. 4** Phylogeny of taxonomically described *Amaranthus* accessions (GRIN) based on chloroplast sequences. The analysis was based on a partitioned maximum likelihood analysis of 45 GRIN amaranth accessions based on whole chloroplast genome sequences. (Evolutionary model GTR + G, 1000 bootstrap replicates). Bootstrap confidence values (> 60%) are indicated at branch nodes. Bayesian probability values above 0.95 are indicated on nodes by (•). Four main clades are identified: (*A*) weedy amaranth (*A. retroflexus, A. powellii*); (*B*) weedy amaranth (*A. dubius, A. spinosus, A. quitensis*), grain amaranth (*A. hypochondriacus, A. caudatus*); (*C*) weedy amaranth (*A. dubius*), grain amaranth (*A. hypochondriacus, A. cruentus*), leafy amaranth (*A. hybridus*) and (*D*) leafy amaranth (*A. blitum, A. tricolor, A. graecizans, A. viridis*)

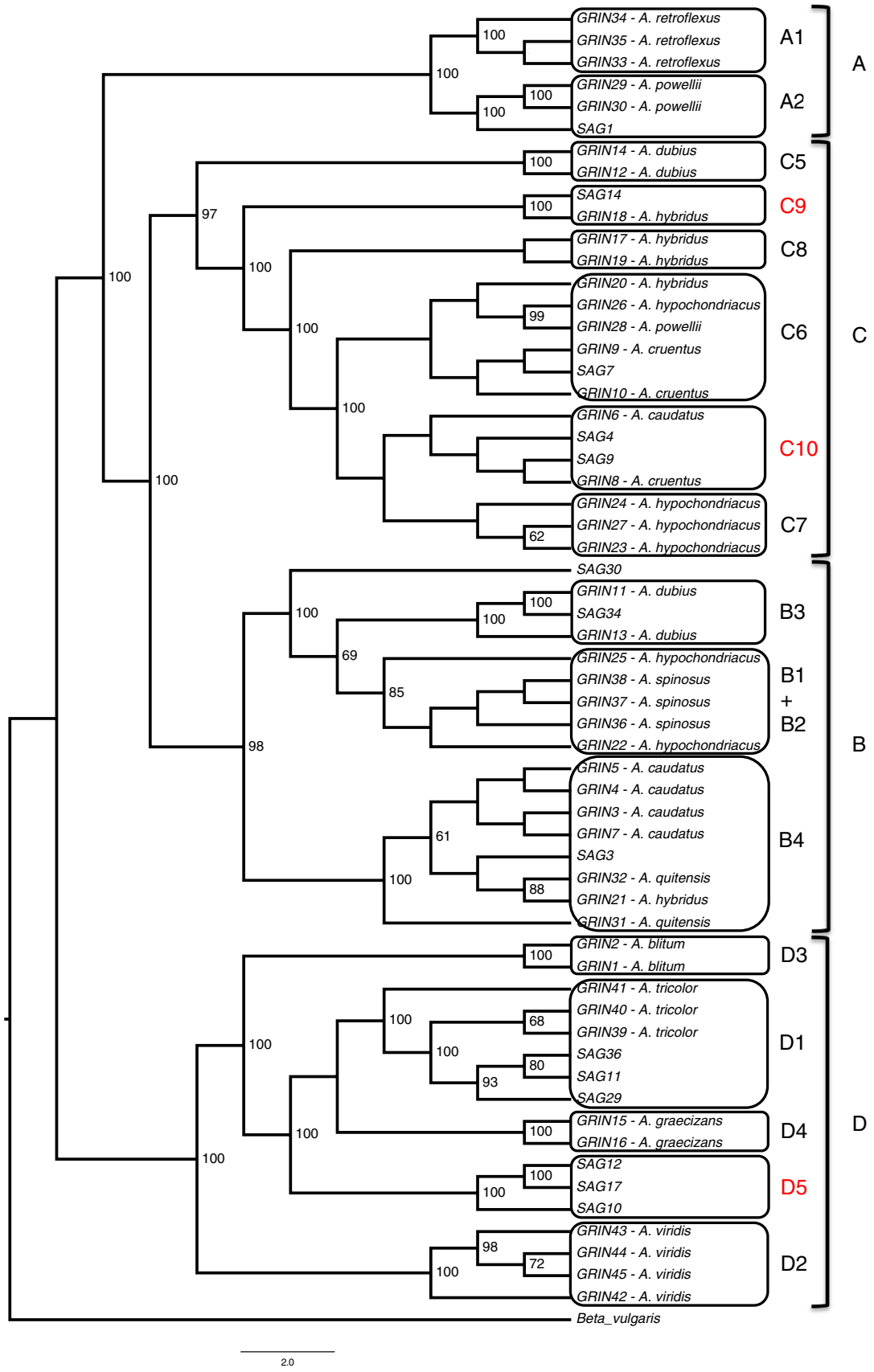
inter- and intra-species accessions as previously reported (Hollingsworth et al. 2016). The low resolution obtained from the barcoding phylogenetic analyses suggested that the diversification of particularly the grain amaranths was fairly recent, and that an incomplete domestication event is observed as was reported by (Stetter et al. 2017). This warrants further investigation into an integrative approach of morphology and additional informative nuclear or chloroplast gene regions to differentiate between the amaranths at interspecies level.

