BMC Genomics

RESEARCH Open Access

Analyses of contiguous reference genomes of *Amaranthus tuberculatus* highlight the landscape of the sex-associated region and PEBP gene family diversity

Damilola A. Raiyemo[1], Luan Cutti[2], Eric L. Patterson[2], Victor Llaca[3], Kevin Fengler[3], Jacob S. Montgomery[4], Sarah Morran[4], Todd A. Gaines[4] and Patrick J. Tranel[1*]

Abstract

Background *Amaranthus tuberculatus* (waterhemp) is a troublesome agronomic weed species that is dioecious with an XY sex-determination system. The evolution of the sex-determining region (SDR), the contiguity of the region, genomic landscape, and the expression pattern of genes within the region remain poorly understood.

Results We assembled a high-quality, chromosome-level nuclear genome and chloroplast and mitochondrial genomes of a male *A. tuberculatus*. Combining the genomes with restriction site-associated DNA genome-wide association (RAD-GWA) analysis, comparative genomics, adaptive evolution analysis, and transcriptomic profiling, we identified a ~ 31.8 Mb region on chromosome 1 that is strongly associated with sex. This region is gene-poor, abundant in long terminal repeat (LTR) retrotransposons, and harbors two inversions and a 3.19 Mb haplotype-specific region. Synteny analysis revealed that chromosome 1 likely originated from the fusion of two ancestral chromosomes, and mRNA analysis indicated 76 genes out of the 528 protein-coding genes within the putative SDR of Hap1 were differentially expressed between mature male and female flowers, with several of the genes enriched for Gene Ontology (GO) terms involved in floral development. We further characterized the phosphatidylethanolamine binding protein (PEBP) family in *A. tuberculatus* and related species to gain insights into *FLOWERING LOCUS T* diversity, as well as identified nuclear insertions of organellar origin in the species.

Conclusion Our results provide insight into the evolution of a sex-associated region in a weedy dioecious species, and the diversity of the PEBP protein family in amaranths. The genomic resources from this study will also be valuable for addressing further questions on adaptive trait evolution within the genus as well as questions surrounding dioecy in this and other plant species.

Keywords Chromosomal fusion, Dioecy, Retrotransposon, Sex-associated region, Waterhemp genome, Weedy amaranth

*Correspondence:
Patrick J. Tranel
tranel@illinois.edu
[1]Department of Crop Sciences, University of Illinois, Urbana, IL 61801, USA
[2]Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI, USA
[3]Genome Center of Excellence, Corteva Agriscience, Johnston, IA, USA
[4]Department of Agricultural Biology, Colorado State University, Fort Collins, CO, USA

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Background

Dioecy, the separation of male and female reproductive systems on different plants, has evolved multiple times independently across many lineages [1–3], and via several mechanisms [4–7]. One model of dioecy evolution postulates that dioecy could evolve via a gynodioecy pathway requiring two mutations [4, 8, 9] while the second model (one-gene model) postulates that dioecy could evolve via a single regulatory factor [7]. Both models have been supported by the discovery of either two [e.g., in *Actinidia* spp [10]. and *Asparagus officinalis* [11]] or one sex-determining gene(s) [e.g., in *Diospyros lotus* [12] and several *Populus* species [13]].

The sex-determining genes could either be contained within a small region (e.g., ~ 150 kbp in *Vitis* spp.) [14] or a larger non-recombining region that may have different strata (i.e., evolutionary histories) (e.g., 17.42 Mb in *Spinacia oleracea*) [15]. These regions typically accumulate repetitive sequences, possibly due to the low rates of crossing over [16–18]. The study of sex-determining regions (SDR) or the chromosomes containing the regions has been complicated due to assembly challenges posed by gene turnovers, structural variations, and repetitive sequences common within SDRs [19]. Nevertheless, the advances in sequencing technologies and overall improvements in computational approaches have made possible the sequencing and assembly of whole genomes and, more importantly, contiguous sex chromosomes, of non-model species such as weeds [20–24].

Amaranthus tuberculatus (Moq.) J.D. Sauer (waterhemp) is a troublesome dioecious weed of agronomic crops, native to the Midwest United States, but now with a global range [25, 26]. It is one of nine dioecious amaranths in the subgenus *Acnida* (L.) [25, 27, 28]. Previous studies investigating sex determination in *A. tuberculatus* confirmed males were the heterogametic sex [29, 30], but the sex chromosome pair were indistinguishable via microscopy [31]. Furthermore, a draft genome of *A. tuberculatus* was assembled into 841 contigs and, together with a *k*-mer-based approach, was used to identify male-specific Y contigs totaling ~ 4.6 Mb in length and containing 147 gene models [32, 33]. One of the male-specific Y contigs contained *FLOWERING LOCUS T* (*FT*), which has been hypothesized to contribute to male fitness or early flowering, given that males flower earlier than females in this species [32]. Analysis of short-read sequencing from three dioecious amaranths, *A. acanthochiton*, *A. cannabinus*, and *A. greggii*, within the same clade as *A. tuberculatus* also revealed the conservation of male-specificity of a copy of *FT* [34].

FT-like genes are diverse in their functions with roles mainly in the regulation of flowering time and morphogenesis [35, 36], and are members of the phosphatidylethanolamine binding protein (PEBP) family. The PEBP protein family consists of well-conserved proteins that have been reported across the three domains of life – plants, bacteria and archaea [37, 38]. The PEBP protein family is generally divided into three subfamilies in Arabidopsis: TFL-like, FT-like and MFT-like [39–42]. The TFL-like subfamily comprises three genes, *TERMINAL FLOWER1* (*TFL1*), *ARABIDOPSIS THALIANA CENTRORADIALIS* (*ATC*) AND *BROTHER OF FT AND TFL1* (*BFT*); the FT-like subfamily consists of two genes, *FLOWERING LOCUS T (FT*) and *TWIN SISTER OF FT* (*TSF*); and the MFT-like subfamily contains only one gene, *MOTHER OF FT AND TFL1* (*MFT*). A novel PEBP family member in plants, highly similar to the YY-PEBP (YbhB/YbcL family Raf kinase inhibitor-like protein) in prokaryotes and designated *STEPMOTHER OF FT AND TFL1* (*SMFT*), was recently reported [43]. While several studies have investigated genome-wide PEBP gene family diversity in multiple plant species and reported male-specific copies of *FT*-like genes in *Ficus hispida* [44], *Cannabis sativa* [45] and *Amaranthus* species [32, 34], little is known about the diversity of *FT* genes or the broader PEBP gene family in amaranths.

Sex-determining, or male-specific regions, are characterized by the accumulation of DNA fragments, repeats and pseudogenes, a phenomenon observed in liverwort, *Silene*, *Rumex*, *Coccinea*, and papaya [46–49]. Among these DNA fragments are organelle-derived nuclear insertions, termed NUMTs (nuclear mitochondrial DNA segments) and NUPTs (nuclear plastid DNA segments) [50–52]. The availability of organellar genomes enables the study of nuclear insertions originating from plastids or mitochondria, enhancing the understanding of organelle genome structure while aiding in the decoding of their function, inheritance and evolutionary history [53]. Using a combination of Oxford Nanopore Technologies (ONT) Simplex reads sequencing, Hi-C scaffolding, and Bionano optical mapping, we present haplotype-resolved, chromosome-level nuclear genomes alongside complete chloroplast and mitochondrial genomes of *A. tuberculatus*. This comprehensive dataset facilitates the characterization of the candidate sex-determining region in the genome, including its gene content and repeat structure. It also allows comparison of structural rearrangements between the two haplotype assemblies, provides an in-depth overview of PEBP gene family diversity, and identifies organelle-derived nuclear DNA fragments within the species.

Results

Assembly metrics and genome repetitive landscape

Two haplomes (Hap1 and Hap2) consisting of 16 chromosomes each were assembled for a single male individual, representing 90.42% (Hap1) and 86.66% (Hap2) (Table 1) of the estimated genome size of 675.6 Mb from

Table 1 Comparison of assembly statistics between the two haplomes of *A. tuberculatus* genome assemblies

Genome characteristics	Hap1	Hap2
Contig assembly size (Mbp)	610.86	585.50
Contig N50 (Mbp)	30.09	35.80
Scaffold assembly size (Mbp)	611.34	585.92
Scaffold N50 (Mbp)	37.57	38.80
Scaffold L50	7	7
GC content (%)	34.94	34.67
Assembly complete BUSCO (%)	97.5	97.6
Size of gaps (Mbp)	0.482	0.420
LTR assembly index (LAI)	17.98	19.28
Annotation characteristics	Hap1	Hap2
Protein-coding genes	28,610	27,824
Mean gene length (bp)	4,808	4,902
Mean CDS length (bp)	1,087	1,094
Mean exon length (bp)	293	292
Mean exon per gene	5.3	5.4
Number of tRNA	2,149	2,071
Total annotated genes	30,759	29,895

flow cytometry reported by Stetter and Schmid [54]. Analysis of assembly completeness for both Hap1 and Hap2 using BUSCO with embryophyte_odb10 database revealed 97.5% and 97.6% complete BUSCOs, respectively. LTR assembly index (LAI) also revealed high quality assemblies with average LAI scores for Hap1 and Hap2 at 17.98 and 19.28, respectively (Table 1). Repeat analysis using RepeatMasker revealed that 72.01% of Hap1 and 70.88% of Hap2 were made up of repetitive elements. The LTR/*Ty3* elements were the most abundant retrotransposons, representing 13.3% and 11.99% of the Hap1 and Hap2 assemblies, respectively (Table S1). Analysis of high-copy tandem repeats using StainedGlass heatmaps revealed that chromosomes 1–7 appear to be submetacentric, whereas chromosomes 8–16 appear to be telocentric in both haplomes (Fig. S1, Fig. S2). BLAST search of the simple telomeric repeat, TTTAG GG against both haplotypes revealed telomeric repeat sequences at 30 out of the possible 32 telomeric ends for both haplotypes (Table S2), which is comparable to 30 out of 34 telomeric ends reported for *A. tricolor* [55]. Further annotation of the genomes revealed Hap1 and Hap2 had 1,434 and 1,502 genes annotated as transcription factors (TF), respectively (Table S3 and Table S4). Visualization of the analyzed genomic features indicates an inverse relationship between gene density and LTR proportions, whereby gene-rich regions are LTR-poor and LTR-rich regions are gene-poor (Fig. 1).

Identification of the candidate sex-determining region

Of the 353 individuals genotyped in Montgomery et al. [30], we retained 43 females and 45 males from the population after excluding individuals with more than 60% missing data. Variant calling on these 88 individuals yielded 193,455 variants. Genome-wide association (GWA) analysis with these individuals and variants identified single-nucleotide polymorphisms (SNPs) significantly associated with sex above the Bonferroni threshold ($-\log_{10}P$ = 6.5876). These SNPs were located on chromosome 1, spanning from 14.02 to 45.81 Mb (Fig. 2a). Additional complementary GWA analysis using BLINK revealed a single SNP at position 16,139,826 on chromosome 1 as significantly associated with sex. However, the SNP resides within an intergenic region. The SNP was also the 7th most significant SNP from the first analysis with EMMAX. QQ plots comparing the two GWA approaches revealed underlying population structure, which was accounted for by the different models (Fig. 2b). Genetic differentiation (F_{ST}) along chromosome 1 between females and males exceeding the top 5% threshold spanned from 15.1 to 44.6 Mb, with a second peak at 52.6–52.7 Mb (Fig. 2a).