In our study, we attempted to improve the species tree observed with the barcoding analysis by conducting a whole chloroplast phylogeny. The results revealed that the inclusion of more informative sites (and a diverse array of species) leads to much greater clade resolution and support in the genus representation. Based on entire plastomes, *A. dubius*/*A. hypochondriacus*/*A. cruentus*/*A. hybridus* (clade C) is sister to *A. caudatus*/*A. quitensis*/*A. spinosus*/*A. hybridus* (clade B), while *A. powellii*/*A. retroflexus* (clade A) forms a paraphyletic relationship to clades B and C. The aforementioned clades (A, B and C) represent members of the *Amaranthus* subgenus *Amaranthus* (Mosyakin and Robertson 1996; Wassom and Tranel 2005). Clade B represents the grain amaranth species *A. caudatus*, together with its putative progenitors *A. quitensis* and *A. hybridus* (representing accessions collected from South America), while the remaining two grain *Amaranthus* species (*A. cruentus* and *A. hypochondriacus*) and a putative progenitor (*A. hybridus*) are present in clade C (representing species from Central America). The split between South American and Central American species was highly supported in the phylogeny, indicating that geographical separation probably led to allopatric speciation. Grain amaranths were mostly produced through human intervention, deliberate domestication and occasional accidental wild interspecies crosses (Stetter and Schmid 2017), which is evident in the low levels of genetic diversity seen between *A. hypochondriacus, A. cruentus, A. caudatus* and *A. hybridus*. The *A. powellii* and *A. retroflexus* sister lineage (clade A) was also observed when analysing genetic diversity with nuclear SSR markers (Mallory et al. 2008). Since clade A (*A. powellii* accessions) is separated from clade B/C (*A. hypochondriacus* accessions) with strong statistical support, the previously reported hypothesis that *A. powellii* is a progenitor to *A. hypochondriacus* becomes highly unlikely (Sauer 1967). An exception is the *A. powellii* accession from Mexico (GRIN28); however, it groups with *A. hypochondriacus* GRIN26 from Mexico in clade C6 and other *A. hypochondriacus* accessions in clade C, and thus GRIN28 may be misidentified.

The groupings observed for clade B and C closely mirror results presented in a recent study utilizing whole genome SNP marker data for different amaranth species (Stetter and Schmid 2017). The combination of *A. caudatus, A. cruentus, A. hypochondriacus, A. hybridus* and *A. quitensis* are often referred to as the Hybridus complex and are notoriously difficult to classify (Stetter and Schmid 2017). Clade C and clade B5 of our chloroplast phylogeny represents the Hybridus complex, since it contains accessions of all five of these species (Fig. 4), and we ascribe the distinction between the two clades to geographical separation during the ongoing domestication process. In this vein, the chloroplast phylogeny supported the hypothesis that *A. hybridus* and *A. quitensis* contributed to the incomplete domestication of *A. caudatus* in a particular geographic region (Stetter et al. 2017), since all three species are grouped together in clade B5, including closely related *A. hybridus* and *A. quitensis* accessions from Brazil (Fig. 4). The *A. hybridus* accession GRIN19 from a different geographic region (Mexico in Central America) grouped in a different clade C8. Stetter et al. (2017 also highlighted the role of geographic separation,

Table 6 Nucleotide site analysis of chloroplast sequences from *Amaranthus* GRIN and SAG accessions

	Length (bp)	Constant sites	Variable sites	Parsimoniously informative sites	
Unpartitioned analysis					
Chloroplast alignment with gaps	246,850	115,312	19,154	5246	2.1%
Chloroplast alignment with gaps removed	76,857	72,013	4844	914	1.2%
Partitioned analysis					
Chloroplast alignment with gaps	58,792	53,289	4202	851	1.4%
Chloroplast alignment with gaps removed	45,305	42,541	2764	537	1.2%


GRIN34 - A. retroflexus
GRIN35 - A. retroflexus
GRIN33 - A. retroflexus
A1
GRIN29 - A. powellii
GRIN30 - A. powellii
SAG1
A2
A
GRIN14 - A. dubius
GRIN12 - A. dubius
C5
SAG14
GRIN18 - A. hybridus
C9
GRIN17 - A. hybridus
GRIN19 - A. hybridus
C8
GRIN20 - A. hybridus
GRIN26 - A. hypochondriacus
GRIN28 - A. powellii
GRIN9 - A. cruentus
SAG7
GRIN10 - A. cruentus
C6
C
GRIN6 - A. caudatus
SAG4
SAG9
GRIN8 - A. cruentus
C10
GRIN24 - A. hypochondriacus
GRIN27 - A. hypochondriacus
GRIN23 - A. hypochondriacus
C7
SAG30
GRIN11 - A. dubius
SAG34
GRIN13 - A. dubius
B3
GRIN25 - A. hypochondriacus
GRIN38 - A. spinosus
GRIN37 - A. spinosus
GRIN36 - A. spinosus
GRIN22 - A. hypochondriacus
B1 + B2
GRIN5 - A. caudatus
GRIN4 - A. caudatus
GRIN3 - A. caudatus
GRIN7 - A. caudatus
SAG3
GRIN32 - A. quitensis
GRIN21 - A. hybridus
GRIN31 - A. quitensis
B4
B
GRIN2 - A. blitum
GRIN1 - A. blitum
D3
GRIN41 - A. tricolor
GRIN40 - A. tricolor
GRIN39 - A. tricolor
SAG36
SAG11
SAG29
D1
GRIN15 - A. graecizans
GRIN16 - A. graecizans
D4
SAG12
SAG17
SAG10
D5
GRIN43 - A. viridis
GRIN44 - A. viridis
GRIN45 - A. viridis
GRIN42 - A. viridis
D2
D
Beta_vulgaris
2.0

◂**Fig. 5** Phylogenetic distribution of unknown *Amaranthus* accessions (SAG) in relation to previously described *Amaranthus* species (GRIN) based on chloroplast sequences. The analysis was based on a partitioned maximum likelihood analysis of 45 GRIN and 14 SAG amaranth accessions (evolutionary model GTR + G, 1000 bootstrap replicates) of whole chloroplast genome sequences. Bootstrap confidence values (> 60%) are indicated at branch nodes. Four main clades are identified: (*A*) weedy amaranth (*A. retroflexus, A. powellii*); (*B*) weedy amaranth (*A. dubius, A. spinosus, A. quitensis*), grain amaranth (*A. hypochondriacus, A. caudatus*); (*C*) weedy amaranth (*A. dubius*), grain amaranth (*A. hypochondriacus, A. cruentus*), leafy amaranth (*A. hybridus*) and (*D*) leafy amaranth (*A. blitum, A. tricolor, A. graecizans, A. viridis*). The addition of unknown amaranth accessions (SAG) results in the formation of additional sub-clades C9, C10 and D5

since they observed a population of closely related *A. hybridus* and *A. quitensis* accessions (Peruvian amaranth) that was distinct from a population of *A. hybridus* and *A. quitensis* in Ecuador.