A BLAST query of previously reported primer sets, used to amplify male-specific regions [32], against both haplotypes also revealed perfect matches to chromosome 1 on Hap1, but none on Hap2. One set of primers (WHMS) matched to a 572 bp region (33,273,720–33,274,292 bp) (Fig. 2c) while primer sets MU-976, MU-657.2, and MU-533 also matched to regions within the candidate SDR on Hap1 (Table S5). The 572 bp male-specific marker had homology to an LTR/*Ty3* retrotransposon (Fig. S3). Similarly, MU-657.2 was previously reported to have homology to a LTR/*Ty3* element in sugar beet [30], indicating the abundance of the *Ty3* superfamily within the SDR. A BLAST search of a 200 bp *FT* gene, reported to be male-specific across three dioecious amaranths related to *A. tuberculatus* [34], against both haplotype assemblies also revealed a perfect match to the fourth exon (15,545,318–15,545,518 bp) of a full-length, four-exon gene, AmaTuChr01Ag009380 on Hap1. We therefore consider chromosome 1 of Hap1 the Y and chromosome 1 of Hap2 the X (Fig. 2b).

Reciprocal best hit (RBH) search between 147 genes from previously identified male-specific contigs in a draft *A. tuberculatus* genome [32] and the haplotype assemblies from this study revealed 18 and 12 genes had best matches on chromosome 1 of Hap1 and Hap2 assemblies, respectively (Table S6). The genes are thus conserved between the draft assembly and the new assembly. Taken together, the line of evidence above supports chromosome 1 as the likely sex chromosome, and we defined an approximate boundary of the candidate SDR as a region of chromosome 1 on Hap1 between 14.02 and 45.81 Mb (~ 31.79 Mb). There are 2,614 protein-coding genes on chromosome 1 of Hap1, of which 796 are within the candidate SDR.

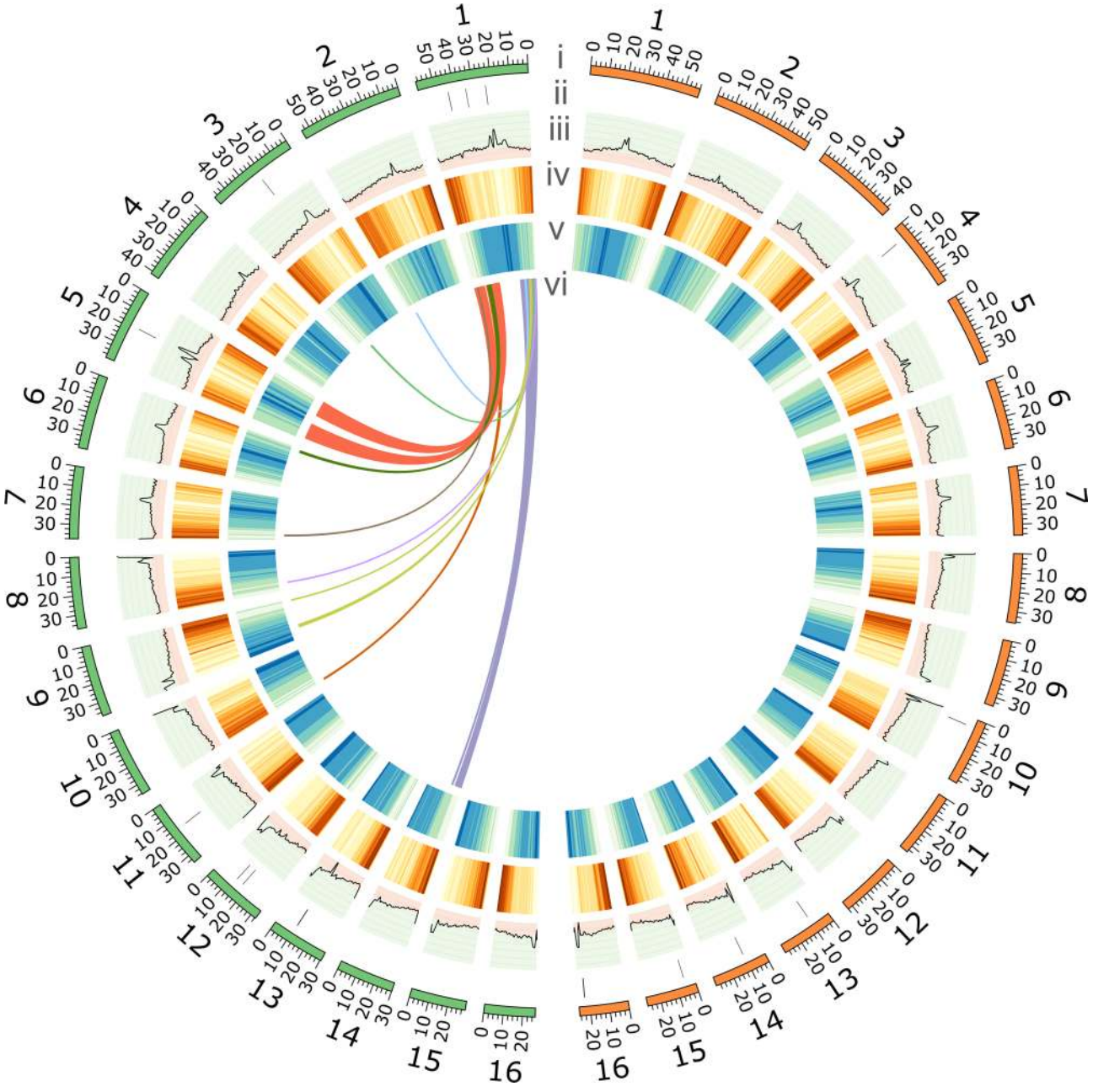


Fig. 1 Genomic features of *A. tuberculatus* Hap1 (left) and Hap2 (right) assemblies. Circos plot depicts (i) number and length (Mb) of chromosomes, (ii) distribution of gaps in the genomes, (iii) GC content along the chromosomes, with peaks in light green area representing GC content greater than the median (0.3442) and peaks in light red area representing GC content less than the median, (iv) gene density across the chromosomes, with brown representing gene-rich regions and yellow representing gene-poor regions, (v) LTR (long terminal repeats) density along chromosomes, with blue representing LTR-rich regions and green representing LTR-poor regions, (vi) inner ribbons represent collinear gene blocks on chromosome 1 of Hap1. Collinear genes on chromosome 1 of Hap2 are not shown in the figure to avoid redundancy. Window size of 1 Mb and step size of 500 kbp for ii-iv

Comparative analysis between the candidate SDR of the two haplomes

Synteny patterns between the two haplomes indicate a 1:1 relationship in gene content (Fig. S4). Out of the 20,763 gene pairs representing the reciprocal best matches between the two haplotypes, there were 1,822 pairs for chromosome 1, which were used to define the boundaries of regions within the candidate SDR (Collinear 1: 663, INV 1: 131, Collinear 2: 21, INV 2: 94, Collinear 3: 913). The synteny analysis revealed two inversions on either side of a collinear region (Collinear 2) within the candidate SDR (Fig. 2c). One inversion (INV 1) is from 17,907,092–33,155,961 bp on Hap1, and from 15,719,009–31,029,795 bp on Hap2, based on

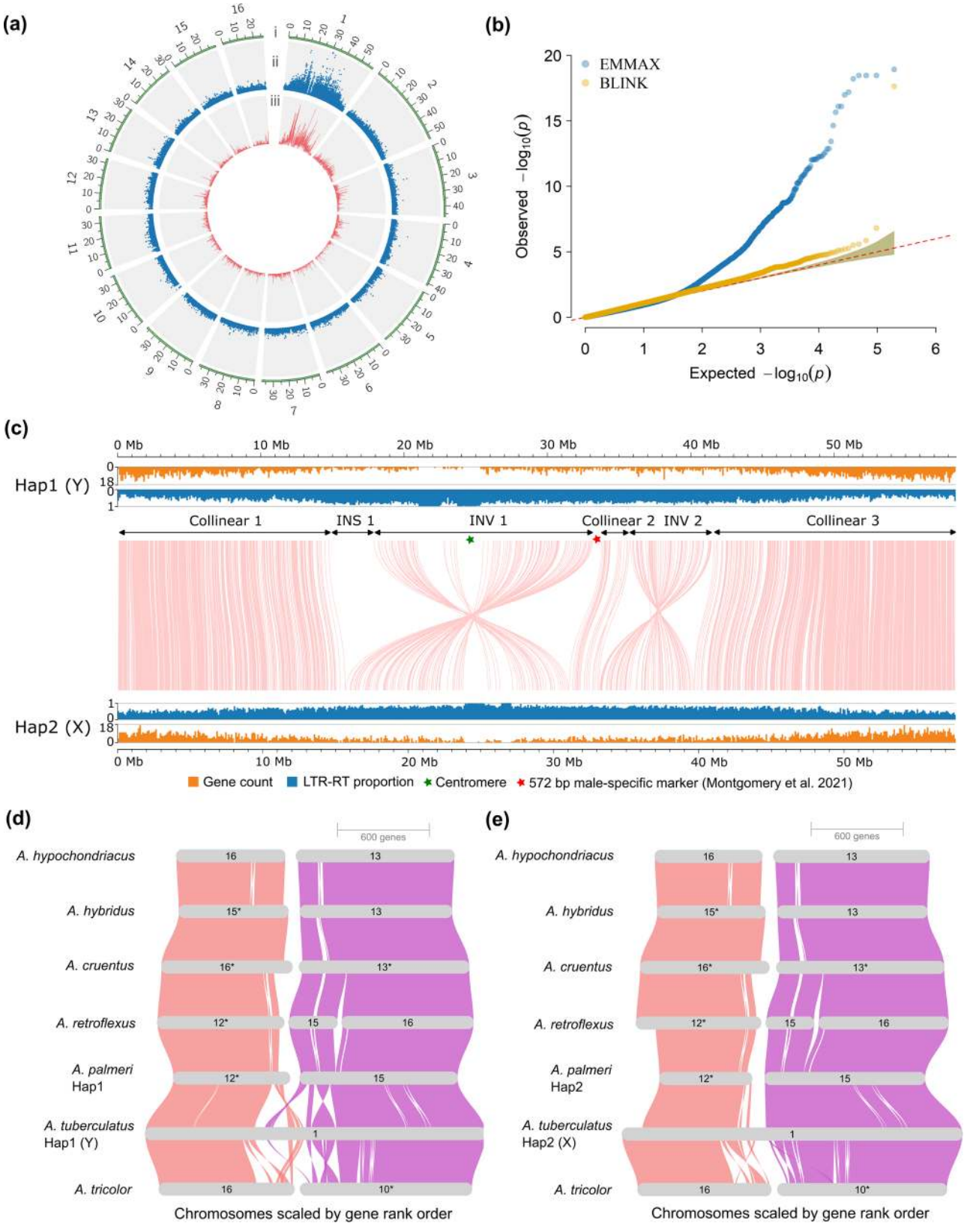


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Identification of the sex-associated region on chromosome 1. **a** Circos plot represents (i) number and length (Mb) of chromosomes. (ii) SNPs from GWA analysis using RAD-seq data from 43 females and 45 males. (iii) Fixation index (F_{ST}) between females and males across all 16 chromosomes (window size 100 kbp; step size 50 kbp). **b** Quantile-quantile (QQ) plot of two complementary GWA analyses. **c** Synteny plot based on gene order, showing collinear regions, an insertion (INS 1), and two inversions (INV) on chromosome 1. The red asterisk represents a previously reported 572 bp male-specific marker [32] that matched to a region on Hap1. Gene count and LTR-RT proportion were calculated based on 100 kbp non-overlapping windows. **d**, **e** Synteny between the haplotype assemblies of *A. tuberculatus* and chromosome-level assemblies of other *Amaranthus* species highlighting likely fusion of two separate chromosomes in *A. tuberculatus*. Asterisks indicate chromosomes that were manually inverted to keep the gene order consistent with *A. tuberculatus*

the first and last gene pairs for the inversion. The other inversion (INV 2) ranges from 35,809,637–41,407,537 bp on Hap1, and from 33,229,536–39,616,194 bp on Hap2 (Fig. 2c). Further analysis of structural rearrangements using SyRI revealed 274 inversions spread across the 16 chromosomes between Hap1 and Hap2 (Table S7); however, INV 1 and INV 2 within the SDR were the two largest inversions out of the 274 inversions identified using SyRI (Table S7). Comparison of chromosome 1 across *A. tuberculatus* and six other *Amaranthus* species indicates a conserved gene order between the species and Hap2; however, it appears the inversions likely occurred on Hap1 (Fig. 2d and e). Also, chromosome 1 appears to have originated from the fusion of two ancestral chromosomes (Fig. 2d and e).

Analysis of the inversions revealed INV 1 on Hap1 and Hap2 contain 326 and 313 protein-coding genes, respectively, although only 130 genes are syntenic between the haplotypes. INV 2 on Hap1 and Hap2 contain 191 and 215 protein-coding genes, respectively, with only 93 syntenic genes between the haplotypes. A region upstream of Hap1 between 14,660,708–17,851,789 bp (3.19 Mb) was found to lack synteny to any region of chromosome 1 on Hap2, and no homologous collinear block of sequences elsewhere in the genome. This region was designated INS 1 and contains 58 protein-coding genes (Fig. 2e; Table S8). However, only 5 of the genes (encoding EG45-like domain containing protein 1, FT/HEADING DATE 3 A, FLOURY 1, polyadenylate-binding protein 5, and binding partner of ACD11 1) were annotated while the rest were uncharacterized or unknown. Additional analysis of the 58 genes within INS 1 revealed 35 genes had homologous copies across the 16 chromosomes with greater than 70% similarity at the protein sequence level (Table S9). Although the *FT/HEADING DATE 3A* and its homologs vary in their intron length, they appear to have a conserved gene structure (Fig. S5). Phylogenetic analysis of the encoded protein sequences of the *FT/HEADING DATE 3A* and its homologs indicate that the copy (AmaTuChr01Ag009380) within the male-specific Y region (MSY) of *A. tuberculatus* Hap1 genome is more related to the copy (AmaPaChr00Ag260700) in unplaced chromosome (Chr00) of the *A. palmeri* genome rather than the copy (AmaPaChr03Ag064170) within the MSY of *A. palmeri* (Fig. S5). Gene density for the regions (collinear 1, INS 1, INV 1, collinear 2, INV 2, and collinear 3) on chromosome 1 of Hap1 and Hap2 were statistically different with $p < 6.056e\text{-}14$ and $p < 1.133e\text{-}13$, respectively (Fig. 3a). Similarly, intact LTR proportions for the regions were statistically different for both haplotypes (Hap1; $p < 2.341e\text{-}16$ and Hap2; $p < 2.85e\text{-}15$) (Fig. 3b). Pairwise comparison indicated that collinear 1 and 3 are gene-rich but LTR-poor while the regions (INS 1, INV 1, collinear 2, and INV 2) are LTR-rich but gene-poor for Hap1, which was same for Hap2 (excluding the insertion which was not present in the haplotype). Although analysis of insertion times revealed slight variation in the timing of intact long terminal repeat retrotransposons (LTR-RTs) insertions for collinear 2 and INV 2 compared with the other regions, both haplotypes appear to have accumulated several intact LTR-RTs in these regions within the last 0.5 MYA, as indicated by the peaks around this time (Fig. 3c and d).

Sequence divergence and detection of genes under positive selection

Analysis of adaptive evolution using the M1a vs. M2a model revealed 24 genes out of 608 single-copy genes on chromosome 1 were significant ($\alpha = 0.05$) for sites under positive selection (Table S10). Using the branch model, where Hap1 was used as the foreground branch to determine if ω differs for this branch relative to the background branches (i.e., the M0 vs. two-ratio model), 5 genes, including a serine/arginine-rich SC35-like splicing factor 33 (*SCL33*) had significantly different ω, indicating it confers some fitness benefit. When Hap2 was used as the foreground branch, only three genes had significantly different ω. When both haplomes were specified as the foreground branch [i.e., (Hap1 #1, Hap2 #1)], eight genes had ω that differed for the specified branch. Similarly, when the branch leading to the common ancestor of the two haplomes were included in the foreground branch (i.e., (Hap1 #1, Hap2 #1) #1)], eight genes, including ones encoding bark storage protein B and polygalacturonase, had ω that differed for the foreground branch compared to the background branches.

Adopting the foreground and background branches used for the branch model above but utilizing a branch-site model (i.e., MA_{null}, $\omega = 1$ vs. MA, $\omega > 1$), indicated 24 genes, 35 genes, 50 genes, and 46 genes were more likely to contain sites with $\omega > 1$ when Hap1, Hap2, both haplotypes, the branch leading to the common ancestor, and the branch leading to the haplomes was used as the foreground branch, respectively (Fig. S5, Table S10, and

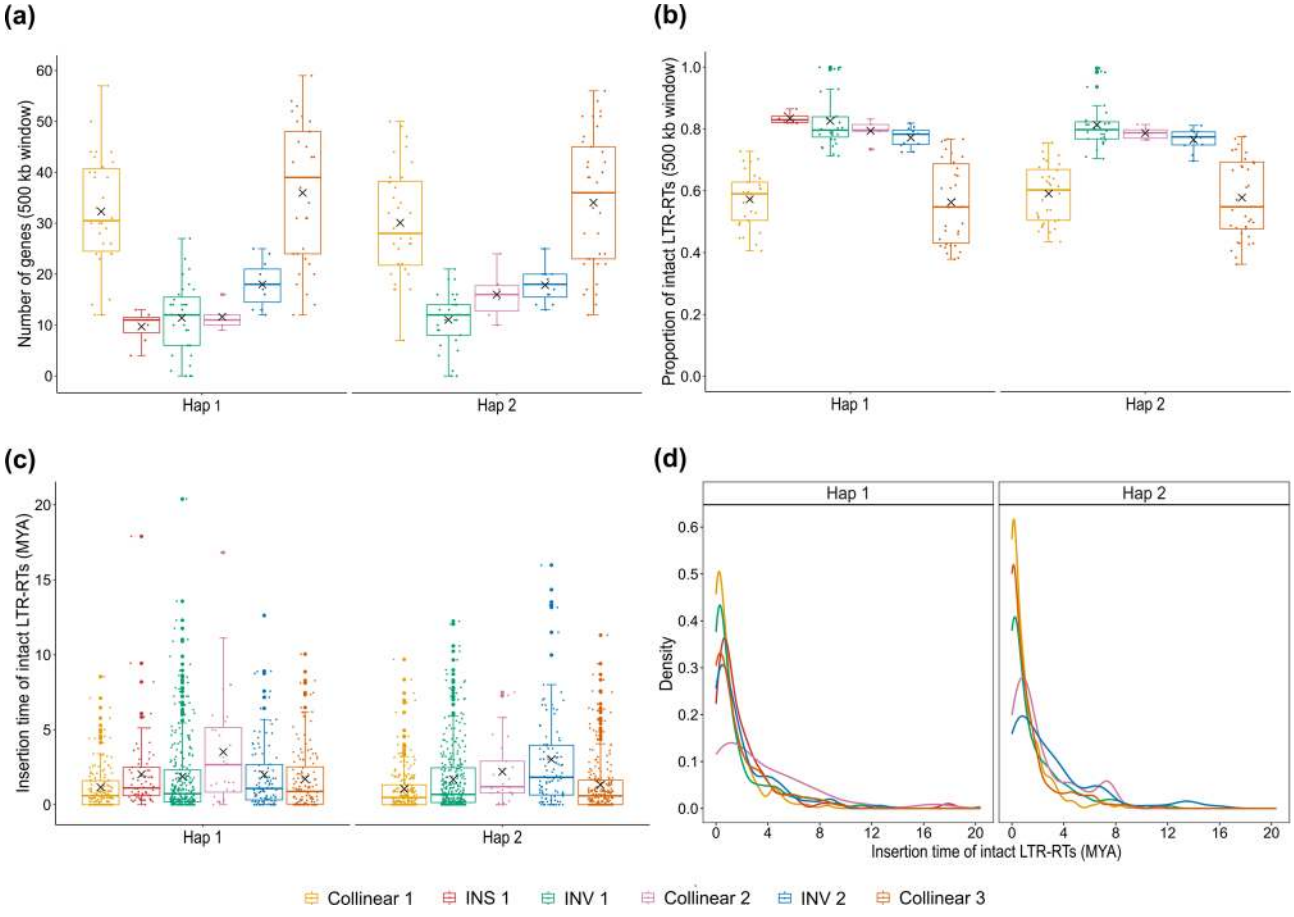


Fig. 3 Comparative analysis between chromosome 1 of the two haplomes. **a** Number of genes across five regions (three collinear regions and two inversions) plus an insertion in Hap 1 on the chromosome. Gene densities are calculated per 500 kbp non-overlapping windows. **b** Proportion of intact LTR-RTs across the five regions on the chromosome. LTR-RT proportions are calculated per 500 kbp non-overlapping windows. **c** Insertion time of intact LTR-RTs across the five regions. **d** Density distribution of intact LTR-RTs insertion times across the five or six regions on the chromosome

Table S11). Within the candidate SDR, genes encoding E3 SUMO-protein ligase SIZ1-like, arabinogalactan peptide 16, and several unknown proteins were more likely to contain sites under positive selection based on the branch-site model if Hap1 was the foreground branch (Fig. S6, Table S10, and Table S11). Taken together, our analysis above reveals genes that are potentially important in sex-specific adaptation.