The grouping of accessions into sub-clades B3 and B1 + B2 are indicative of a deeper level of identification that would not be possible based solely on morphological analysis. In this study, GRIN11 and GRIN13 (sub-clade B3) were originally identified as *A. dubius* accessions but did not cluster together with the other *A. dubius* accessions (GRIN12 and GRIN14, sub-clade C5). Being a known allotetraploid, it has been suggested that *A. dubius* could be a hybrid to which *A. spinosus* had contributed one chromosome set and the other parent could be either *A. quitensis* or *A. hybridus* (Sauer 1967). This hypothesis was supported by the chloroplast tree, as GRIN11 and GRIN13 form a confident sister lineage to *A. spinosus*. Further studies are, however, needed to confidently confirm the identification of GRIN12 and GRIN14 as *A. dubius* (due to their placement in sub-clade C5 instead of sub-clade B3) by determining their genome size. *Amaranthus dubius* is the only allotetraploid amaranth species (Stetter and Schmid 2017) investigated during this study; therefore, similar genome sizes of GRIN12 and GRIN14 would be indicative of these accessions being conspecific with *A. dubius*. The inclusion of two *A. hypochondriacus* accessions within sub-clade B1 + B2 together with *A. spinosus* has been observed before during chloroplast and nuclear gene phylogenies and it is suspected that these may be spineless versions of *A. spinosus* rather than true *A. hypochondriacus* accessions (Waselkov 2013).

Within clade C of the whole chloroplast phylogeny, sub-clades C6, C7 and C8 formed poorly resolved clades. In these clades, the species boundaries between the *A. hybridus, A. hypochondriacus* and *A. cruentus* accessions could not be determined with confidence. The low level of genetic diversity with sub-clades C6, C7 and C8 could indicate very recent hybridization or domestication events, possibly due to self-hybridization of *A. hybridus* as previously suggested (Sauer 1967). The low genetic diversity could also be attributed to low temporal resolution provided by chloroplast data. To further validate the genetic diversity, additional nuclear markers should be included.

Accessions residing in clade D formed a distant lineage to clades A, B, and C. Within this clade, *A. tricolor* and *A. graecizans* (leafy amaranths) were monophyletic, while *A. blitum* and *A. viridis* formed close sister groups. These species are thought to be introductions of Asian and European origin (classified within *Amaranthus* subgenus *Albersia*), which would support their robust separation from the Central and South American grain/weedy amaranth types (Mosyakin and Robertson 1996). The high resolution of terminal nodes in leafy and weedy amaranths (clades A and D) indicate older, more stably integrated diversification events.

Conclusion

This study is the first to report a complete chloroplast sequence of the leafy vegetable, *A. tricolor*. Phylogenetic trees based on additional chloroplast assemblies of a diverse range of amaranth accessions confirmed a moderately resolved phylogeny for the grain amaranths and a highly resolved phylogeny for most of the weedy and leafy amaranths. The two main weedy amaranths (*A. retroflexus* and *A. powellii*) resided within the same clade based on three different phylogenetic analyses (original barcoding, new barcoding and whole chloroplast phylogeny). The *A. retroflexus* and *A. powellii* species formed a sister lineage to the commonly known Hybridus complex consisting mainly of *A. hypochondriacus, A. cruentus, A. caudatus, A. hybridus, A. quitensis* and *A. dubius*. Within the Hybridus complex, separation of species groups was still not adequate to make confident classifications. Broadly, *A. caudatus* consistently grouped with *A. quitensis* and selected *A. hybridus* accessions, while *A. cruentus* and *A. hypochondriacus* grouped with alternative *A. hybridus* accessions. This is in line with the conclusion of Stetter et al. (2017) that *A. caudatus* accessions are the result of partial domestication from *A. hybridus* including gene flow from *A. quitensis*. *A. hypochondriacus* and *A. cruentus* could be domesticated versions of different geographical isolates of *A. hybridus*, as previously reported (Mallory et al. 2008; Stetter and Schmid 2017). The grouping of all the aforementioned species (*A. caudatus, A. cruentus, A. hypochondriacus, A. hybridus, A. quitensis, A. powellii* and *A. retroflexus*) conform to the initially described *Amaranthus* subgenus *Amaranthus* (Mosyakin and Robertson 1996). The leafy amaranths seem to be stable in their genetic content, by revealing the same robust topology between the barcoding and the whole chloroplast analysis. Whole chloroplast sequence analysis also facilitated the identification of unknown *Amaranthus* accessions in the South African genebank (Fig. 5; Table 1).

Further investigations of the whole chloroplast sequence to identify additional gene barcodes for in-depth phylogenies, however, proved unsuccessful. None of the potential "new" barcoding regions possessed adequate polymorphic content to discriminate between accessions to the level of whole chloroplast analysis. The leafy amaranth accessions (*A. tricolor, A. viridis, A. blitum* and *A. graecizans*) conform to the initially described *Amaranthus* subgenus *Albersia* (Mosyakin and Robertson 1996), and was separated from the subgenus *Amaranthus* with high support in the current phylogeny (BS = 100%).

The results of this study indicate that plastomes contain the discriminatory power to separate *Amaranthus* accessions into different species groups with a high level of confidence. However, a number of nodes in the phylogenetic trees obtained in this study suffered from low statistical support. A future step to obtain a species tree with stronger resolution (especially within the Hybridus complex) would be to incorporate nuclear (sequence and marker) data and whole mitochondrial sequence phylogeny. Nuclear data would be particularly useful to investigate incongruent species placements due to their divergent histories, especially since most of the species were collected from geographical areas with no prior known selection pressures. Previous studies have shown that the development of nuclear SNP markers are extremely useful in constructing a highly informative phylogeny of the *Amaranthus* genus, both to group the species within their respective sub-genera and to identify highly differentiated groups within each sub-genus (Stetter and Schmid 2017). Whereas the chloroplast is mostly maternally inherited, the nuclear genome can shed light in historical recombination and hybridization events (Nikiforova et al. 2013). Since evidence of hybridizations and ongoing gene flow between amaranth species exist, care should be taken when investigating nuclear markers. Polyploid species (such as *A. dubius*) may be characterized by a high level of heterozygosity, and the presence of multiallelic SNPs can influence phylogenetic interpretations.

The plastome is extremely useful in phylogenetic analysis due to its relatively small size, the conserved gene order and content across different plant families, high copy number in plant cells, the absence of recombination and mostly uniparental inheritance (Davis et al. 2014; Hollingsworth et al. 2016). In addition, the intergenic regions of the chloroplast usually have higher mutation rates than the genic regions, allowing more informative phylogenies over a longer time scale (Nock et al. 2011). Furthermore, the chloroplast sequences could be used to investigate unique SNP markers for each species group. Instead of a genic or intergenic barcode, a SNP barcode could be developed. By sequencing and assembling whole chloroplasts of representative species from a genus, it would be feasible to identify species-specific SNPs. As illustrated by Chaney et al. (2016) and Lightfoot et al. (2017), long read next-generation sequencing technologies such as PacBio or Oxford Nanopore provide high-quality sequences in highly repetitive regions such as the chloroplast IRa and IRb regions. Consequently, phylogenomics has potential to resolve phylogenetically difficult plant families.

Future studies of the *Amaranthus* genus should also focus on the addition of more species within the Hybridus complex. These additional species should include accessions representing their native and introduced geographical ranges to increase species sampling and genetic variation associated by geographical separation. Care should also be taken to ensure that one species is not overrepresented, to fully understand the underlying genetic diversity and the complex relatedness of different species groups. In this manner, a more complete picture will be obtained of grain amaranth domestication and the role of their weedy ancestors.

This study highlights the great potential of next-generation sequencing for the study of plant species evolution. The identification and classification of *Amaranthus* accessions in this study will be an important tool to provide resources in terms of positively identified breeding lines for investigating nutritional, biochemical, biotic and abiotic resistance and medicinal traits naturally found in the *Amaranthus* genus.

Acknowledgements The authors wish to thank the Department of Science and Technology of South Africa, the National Research Foundation and the Professional Development Program of the Agricultural Research Council (ARC) in South Africa for providing funding for the PhD study from where this work originated. The authors also thank Dr Charles Hefer at the ARC for bioinformatics support. The authors thank Mr Willem Jansen van Rensburg and his staff at the ARC Vegetable and Ornamental Plant Institute for providing the *Amaranthus* germplasm set (SAG) as well as plant maintenance. The authors thank the Core Facility team at the ARC Biotechnology Platform for DNA sequencing.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Achigan-Dako EG, Sogbohossou OE, Maundu P (2014) Current knowledge on *Amaranthus* spp.: research avenues for improved nutritional value and yield in leafy amaranths in sub-Saharan Africa. Euphytica 197:303–317