Expression analysis identifies genes involved in floral development

Cleaned mRNA reads (i.e., adapter-trimmed and low-quality bases removed) from three tissue types (shoot apical meristem, floral meristem, and mature flower) reported by Bobadilla et al., [56] that were mapped to the Hap1 assembly had uniquely mapped reads for males ranging from 84.01 to 89.45%, and uniquely mapped reads for females ranging from 74.82 to 90.54%. Out of the 28,610 annotated Hap1 protein-coding genes, 21,530 genes were retained for differential expression analysis after filtering and TMM normalization (Fig. 4a). Among the retained genes, 5,153 were differentially expressed between male and female individuals at the mature flower stage, representing the majority of differentially expressed genes across all sampled tissue stages in the study (Fig. 4b). Among the 1,904 genes retained on chromosome 1, two at the shoot apical meristem stage, four at the floral meristem stage, and 446 (244 upregulated and 202 downregulated) at the mature flower stage were differentially expressed between males and females (Fig. 4c; Table S12 – S14). Within the candidate SDR on chromosome 1, two genes at the shoot apical meristem stage, two genes at the floral meristem stage, and 76 genes (44 upregulated and 32 downregulated) at the mature flower stage were differentially expressed (Table S12 – S14). Two genes within the candidate SDR (specifically within INV 1) encoding MADS-box transcription factor 23 (AmaTuChr01Ag011280) and LOB domain-containing protein 19-like (AmaTuChr01Ag011320) were consistently downregulated across the three tissue types for male plants (Table S12 – S14). Gene ontology (GO) term enrichment analysis was performed to gain

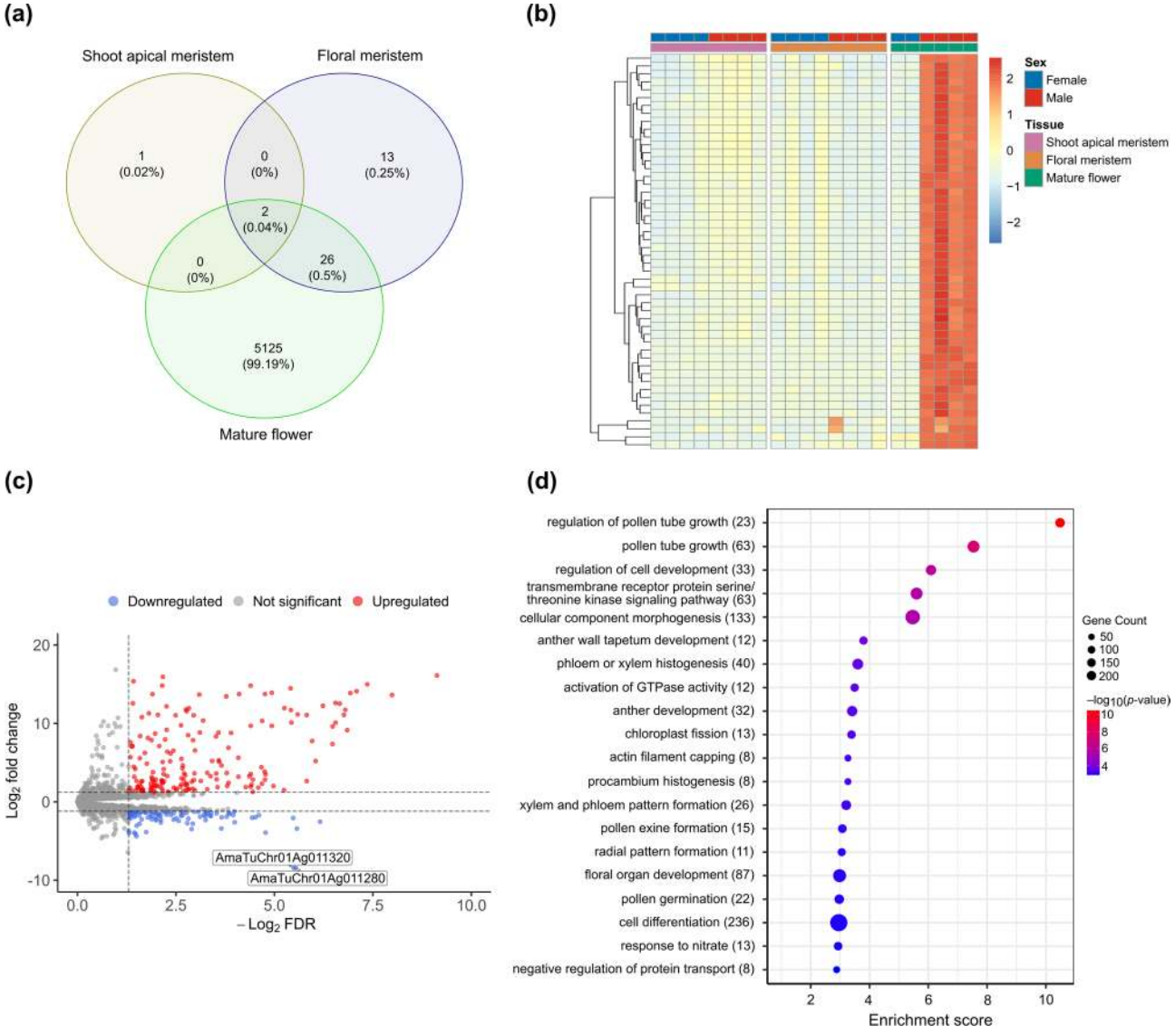


Fig. 4 Differential gene expression analysis between male and female individuals across three tissue types. **a** Numbers of differentially expressed genes for male versus female comparison for shoot apical meristem, floral meristem, and mature flower. **b** Heatmap of rlog (regularized log) transformed raw counts of the 5,153 differentially expressed genes in male versus female comparison for mature flower. Samples are ordered using the hierarchical clustering method. Brown depicts genes with relatively high expression, yellow depicts intermediate expression levels and blue represents genes with relatively low expression levels. **c** Volcano plot of differentially expressed genes on chromosome 1 for mature flower. Highlighted genes were consistently downregulated among the three tissue types highlighted. Dotted lines represent FDR < 0.05 and FC > 1.2 thresholds. **d** Gene ontology enrichment analysis showing significantly overrepresented terms for male versus female differentially expressed genes for mature flower. Values in parentheses represent the number of genes within the term

insight into biological processes that could be involved in sex determination. Biological processes including pollen tube growth, pollen germination, pollen exine formation, regulation of cell development, cellular component morphogenesis, and anther wall tapetum development were identified among the top 20 enriched GO terms (Fig. 4d, Table S15 – S16). The top five terms enriched for molecular function included transmembrane receptor protein serine/threonine kinase activity, GTPase activator activity, small GTPase binding, pectinesterase activity and mannan synthase activity (Table S17), while the top five terms enriched for cellular function included plasma membrane, apical plasma membrane, pollen tube, pollen tube tip, and transport vesicle (Table S18).

PEBP gene family diversity

The identification of a male-specific *FT* copy in *A. tuberculatus*, along with the conservation of the *FT* in three related species and the hypothesis of its evolutionary benefits in male fitness, highlights the need for a

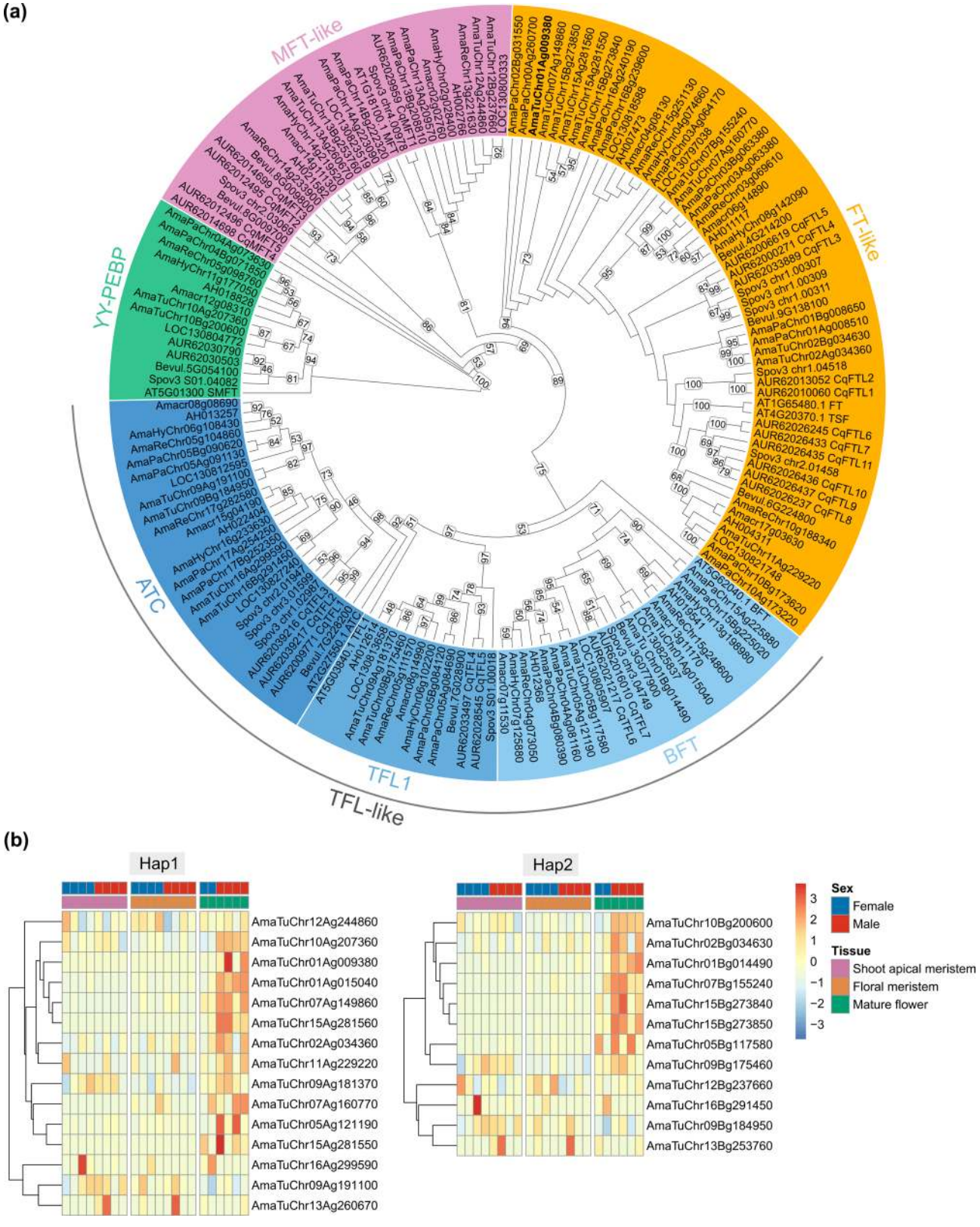


Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 PEBP protein family diversity and corresponding expression patterns in *A. tuberculatus*. **a** Phylogenetic tree of PEBP proteins from *Amaranthus* species, *Chenopodium quinoa*, *Spinacia oleracea*, *Beta vulgaris* and *Arabidopsis thaliana*. Numbers above branches represent RAxML bootstrap support (BS) values. Only BS greater than 45% are shown on tree. Subfamilies are indicated as TFL-like (TERMINAL FLOWER1-like; includes *ATC* [*ARABIDOPSIS THALIANA CENTRORADIALIS*], *TFL1* [*TERMINAL FLOWER1*], and *BFT* [*BROTHER OF FT AND TFL1*]), MFT-like (MOTHER OF FT AND TFL1-like), FT-like (FLOWERING LOCUS T-like), and YY-PEBP (YbhB/YbcL family Raf kinase inhibitor-like protein). The male-specific FT-like copy on Hap1 is shown in bold. **b** Heatmap indicating expression patterns of *PEBP* genes in *A. tuberculatus* Hap1 and Hap2 across three tissue types

comprehensive exploratory analysis of *FT* genes and the PEBP gene family in amaranths. BLAST and HMMsearch analyses of PEBP genes in the amaranths revealed several candidates; however, only genes with PEBP conserved domains were retained for downstream analysis. We identified 9 (*B. vulgaris*), 10 (*A. hybridus*), 11 (*A. hypochondriacus*, *A. cruentus*, *A. retroflexus*, and *A. tricolor*), 12 (*A. tuberculatus* Hap2), 13 (*A. palmeri* Hap2), 14 (*A. palmeri* Hap2), and 15 (*A. tuberculatus* Hap1) PEBP genes across the species. The genes were further classified into subfamilies based on the Arabidopsis-derived classification system, namely MFT-like, TFL-like, and FT-like (Fig. 5a). The *MFT* gene had two copies in *Amaranthus* species, *B. vulgaris* and *S. oleracea*, compared with a single copy in *A. thaliana* and five copies in *C. quinoa*. Our analysis also distinguished clearly between *TFL1*, *ATC*, and *BFT* genes (members of the TFL-like subfamily) across the species (Fig. 5a). All the species had a single copy of *TFL1*, compared to two copies in *C. quinoa*. The *Amaranthus* species had two copies, each of *ATC* and *BFT*, while *B. vulgaris* had a single copy of each. The FT-like subfamily was more diverse than the other two subfamilies. There was also no clear distinction between the *FT* and *TSF* genes (both members of the FT-like subfamily). Our analysis revealed two copies of FT-like genes in *A. hybridus*; three copies in *A. hypochondriacus*, *A. cruentus*, *A. tricolor*, and *A. retroflexus*; four copies in *B. vulgaris*; five copies in *A. palmeri* Hap2; and six copies in *A. palmeri* Hap1, compared with five copies in *S. oleracea* and eleven copies in *C. quinoa*. In Hap1 of *A. tuberculatus*, there were seven copies of FT-like genes while Hap2 had four copies of the FT-like genes. Both AmaTuChr01Ag009380 and AmaTuChr01Ag015040 are within the candidate SDR on chromosome 1 of Hap1; however, only AmaTuChr01Ag009380 was Hap1-specific and previously reported to be male-specific while AmaTuChr01Ag015040 is homologous to AmaTuChr01Bg014490 in Hap2. Among the PEBP genes in *A. tuberculatus*, only two (AmaTuChr10Ag207360 in Hap1 and AmaTuChr10Bg200600 in Hap2; annotated as uncharacterized protein, and AmaTuChr01Ag015040 in Hap1 and AmaTuChr01Bg014490 in Hap2; annotated as CEN-like protein 1) were upregulated in mature male flowers relative to female flowers (Fig. 5b). Based on the NCBI Conserved Domain Database (CDD), the uncharacterized protein contains a domain homologous to the PEBP domain found in bacteria and archaea, and is classified as a YbhB/YbcL family Raf kinase inhibitor-like protein, similar to *E. coli* YbhB and YbcL. Genes from other amaranths clustering with AmaTuChr10Ag207360 also contain a domain homologous to a PEBP domain in bacteria and archaea; we therefore designate this clade YY-PEBP in our phylogenetic tree (Fig. 5a).

Organellar genome features and their homologous sequences in the nuclear genome

Due to several recent reports of chloroplast genomes for *A. tuberculatus* [57, 58], we focused on mitochondrial genome features of the species but further explored insertions from both organelles into the nuclear genome, including within the candidate SDR. The mitochondrial genome of *A. tuberculatus* was assembled to a length of 358,073 bp, and contained 29 protein-coding, 25 transfer RNA (tRNA) and 3 ribosomal RNA (rRNA) genes (Fig. 6a; Table S19). Sixty-four simple sequence repeats were identified in the mitochondrial genome, in which tetranucleotide (37) were most abundant, followed by trinucleotide (11), mononucleotide (9), dinucleotide and pentanucleotide (3), and hexanucleotide (1). Additional repeat analysis using REPuter revealed the mitochondrial genome had 31 forward repeat sequences and 69 palindromic repeat sequences. Comparison of the *A. tuberculatus* mitochondrial genomes to mitochondrial genomes from *C. quinoa*, *S. oleracea*, and *B. vulgaris* indicates a high degree of rearrangements and low conservation among the species (Fig. 6b).

Analyses of mitochondrial DNA insertions revealed 14 mitochondrial genes (*atp1*, *atp8*, *ccmFC*, *ccmFn*, *cob*, *cox1*, *cox3*, *nad3*, *nad4*, *nad5*, *nad9*, *rpl5*, *rps3*, *rps4*) had copies distributed across 9 chromosomes in Hap1 (Table S20A). Similarly, 14 mitochondrial genes (*atp1*, *atp8*, *ccmC*, *ccmFC*, *ccmFn*, *cox1*, *nad2*, *nad4*, *nad5*, *nad9*, *rpl5*, *rps12*, *rps3*, *tatC*) had copies distributed across 12 chromosomes in Hap2 (Table S20B). Analyses of chloroplast DNA insertions revealed Hap1 and Hap2 had 59 and 61 chloroplast genes, respectively, distributed across the 16 chromosomes (Tables S21 and S22). Among the chloroplast genes integrated into the nuclear genomes, *ycf2*, *ycf1*, and *rpoC2* were most abundant with greater than 20 copies (Tables S21 and S22). A search of the chloroplast genes within the mitochondrial genome revealed 24 genes (*accD*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*, *cemA*, *ndhB*, *pafII*, *petA*, *petG*, *psaI*, *psbA*, *psbC*, *psbE*, *psbF*, *psbL*, *rbcL*, *rpl2*, *rpoC1*, *rpoC2*, *rps2*, *ycf1*, and *ycf2*) were found

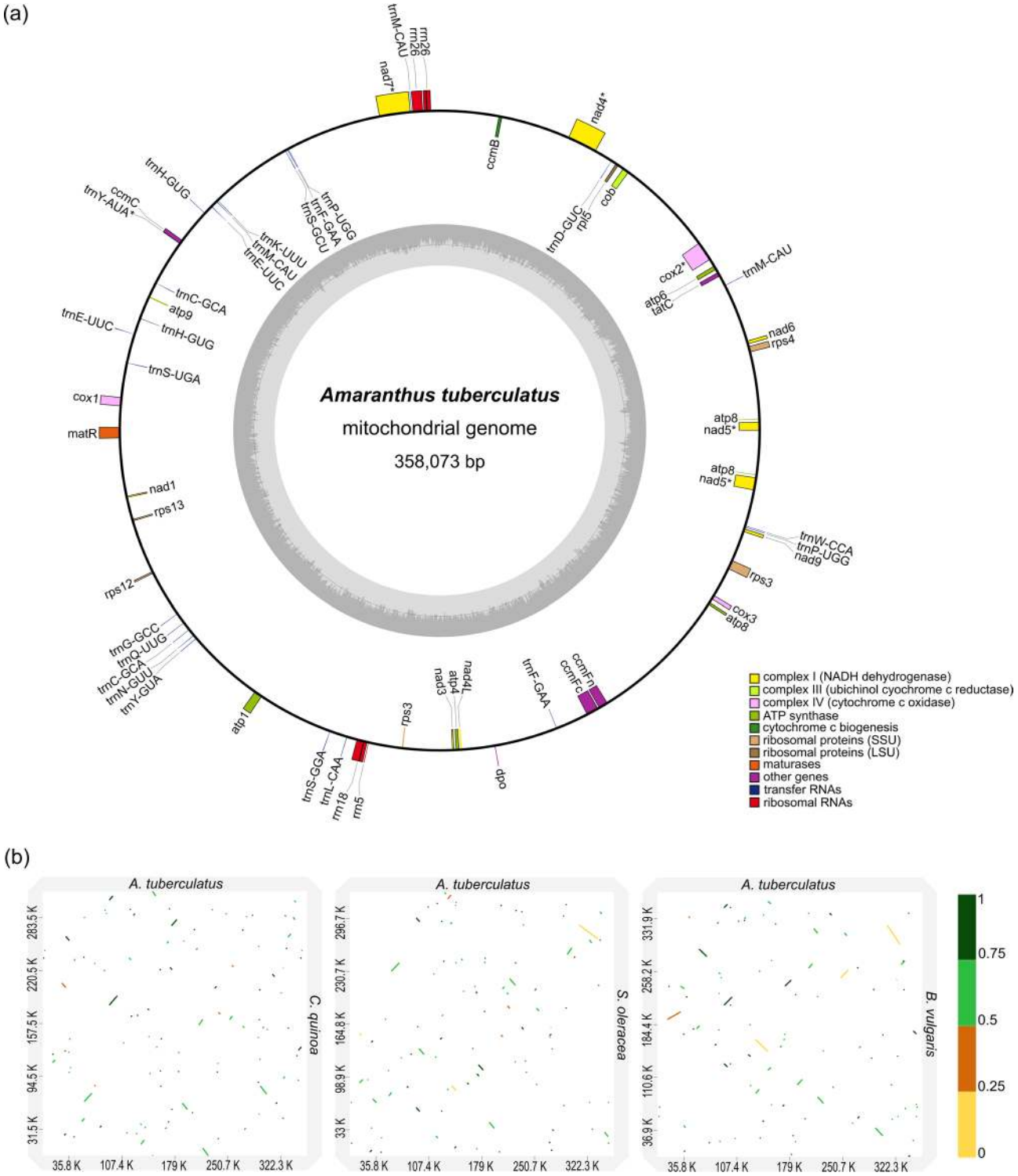


Fig. 6 Mitochondrial genome of *Amaranthus tuberculatus* and comparisons with related species. **a** Circular structure of the *A. tuberculatus* mitochondrial genome. Genes shown on the outside of the map are transcribed clockwise while genes shown on the inside are transcribed counterclockwise. The dark grey region within the circle represents GC content. **b** Dotplot visualization of alignment of the *A. tuberculatus* mitochondrial genome to related species. The scale bar represents percentage identity

within the mitochondrial genome; however, none of the mitochondrial genes had a sequence match in the chloroplast genome (Table S23). Analysis of larger blocks of DNA fragments revealed 29 fragments with a total length of 33.38 kbp from the chloroplast genome had homologs in the mitochondrial genome (9.32% of the mitochondrial genome) (Table S24). Several fragments (>400) from both the mitochondrial and chloroplast genomes were found to have integrated into the nuclear genome with longer stretches in some cases (e.g., an 18.25 kbp chloroplast fragment integrated into chromosome 2 of Hap1, an 11.14 kbp chloroplast fragment integrated into chromosome 6 of Hap2, a 9 kbp mitochondrial fragment integrated into chromosome 3 of Hap1, and a 42.56 kbp mitochondrial fragment integrated into chromosome 13 of Hap2) (Supplementary Table S25 – S28). Within the candidate SDR (14.02–45.81 Mbp) in Hap1, about 5,058 bp of mitochondrial sequences (including *atp1*, *nad3*, and *rps3* gene fragments totaling 1,150 bp) were inserted into the region. About 3,188 bp of chloroplast sequences (including *atpI*, *ndhB*, *ndhH*, *psbC*, and *ycf2* totaling 1,038 bp) were also inserted into the SDR on Hap1 (Supplementary Table S25 – S28). No mitochondrial or chloroplast sequence fragments however had a match to any sequence within the region (14.66–17.85 Mb) designated as male-specific (INS 1) on Hap1. Similarly, within the region on Hap2 (~14.6–43.00 Mb) equivalent to the candidate SDR on Hap1, about 6,731 bp of mitochondrial sequences (including the *rps3* gene fragment totaling 816 bp) were inserted into the region. Also, about 7,383 bp of chloroplast sequences (including *accD*, *matK*, *ndhF*, *psaB*, *ycf1*, and *ycf2* totaling 5,526 bp) were inserted into the region (Supplementary Table S25 – S28).