Akond M, Islam S, Wang X (2013) Genotypic variation in biomass traits and cell wall components among 35 diverse accessions of Amaranthaceae family. J Appl Phytotechnol Environ Sanit 2:37–45

Alamgir M, Kibria M, Islam M (2011) Effects of farm yard manure on cadmium and lead accumulation in Amaranth (*Amaranthus oleracea* L.). J Soil Sci Environ Manag 2:237–240

Alemayehu RF, Bendevis MA, Jacobsen SE (2015) The potential for utilizing the seed crop amaranth (*Amaranthus* spp.) in East Africa

as an alternative crop to support food security and climate change mitigation. J Agron Crop Sci 201:321–329
Barrett CF, Davis JI, Leebens-Mack J, Conran JG, Stevenson DW (2013) Plastid genomes and deep relationships among the commelinid monocot angiosperms. Cladistics 29:65–87
Bell KL, de Vere N, Keller A, Richardson RT, Gous A, Burgess KS, Brosi BJ (2016) Pollen DNA barcoding: current applications and future prospects 1. Genome 59:629–640
Bezeng B, Davies TJ, Daru BH, Kabongo RM, Maurin O, Yessoufou K, van der Bank H, Van der Bank M (2017) Ten years of barcoding at the African Centre for DNA barcoding. Genome 60:629–638
Braukmann TW, Kuzmina ML, Sills J, Zakharov EV, Hebert PD (2017) Testing the efficacy of DNA barcodes for identifying the vascular plants of Canada. PLoS ONE 12:e0169515
Brenner DM, Baltensperger DD, Kulakow PA, Lehmann JW, Myers RL, Slabbert MM, Sleugh BB (2000) Genetic resources and breeding of *Amaranthus*. In: Janick J (ed) Plant breeding reviews, vol 19. Wiley, New York, pp 227–285
Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC, Newmaster SG, Percy DM, Hajibabaei M, Barrett SC (2011) Discriminating plant species in a local temperate flora using the *rbcL* + *mat*K DNA barcode. Methods Ecol Evol 2:333–340
Chan K, Sun M (1997) Genetic diversity and relationships detected by isozyme and RAPD analysis of crop and wild species of *Amaranthus*. TAG 95:865–873
Chaney L, Mangelson R, Ramaraj T, Jellen EN, Maughan PJ (2016) The complete chloroplast genome sequences for four Amaranthus species (Amaranthaceae). Appl Plant Sci 4:1600063
Chung H-J, Jung JD, Park H-W, Kim J-H, Cha HW, Min SR, Jeong W-J, Liu JR (2006) The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. Plant Cell Rep 25:1369–1379
Costea M, Brenner DM, Tardif FJ, Tan YF, Sun M (2006) Delimitation of *Amaranthus cruentus* L. and *Amaranthus caudatus* L. using micromorphology and AFLP analysis: an application in germplasm identification. Genet Resour Crop Ev 53:1625–1633
Cuénoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbc*L, *atp*B, and *mat*K DNA sequences. Am J Bot 89:132–144
Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772–772
Das S (2011) Systematics and taxonomic delimitation of vegetable, grain and weed amaranths: a morphological and biochemical approach. Genet Resour Crop Evol 59:289–303
Davis CC, Xi Z, Mathews S (2014) Plastid phylogenomics and green plant phylogeny: almost full circle but not quite there. BMC Biol 12:11–15
Dong W, Xu C, Cheng T, Lin K, Zhou S (2013) Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales. Genome Biol Evol 5:989–997
Dong W, Liu H, Xu C, Zuo Y, Chen Z, Zhou S (2014) A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. BMC Genet 15:138
Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S (2015))*ycf*1, the most promising plastid DNA barcode of land plants. Sci Rep 5:8348