Discussion

We present a haplotype-resolved nuclear genome and chloroplast and mitochondrial genome assemblies of *A. tuberculatus* with high contiguity. We comprehensively demonstrate that chromosome 1 of the assembled genome contains the candidate region that is strongly associated with sex, harboring the two largest inversions in the genome within a ~ 31.8 Mb region that is gene-poor but abundant in LTR retrotransposons. The regions flanking this candidate sex-determining region are gene-rich and LTR-poor, indicating the possibility that they are still actively recombining between the Y and the X. This observed SDR landscape is also corroborated by evidence from *Spinacea oleracea*, where two inversions within a 17.42 Mb are flanked by pseudoautosomal regions (PAR) that are syntenic between the Y and X [15]. The concentration of high-order repeats, typical of centromeric regions [59], in the middle of chromosome 1 indicates the chromosome is either submetacentric or metacentric, with the first inversion (INV 1) located in the peri- or centromeric region of the chromosome. Inversions have been observed within sex chromosomes of different species, including plants and animals [15, 47, 60, 61], and are known evolutionary drivers of sex determination systems [60]. Chromosome 1 also appears to have originated from the fusion of two ancestral chromosomes following the divergence of the *Acnidia* subgenus from *Albersia*. This fusion (referred to as Robertsonian translocation) was independently observed by Kreiner et al. [62] in their study of sex evolution in *A. tuberculatus*, in which they also identified chromosome 1 as the sex chromosome. In that study, the authors report a smaller (~ 3 Mb) region on the Y chromosome that is strongly but incompletely associated with maleness. While Robertsonian translocation typically involves the fusion of two telocentric or acrocentric chromosomes [63], we note that in *A. tricolor*, chromosome 10 appears to be telocentric [55], and therefore suitable for fusion; however, chromosome 16 appears to be metacentric or submetacentric [55], and initial shortening of the chromosome arm or a complex mechanism would be required for the fusion of both chromosomes. Interestingly, chromosome 1 of *A. tuberculatus* shows two regions with high-order repeats and thus appears dicentric; however, whether both centromeres are active, functional or stable would require further investigation. It is worth noting that Kreiner et al. [62] reported varying SDR architectures among several populations, with an inversion segregating between X and Y across three haplotypes and thus reflects population-level diversity of the SDR in *A. tuberculatus*.

In multiple species, insertions, notably TE accumulation, have contributed to the expansion of sex-determining regions [16, 64–66]. Although the haplotype-specific insertion (INS 1) upstream of the first inversion has accumulated retrotransposons, the region still contained fifty-eight genes, including *FT*/*HEADING DATE 3A*, which we previously reported as male-specific in this species [34]. Given that 50% of the genes have homologous copies elsewhere in the genome, and none of the genes are expressed across three tissue types, it is possible that the genes are not functional or perhaps their expression was not captured at the stage during which tissues were sampled. Furthermore, only two genes within the sex-determining region on chromosome 1, one encoding a MADS-box transcription factor 23 and the other, a LOB domain-containing protein 19-like were down-regulated in males across the three tissue types. Further comparison between this work and Bobadilla et al. [56] who investigated the expression of genes between male and female individuals using a draft *A. tuberculatus* contig assembly indicates high congruency in the expression patterns observed in both studies. While the authors reported no differentially expressed genes within the

putative male-specific Y contigs from the draft assembly, they found autosomal genes, including MADS-box transcription factor 18 (*MADS18*), LOB domain-containing protein 31 (*LOB31*), *SOC1*, and *JAGGED* that were downregulated in males. These genes were also downregulated in our reanalysis of their transcriptome data; however, the genes are now mapped to their respective chromosomal locations. Similar to their finding, several genes, including *PMADS2*, *CMB2*, *bHLH91*, and *CYP120A11*, reported as upregulated in males were also upregulated in males in this study. It has been suggested that the transition to dioecy in *A. tuberculatus* is incomplete and ongoing, with the Y and X chromosomes still in an early stage of divergence (proto-Y/X) [62]. Consistent with this, our finding that intact LTR retrotransposon accumulation dated to less than 0.5 MYA within the candidate region associated with sex supports the hypothesis that recombination suppression within this region is relatively recent. Waselkov et al. [67] estimated the stem age of *Amaranthus* at 16.2–11.1 MYA, while Wang et al. [55] estimated that *A. tuberculatus* diverged from *A. palmeri* and monoecious amaranths 5.29 MYA. It therefore appears that most of the intact LTR insertions into the sex-associated region occurred after speciation. Nevertheless, the conservation of sex markers, including *FT*, among three species (*A. acanthochiton*, *A. cannabinus*, and *A. greggii*) closely related to *A. tuberculatus* [68] suggests that the sex-associated region is more ancient, potentially preceding the divergence of these species.

While numerous studies have demonstrated the significance of PEBP proteins in the growth and development of multiple plant species, few studies have investigated the diversity of the PEBP protein family in *Amaranthus* species. We leveraged the high-quality chromosome-level genomes of species within the genus to bridge this knowledge gap. Our genome-wide PEBP protein analysis suggests several factors (e.g., lineage-specific duplications, retention bias following whole genome duplication, or transposon-mediated duplications) could have given rise to the additional copies of *PEBP* genes (specifically *FT*-like) in *A. tuberculatus* and *A. palmeri*, compared to the other *Amaranthus* species. The imbalance in *FT*-like gene copies between Hap1 and Hap2 of *A. tuberculatus* may however be due to either haplotypic differences or fragmentation of the Hap2 assembly. Despite significant advances in genome sequencing and assembly, contiguous assembly and accurate phasing are still major challenges for complex genomes [69, 70]. Although only two genes; a *BFT* gene (AmaTuChr01Ag015040 in Hap1 or AmaTuChr01Bg014490 in Hap2) and a *YY-PEBP* gene (AmaTuChr10Ag207360 in Hap1 or AmaTuChr10Bg200600 in Hap2), among the *PEBP* genes were upregulated in mature male flowers, it is possible that the *FT*-like genes have time-specific or tissues-specific expression patterns not detected in our mRNA data. While *BFT* promotes meristem growth in plants [71], *YY-PEBP* (designated *STEPMOTHER OF FT AND TFL1*) represents a novel member of the PEBP gene family that appear to have originated via horizontal gene transfer (HGT) from a prokaryote into an ancestral streptophyte, although its evolutionary origin within Viridiplantae remains unclear [43]. Further analyses suggest that this *YY-PEBP* gene may contribute to the maintenance of reproductive structure vitality and is present as a single copy [43]. Although the function of *YY-PEBP* in amaranths has not yet been determined, our results indicate that it also exists as a single copy in this lineage. Additional analysis of the FT protein sequence in the amaranths revealed that the copy within the MSY of *A. tuberculatus* is closely related to a copy located on Chr00 (unplaced contigs) in *A. palmeri*, with 67% bootstrap support. These two copies likely share a common ancestor but have since diverged. It is thus possible that these *FT* gene copies confer some male fitness advantage by promoting earlier flowering of males in both species, as previously hypothesized [32, 34]. Overexpression of *FT1* has been shown to induce early flowering in several species e.g., *Manihot esculenta* [72], and *Perilla frutescens* [73]. In *Beta vulgaris*, *BvFT1* (Bevul.9G138100) represses flowering and contributes to vernalization, whereas *BvFT2* (Bevul.4G214200) promotes flowering [74]. *BvFT1* is orthologous to *CqFTL1/CqFTL2* in *C. quinoa* [75] and to *CrFTL2* in *C. rubrum* [76], while *BvFT2* is orthologous to *CqFTL4/CqFTL5* or *CrFTL1* [75, 76]. Notably, *CqFTL1/CqFTL2* are also referred to in the literature as *CqFT1A/CqFT1B-1/CqFT1B-2*, and *CqFTL4/CqFTL5* as *CqFT2A/CqFT2B* [77]. Our analysis revealed that *CqFTL1/CqFTL2* and *CqFTL4/CqFTL5* have homologous copies in the amaranths; however future work incorporating codon-based alignments may capture more recent substitutions and thereby improve the resolution of phylogenetic relationships among taxa.

Understanding how mitochondrial or chloroplast DNA segments integrate into the nuclear genome can reveal the mechanisms of genetic transfers, how gene functions are regulated, or provide insights into genome evolution and adaptation [78–81]. Numerous studies have revealed DNA insertions of organellar origin into the nuclear genome, including sex chromosomes or sex-determining regions, in multiple species [82–84]. The mitochondrial and chloroplast genomes of *A. tuberculatus* (~ 360 kbp) assembled in this study allowed for investigation of these nuclear integrants originating from the organelles. While the mitochondrial genome is comparable in length to other species within the Amaranthaceae family (*B. vulgaris* ssp. *vulgaris*; 368.8 kbp, *B. vulgaris* ssp. *maritima*; 364.96 kbp, *S. oleracea*; 329.6 kbp, and *C. quinoa*; 315 kbp) [85–87], there were differences in the number

of protein-coding genes, specifically, *rps7* was not annotated in the *A. tuberculatus* mitogenome, or detected via a BLAST search of the *Chenopodium quinoa rps7* copy to the mitogenome. It is possible that the *rps7* gene is lost in *A. tuberculatus*; however, additional mitochondrial genomes of species within the genus would be required to ascertain if the absence of the gene is lineage specific. Like *B. vulgaris* and *C. quinoa*, the *A. tuberculatus* mitochondrial genome also contains two $tRNA^{Cys}$ genes (*trnC-GCA*): one around 151.7 kbp and the other at 217.3 kbp, which had only been reported in *B. vulgaris* [86] and *C. quinoa* [85]. Although mitochondrion-to-plastid DNA transfer has been documented for some lineages [88], our analysis revealed a unidirectional migration of DNA segments from chloroplast to the mitochondrial genome, as reported for multiple plant species [88–91]. Overall, < 0.02% of the SDR on Hap1 is made up of NUMTs, while 0.01% of the region is made up of NUPTs. The absence of both NUMTs and NUPTs within the 3.19 Mbp male-specific Y region is congruent with findings in species such as *Asparagus officinalis*, where the male-specific region of the Y chromosome lacked NUPT insertions [83].

Conclusions

We present high-quality, contiguous assemblies of the nuclear, chloroplast, and mitochondrial genomes of *A. tuberculatus*, including a chromosome-level nuclear assembly. These genomic resources provide a valuable foundation for exploring the evolution of the sex-associated region, diversity within the PEBP protein family, and nuclear DNA segments of organellar origin in this troublesome dioecious weed. Our analyses suggest that chromosome 1, which contains the sex-associated region, likely originated from the fusion of two ancestral chromosomes. While the absolute age of the sex-associated region is unknown, the ~0.5 MYA estimate for intact LTR insertions reflects subsequent TE accumulation dynamics rather than the origin itself. We further identified sex-linked and sex-biased differentially expressed genes with potential roles in sex determination and flowering. Together, these findings establish a framework for systematic comparative genomics across dioecious *Amaranthus* species to better understand the origin, variation, and evolution of sex-determining regions in the genus.

Materials and methods

Plant material and growth conditions

An accession of *A. tuberculatus*, designated WUS-BM and available through the USDA Germplasm Resources Information Network (GRIN: PI 698378), was used in this study. The original source population, designated the ABD population, from which WUS-BM was derived, was collected by Katherine E. Waselkov in 2009 at Aberdeen Community Park on the north bank of the Ohio River (GPS coordinates: 38.6543, −83.7623). Seeds from the accession were sown in presoaked potting soil (Lambert LM-GPS). Seedlings were grown to ~ 5 cm in height under a 16/8-h day/night photoperiod and 25/20 C temperature in the greenhouse prior to transplanting to 16-cm pots (America Clay Works I-A650MP). A single flowering male plant was placed in the dark for 72 h, after which 4 g and 2 g of fresh leaf tissues were sampled for PacBio HiFi and Hi-C library preparation, respectively. In each case, tissue was flash frozen in liquid nitrogen and stored at −80 C until use. Four grams of fresh leaf tissue was harvested from the same individual and shipped immediately on damp paper towels at 4 C for Bionano library preparation. Finally, a mix of four male and four female plants, assigned genotypically as described by [32], were grown as previously described. Root, stem, leaf, and meristem tissue were sampled from young (~ 6–10 leaf stage) and flowering plants and combined with floral tissue from several stages of development. All tissues were ground together in equal amounts by weight, and RNA was extracted using Zymo Direct-zol RNA Miniprep kit. Following quantification of concentration and quality, three µg of RNA was used for PacBio Iso-Seq library preparation. All samples were shipped on dry ice (except the tissue for Bionano library preparation) to the Genome Center of Excellence at Corteva Agriscience.

Library preparation, genome sequencing, and assembly

DNA from a single *A. tuberculatus* individual was isolated from approximately 1.2 g of frozen leaf tissue using the Nucleobond HMW DNA kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) following the manufacturer's protocol with some modifications [92]. Oxford Nanopore Technologies (ONT) ligation sequencing libraries were constructed using the LSK114 Kit (Oxford Nanopore Technologies, Oxford, UK) as per standard protocol with a few modifications. The libraries were loaded into FLO-PRO114M Flowcells, and data were collected using a PromethION 24 sequencer with Configuration v.5.8.6, MinKNOW v.23.11.7, Bream 7.8.2, and Dorado 7.2.13. In total, six flowcells were used to complete 12 runs yielding 2,305,539 reads with minimum length of 40 kbp, N50 of 54.7 kbp and total sequence length of 128,275,261,579 bp. Hi-C library was constructed using the Phase Genomics Proximo system following the manufacturer's protocol (Phase Genomics, Seattle, WA). The library was then sequenced on an Illumina Novaseq 6000 system, yielding 142,208,874 PE150 paired clusters for a total sequence length of 43,862,662,200 bp. For Bionano optical mapping, nuclei were extracted from approximately 500 mg of young leaf tissue using Prep™ Plant Tissue DNA Isolation Kit (Bionano Genomics, San Diego, CA, USA). Ultra-high molecular weight (uHMW) DNA was then obtained from the resulting nuclei following resuspension and

centrifugations (first at low speed of 100 x g, and subsequently at 2,500 x g). Bionano Direct Label and Stain (DLS) protocol was then carried out with 800 ng of the uHMW using a DLS kit. After the dye cleaning step, the DNA was loaded into a single Bionano chip in a Saphyr system following the manufacturer's recommendations.

Filtered ONT simplex reads (104x) were assembled with hifiasm v0.21.0 with default parameters using the new module for ONT only assembly (--ont) [93]. The resulting primary contigs were filtered for minimum length (100 kbp) and minimum coverage (20x) thresholds yielding 64 filtered contigs (contig N50 = 34.6 kbp, sum = 1.20 Gbp). Contigs were combined with 70 Bionano maps (map N50 = 29.1 Mbp, sum = 1.20 Gbp) to create hybrid scaffolds using Bionano Access v1.8 and Bionano Solve v3.8, which were manually curated to a set of 32 scaffolds (scaffold N50 = 36.7 Mbp, sum = 1.195 Gbp). Hi-C data were aligned to scaffolds using Juicer v1.6 [94] to generate Hi-C contact maps. Visual inspection of the maps with Juicebox v1.11.08 [95] revealed that the 32 scaffolds represented 32 chromosomes. Homeologous scaffold/chromosome pairs were identified and oriented by chunking each scaffold into 100 bp "reads" and aligning to the other scaffolds using minimap2 v2.26 (using short read mode) [92]. The top matches were identified by visual inspection of the alignment coordinates with Tibco Spotfire (v14.0). Scaffolds were organized into 16 chromosome pairs, randomly designated as A (Hap1) and B (Hap2), with A representing the longer chromosome of each pair. Corrected ONT reads from hifiasm were aligned to the chromosome assembly using minimap2 (HiFi mode) to correct spurious consensus SNP/INDEL errors using samtools v1.3.1 [96]. In addition, manual curation was performed to resolve two overlapping contigs and one collapsed 327 kbp tandem repeat. The final assembly contains only 15 gaps, and the remaining gaps appear to be associated with AT-rich repeats and reduced ONT coverage.

Genome annotation

The annotation pipeline had been previously described [97]. Briefly, transposable element families were first identified in the two chromosome-level assemblies using RepeatModeler v2.0.2 [98], RepeatMasker v4.1.2 (http://www.repeatmasker.org/RepeatMasker/), and BEDtools v2.30.0 [99]. Iso-Seq reads were then mapped to the softmasked assemblies generated from the repetitive elements' annotation using pbmm2 v1.10.0. Redundant transcripts were collapsed using IsoSeq3 v3.8.2 (https://github.com/PacificBiosciences/IsoSeq). The collapsed gene models together with protein sequences from *A. hypochondriacus* [100] were used for the final gene model predictions in MakerP v1.0. Functional annotation and subcellular localization of proteins were carried out utilizing series of databases, including InterProScan5 [101], Uniref50 [102], and NCBI nonredundant protein database. Transfer RNA (tRNA) genes were predicted using tRNAscan-SE v2.0.12 [103] within Maker. Transcription factors were identified using PlantTFDB v5.0 (http://planttfdb.gao-lab.org/prediction.php) [104].

Genome characteristics and repeat analysis

The assembly quality and completeness for both haplotypes were accessed using the Benchmarking Universal Single-Copy Orthologs (BUSCOs) v5.8.2 [105]. Genome characteristics were computed using "agat_sp_statistics.pl" from AGAT Toolkit v1.4.1 [106] and "bbstats.sh" from BBTools [107]. Species-specific repeats in the two haplotype assemblies were first identified using RepeatModeler v2.0.6 [98]. The RepBase database (RepeatMaskerEdition-20181026) [108] was then combined with Dfam3.8 database in RepeatMasker, and queried using the famdb.py utility in RepeatMasker to obtain 'viridiplantae' repeats library. LTR structural analysis was carried out separately using LTR_harvest from genometools v1.6.0 with parameters: -minlenltr 100 -maxlenltr 7000 -mintsd 4 -maxtsd 6 -motif TGCA -motifmis 1 -similar 85 -vic 10 -seed 20 -seqids yes [109], and LTR_FINDER_parallel with default parameters [110]. The output from both LTR_harvest and LTR_FINDER_parallel were combined and passed to LTR_retriever v3.0.1 [111] to generate a non-redundant LTR library for each haplotype.

The substitution rate of 2.81e-9 synonymous substitutions per site per year for the Chenopodioideae plastid *rbcL* gene was specified with -u parameter within LTR_retriever to estimate the insertion age of LTR elements. The three libraries (species-specific consensus libraries of repeats, viridiplantae repeats library, and non-redundant LTR libraries) were combined for each of the haplotypes, and clustered using CD-HIT-EST v4.8.1 [112] with parameters: -c 0.8 -G 0.8 -s 0.9 -aL 0.9 -aS 0.9 -M 0 -T 48. The non-redundant libraries were then used to annotate repeats in the haplotype assemblies using RepeatMasker v4.1.7 with parameters: -s -gff -lib -e rmblast (http://www.repeatmasker.org/RepeatMasker/). Annotated repeats were summarized with RepeatMasker's "buildSummary.pl" script, and the distribution of LTRs was estimated using the perl script "LTR_sum.pl" within LTR_retriever.

De novo quality of intergenic and repetitive sequence space of both haplotypes was also accessed using LTR assembly index (LAI) from the LTR retriever pipeline [111, 113]. Putative centromeric repeats in both haplotypes were identified by running StainedGlass [114] with default parameters (window = 5000 and mm_f = 10000) to determine sequence identity within tandem repeat arrays. Telomeric repeats were identified by searching the telomere repeat motif $(TTTAGGG)_n$

(where n = 4) against both haplotypes using BLASTN [115] with parameters: -task blastn-short -db haplotypes -query primers -outfmt 6 -out output.

Genome-wide association analysis for sex in waterhemp populations

A total of 353 individuals (175 females and 178 males) derived from an artificially generated mapping population [116] and previously genotyped using RAD-seq [30] were used for GWA analysis. The single-end reads were demultiplexed and processed using Stacks v2.68 [117]. Each sample was then aligned to Hap1 of the *A. tuberculatus* genome assembly using BWA-MEM v0.7.18 [118]. Individuals were grouped by sex, and the "gstacks" command in Stacks was used to build loci. We ran the "populations" command employing a minimal filtering (--min-maf 0.05) and then assessed the level of missingness in the two data groups using VCFtools (--missing-indv) following a previously described approach [119]. Samples having more than 60% missing data were excluded from further analysis, and the "populations" command rerun on the streamlined samples using more filtering criteria (--min-maf 0.05, -p 2, -r 0.4). GWA analysis was carried out using complementary approaches with the set of variants obtained after the filtering step above. First, we obtained kinship matrix using emmax-kin, and used the matrix as covariates for EMMAX association, which utilizes a mixed-model association method (parameters: -d 10, -v). Second, we obtained principal components (PC) using PLINK v1.90b7 [120], and then used the first 20 PCs as covariates for the association analysis in BLINK-C [121]. Sex of the individuals were used as phenotypes in the analysis. Calculation of f-statistics (F_{ST}) was performed using VCFtools v0.1.16 [122].