Ebert AW (2014) Potential of underutilized traditional vegetables and legume crops to contribute to food and nutritional security, income and more sustainable production systems. Sust 6:319–335
Gerrano AS, van Rensburg WSJ, Adebola PO (2015) Genetic diversity of *Amaranthus* species in South Africa. S. Afr J Plant Soil 32:39–46
Gudu S, Gupta V (1988) Male-sterility in the grain amaranth (*Amaranthus hypochondriacus* ex-Nepal) variety Jumla. Euphytica 37:23–26
Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704
Hackett SJ, Kimball RT, Reddy S, Bowie RC, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J (2008) A phylogenomic study of birds reveals their evolutionary history. Science 320:1763–1768
Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS ONE 6:e19254
Hollingsworth PM, Li D-Z, van der Bank M, Twyford AD (2016) Telling plant species apart with DNA: from barcodes to genomes. Phil Trans R Soc B 371:20150338
Huang H, Shi C, Liu Y, Mao S-Y, Gao L-Z (2014) Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol Biol 14:151–168
Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755
Kavita P, Gandhi P (2017) Rediscovering the therapeutic potential of *Amaranthus* species: a review. Egypt J Basic Appl Sci 4:196–205
Kim JS, Kim JH (2013) Comparative genome analysis and phylogenetic relationship of order Liliales insight from the complete plastid genome sequences of two Lilies (*Lilium longiflorum* and *Alstroemeria aurea*). PLoS One 8:e68180
Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbc*L gene complements the non-coding *trn*H-*psb*A spacer region. PLoS ONE 2:e508
Kuzoff RK, Gasser CS (2000) Recent progress in reconstructing angiosperm phylogeny. Trends Plant Sci 5:330–336
Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V (2008) DNA barcoding the floras of biodiversity hotspots. PNAS 105:2923–2928
Li H, Cao H, Cai Y-F, Wang J-H, Qu S-P, Huang X-Q (2014) The complete chloroplast genome sequence of sugar beet (*Beta vulgaris ssp. vulgaris*). Mitochondrial DNA 25:209–211
Lightfoot DJ, Jarvis DE, Ramaraj T, Lee R, Jellen EN, Maughan PJ (2017) Single-molecule sequencing and Hi-C-based proximity guided assembly of amaranth (Amaranthus hypochondriacus) chromosomes provide insights into genome evolution. BMC Biol 15:74
Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young HA, Feng X, Gu YQ (2013) Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. PLoS ONE 8:e57533
Lubbe E, Rodda N (2016) Effects of greywater irrigation on germination, growth and photosynthetic characteristics in selected African leafy vegetables. Water SA 42:203–212
Ma P-F, Zhang Y-X, Zeng C-X, Guo Z-H, Li D-Z (2014) Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). Syst Biol 63:933–950
Mallory MA, Hall RV, McNabb AR, Pratt DB, Jellen EN, Maughan PJ (2008) Development and characterization of microsatellite markers for the grain amaranths. Crop Sci 48:1098–1106
Mandal N, Das P (2002) Intra-and interspecific genetic diversity in grain *Amaranthus* using random amplified polymorphic DNA markers. Plant Tissue Cult 12:49–56
Maughan PJ, Yourstone SM, Jellen EN, Udall JA (2009) SNP discovery via genomic reduction, barcoding, and 454-pyrosequencing in amaranth. The Plant Gen J 2:260–270

Mlakar SG, Turinek M, Jakop M, Bavec M, Bavec F (2010) Grain amaranth as an alternative and perspective crop in temperate climate. J Geogr 5:135–145

Mnkeni A, Masika P, Maphaha M (2007) Nutritional quality of vegetable and seed from different accessions of *Amaranthus* in South Africa. Water SA 33:377–380

Mosyakin SL, Robertson KR (1996) New infrageneric taxa and combinations in *Amaranthus* (Amaranthaceae). A Bot Fennici 33:275–281

Nikiforova SV, Cavalieri D, Velasco R, Goremykin V (2013) Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. Mol Biol Evol 30:1751–1760

Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ (2011) Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol J 9:328–333

Panero JL, Funk V (2008) The value of sampling anomalous taxa in phylogenetic studies: major clades of the Asteraceae revealed. Mol Phylogenet Evol 47:757–782

Parks M, Cronn R, Liston A (2009) Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol 7:84

Patil SM, Rane NR, Adsul AA, Gholave AR, Yadav SR, Jadhav JP, Govindwar SP (2016) Study of molecular genetic diversity and evolutionary history of medicinally important endangered genus Chlorophytum using DNA barcodes. Biochem Syst Ecol 65:245–252

Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465:1033–1038

Puigbò P, Garcia-Vallvé S, McInerney JO (2007) TOPD/FMTS: a new software to compare phylogenetic trees. Bioinformatics 23:1556–1558

Raju M, Varakumar S, Lakshminarayana R, Krishnakantha T, Baskaran V (2007) Carotenoid composition and vitamin A activity of medicinally important green leafy vegetables. Food Chem 101:1598–1605

Rastogi A, Shukla S (2013) Amaranth: a new millennium crop of nutraceutical values. Crit Rev Food Sci Nutr 53:109–125

Sangeetha RK, Baskaran V (2010) Carotenoid composition and retinol equivalent in plants of nutritional and medicinal importance: efficacy of β-carotene from *Chenopodium album* in retinol-deficient rats. Food Chem 119:1584–1590

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res 6:283–290

Sauer JD (1967) The grain amaranths and their relatives: a revised taxonomic and geographic survey. Ann Mo Bot Gard 54:103–137

Schmitz-Linneweber C, Maier RM, Alcaraz J-P, Cottet A, Herrmann RG, Mache R (2001) The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. Plant Mol Biol 45:307–315

Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB (2014) Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. AmJBot 101:1987–2004

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. The EMBO J 5:2043–2049

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539–545

Soltis PS, Soltis DE, Chase MW (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature 402:402–404

Srivastava R (2017) An updated review on phyto-pharmacological and pharmacognostical profile of *Amaranthus tricolor*: A herb of nutraceutical potentials. Pharma Innov J 6:127–129

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 22:2688–2690

Stetter MG, Schmid KJ (2017) Analysis of phylogenetic relationships and genome size evolution of the *Amaranthus* genus using GBS indicates the ancestors of an ancient crop. Mol Phylogenet Evol 109:80–92

Stetter MG, Müller T, Schmid KJ (2017) Genomic and phenotypic evidence for an incomplete domestication of South American grain amaranth (*Amaranthus caudatus*). Mol Ecol 26:871–886

Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am J Bot 99:349–364

Stull GW, Moore MJ, Mandala VS, Douglas NA, Kates HR, Qi X, Brockington SF, Soltis PS, Soltis DE, Gitzendanner MA (2013) A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. Appl Plant Sci 1:1200497

Sugiura M (1992) The chloroplast genome. In: Schilperoort RA, Dure L (eds) 10 Years plant molecular biology. Springer, Dordrecht, pp 149–168

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Timofte I, Timofte N, Brega V (2009) Development of bioenergy in Moldova. Problemele Energeticii Regionale 2:1–12

Van Rensburg WJ, Van Averbeke W, Slabbert R, Faber M, Van Jaarsveld P, Van Heerden I, Wenhold F, Oelofse A (2007) African leafy vegetables in South Africa. Water SA 33:317–326

Venskutonis PR, Kraujalis P (2013) Nutritional components of amaranth seeds and vegetables: a review on composition, properties, and uses. Comp Rev Food Sci Food Saf 12:381–412

Waselkov K (2013) Population Genetics and Phylogenetic Context of Weed Evolution in the Genus *Amaranthus*: Amaranthaceae. PhD Thesis, University of Washington

Wassom JJ, Tranel PJ (2005) Amplified Fragment Length Polymorphism-Based genetic relationships among weedy *Amaranthus* species. J Hered 96:410–416

Williams AV, Miller JT, Small I, Nevill PG, Boykin LM (2016) Integration of complete chloroplast genome sequences with small amplicon datasets improves phylogenetic resolution in Acacia. Mol Phylogenet Evol 96:1–8

Xu F, Sun M (2001) Comparative analysis of phylogenetic relationships of grain amaranths and their wild relatives (*Amaranthus*; Amaranthaceae) using internal transcribed spacer, amplified fragment length polymorphism, and double-primer fluorescent intersimple sequence repeat markers. Mol Phylogenet Evol 21:372–387

Yang J-B, Yang S-X, Li H-T, Yang J, Li D-Z (2013) Comparative chloroplast genomes of *Camellia* species. PLoS One 8:e73053

Zhang YJ, Ma PF, Li DZ (2011) High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). PLoS ONE 6:e20596

Zhang T, Fang Y, Wang X, Deng X, Zhang X, Hu S, Yu J (2012) The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes. PLoS ONE 7:e30531