Location of previous sex markers and genes in the assemblies

BLASTN [115] was used to search previously reported primer sets (WHMS, MU-976, MU-657.2, and MU-533) that amplified a male-specific region of *A. tuberculatus* [30, 32] against both haplotype assemblies. Additionally, a 200 bp *FT* sequence, previously reported as male-specific and conserved among three other dioecious amaranths related to waterhemp [34], was searched against both haplotype assemblies using BLASTN. Reciprocal best hit (RBH) searches between genes on the previously identified male-specific contigs and genes from the haplotype assemblies were also carried out with MMseqs2 [123]. Putative homologs of genes within the male-specific Y region elsewhere in the genome were identified with MMseqs2 using parameters --min-seq-id 0.7 -c 0.8 --cov-mode 1.

Synteny and intragenomic analysis

Collinear gene blocks between haplotype 1 and 2 assemblies were identified using MCScan [124] from the JCVI utility libraries v1.3.5 with a C-score cutoff of 0.99. Blocks on chromosome 1 were visualized and plotted with JBrowse [125]. MCScanX [126] was also used to investigate collinear blocks of genes within the haplotypes, and the output collinearity files, after converting to ".anchors" format, were plotted with Circos [127]. Haplotype sequences were aligned with Minimap2 v2.28-r1209 [128], further processed with SyRI [129] to identify structural rearrangements, and visualized using D-Genies [130]. Syntenic orthologues among the haplotypes and six chromosome-level *Amaranthus* species (*A. hypochondriacus* [100], *A. cruentus* [131], *A. tricolor* [55], *A. palmeri*, *A. retroflexus*, and *A. hybridus* [132]) were evaluated and visualized using GENESPACE v1.2.3 [133].

Sequence divergence and detection of positive selection

We followed previously described protocols to test for the evidence of adaptive evolution in 608 single-copy orthologous genes on chromosome 1 [134, 135]. Briefly, each orthogroup on chromosome 1 from the previous Orthofinder run containing the seven *Amaranthus* species were aligned separately using custom Perl scripts, "multiple_sequence_splitter.pl" and "align_orthologs.pl" from Jeffares et al., (2014). The codon alignments for the 608 orthologs were then used as input for CODEML program in PAML package v4.10.7 [136]. A series of models were adopted to test for adaptive evolution. We first compared the simplest or null model against a nested alternative site model (M0 vs. M1a) that allows neutral sites ($\omega = 1$) to be evaluated. We then asked if adding a third class with $\omega > 1$ fits the data better than a model with only two classes, $\omega < 1$ or $\omega = 1$ (M1a vs M2a), as certain sites could be under positive selection. We further used the branch models to determine if specific foreground branches have different ω from background branches (M0 vs. two-ratio).

Using the branch-site model (MA_{null}, $\omega = 1$ vs. MA, $\omega > 1$), we determined if the previously defined foreground branches are more likely to contain sites under positive selection (see Fig. S14 of schematic representation of trees with defined foreground and background branches). For all hypotheses tested, we implemented the mutation-selection model (FMutSel) with observed codon frequencies used as estimates (CodonFreq = 7, estFreq = 0). The FMutSel model has been indicated to account for mutation bias and selection affecting codon usage, and is preferable over other models [135, 137]. We compared per-gene nested models using likelihood-ratio tests, and p-values were adjusted using the Benjamini–Hochberg method [138] to account for multiple comparisons.

Expression profiling and gene ontology (GO) enrichment analysis

mRNA-sequencing data for three tissue types (mature flowers, shoot apical meristem, and floral meristems) from a previous study [56] were mapped to Hap1 of the *A. tuberculatus* genome using STAR v2.7.11b [139]. Two replicates from the mature flower category were excluded due to low mapping quality (25.05% and 26.86% uniquely mapped reads for replicates 2 and 4), consistent with their removal in the previous study. Gene counts were obtained with featureCounts v2.0.8 from the subread package [140], and the differential expression (DE) between sexes for each tissue type was analyzed using edgeR [141]. Genes with CPM < 1 in at least two samples were removed, and counts were normalized using the TMM method. A negative binomial generalized log-linear model was fitted to the normalized counts, and genes were considered differentially expressed using edgeR's *glmTreat* function with FDR < 0.05 and FC >1.2. Heatmaps of these differentially expressed genes were generated with pheatmap after $\log_2$-transformed normalization of counts in DESeq2 [142]. Translated protein-coding sequences (CDS) of Hap1 were annotated with eggNOG-mapper v2.1.12 [143], and GO enrichment was assessed with topGO (nodeSize = 10) using Fisher's exact test and the "elim" algorithm. GO terms were considered significantly enriched at $p < 0.01$.

PEBP gene family diversity

Protein sequences of *A. tuberculatus* haplotype assemblies from this study, *A. hypochondriacus* [100], *A. cruentus* [131], *A. tricolor* [55], *A. palmeri*, *A. hybridus*, and *A. retroflexus* [132], and *Beta vulgaris* ssp. *vulgaris* [144] were retrieved for candidate PEBP classification. Two complementary approaches were utilized. First, protein sequences encoding the six PEBP genes in *A. thaliana* TAIR10 (Arabidopsis TAIR database) were searched against the *Amaranthus* species and *B. vulgaris* using BLASTP (e-value of 1e-5). Second, the Hidden Markov Model (HMM) for PEBP family (PF01161) retrieved from Pfam (http://pfam.xfam.org, accessed on J anuary 29, 2025) were searched against *Amaranthus* species and *B. vulgaris* protein databases using hmmsearch from HMMER v3.4 [145]. All hits from the HMM search were further verified for the conserved PEBP domain using CDD (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi). Predicted protein sequences lacki ng a PEBP domain were excluded from further analysis. Additionally, the twenty-three PEBP protein sequences previously reported for *Chenopodium quinoa* [75, 77] and PEBP sequences of *Spinacia oleracea* retrieved from Phytozome v14 [146, 147] were included for alignment. PEBP protein sequences from all species were then aligned using MAFFT v7.511 [148]. Phylogenetic tree of the PEBP proteins was constructed using the maximum-likelihood method in RAxML v8.2.13 under the JTT + GAMMA model, selected via PROTGAMMAAUTO, with 1000 rapid bootstrap replicates. The phylogenetic tree was midpoint-rooted, visualized, and annotated in iTOL v7 [149]. The gene structure of *FLOWERING LOCUS T* and its homologs were visualized using GSDS 2.0 [150].

Organellar genomes and their homologous sequences in the nuclear genome

The chloroplast and mitochondrial genomes of *A. tuberculatus* were contiguously assembled as part of the assembly pipeline described previously for the nuclear genome. We, however, focused on the annotation of the mitochondrial genome of *A. tuberculatus* due to previous work on the chloroplast genome of the species [57, 58]. The assembled mitogenome was annotated with GeSeq [151] using BLAT, tRNAscan-SE v2.0.7, and NCBI RefSeq sequences from *B. maritima*, *B. vulgaris*, *S. oleracea* and *C. quinoa*. The BLAT search parameters for protein search identity and rRNA, tRNA, DNA search identity was set to 45% and 85%, respectively. The mitochondrial genome and the annotation were then visualized using the program OGDRAW [152]. Simple sequence repeats were identified using MISA v2.1 (https://webblast.ipk-gatersleben.de/misa/) with parameters: 12, 6, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively [153]. Repetitive sequences were further identified with REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) with parameters: 30 bp (minimum repeat size), 100 (maximum computed repeats), 3 (hamming distance) [154]. Previously reported mitochondrial genomes of *C. quinoa* (GenBank accession MK182703.1), *S. oleracea* (GenBank accession NC_035618.1) and *B. vulgaris* ssp. *vulgaris* (GenBank accession NC_002511.2) were obtained from NCBI and compared to the *A. tuberculatus* mitochondrial genome using minimap2 [92]. Alignments were visualized as dotplots with D-GENIES [130].

To identify nuclear sequences of organellar origin and fragments transferred between organelles (i.e., from chloroplast to mitochondrion), two complementary approaches were utilized. First, coding sequences (CDS) from the mitogenome and plastome were searched against the nuclear genome, and the coding sequences from the plastome were searched against the mitogenome using BLASTN [115]. Second, the complete mitogenome and plastome were searched against the nuclear genome using BLASTN, allowing for the identification of longer homologous blocks or fragments inserted into the nuclear genome. The BLAST parameters used included a minimum identity threshold of 80%, e-value cutoff of 1×10^{-5}, and alignment lengths exceeding 100 bp. The BLAST outputs were manually filtered to remove overlapping or redundant matches.

Statistical analysis

Gene count per 500 kbp, TE proportion per 500 kbp, and LTR insertion times data were analyzed using Kruskal-Wallis test. Post-hoc test of multiple comparisons between regions (Collinear 1, INS 1, INV 1, Collinear 2, INV 2, Collinear 3) was carried out with Conover-Iman test, following the rejection of the null hypothesis from the Kruskal-Wallis test. *P*-values were adjusted for multiple comparisons with Benjamini–Hochberg correction [138]. The statistical analyses were carried out with the PMCMRplus package in R v4.1.2 [155].

Abbreviations

BLINK	Bayesian-information and linkage disequilibrium iteratively nested keyway
CDD	Conserved domain database
EMMAX	Efficient mixed-model association eXpedited
LTR-RTs	Long terminal repeat retrotransposons
MSY	Male-specific Y region
NUMTs	Nuclear mitochondrial DNA segments
NUPTs	Nuclear plastid DNA segments
PAR	Pseudoautosomal region
PEBP	Phosphatidylethanolamine binding protein
RAD-GWA	Restriction site-associated DNA Genome-wide association
SDR	Sex-determining region
TMM	Trimmed Mean of M-values

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-12181-w.

Supplementary Material 1.

Supplementary Material 2.

Authors' contributions

P.J.T., T.A.G., and E.L.P. conceived the research study. J.S.M. and S.M. grew plants, harvested, and prepared samples for shipping to the sequencing center. V.L. and K.F. performed DNA and RNA extraction, ONT sequencing, Bionano DLS, Hi-C Seq, Iso-Seq sequencing, as well as genome data integration and assembly. J.S.M, L.C., and E.L.P. annotated the genome and provided genome assembly metrics. D.A.R. carried out genome-wide association analysis, adaptive evolution analysis, transposable elements analysis, comparative genomics analyses, PEBP protein family diversity, and statistical analysis with supervision of P.J.T. D.A.R. wrote the manuscript. All authors read, revised, and approved the manuscript.

Funding

This work was supported by the International Weed Genomics Consortium with funding from Foundation for Food & Agriculture Research (FFAR grant number DSnew-0000000024), Bayer Crop Science, Corteva Agriscience, Syngenta, BASF, and the Global Herbicide Resistance Action Committee. We also acknowledge funding from the USDA National Institute of Food and Agriculture (Grant number 2022-67013-36142 to PJT).

Data availability

Sequencing data and the genome assemblies in this study are available through the National Center for Biotechnology Information (NCBI) under BioProject PRJNA1089550. The genome assemblies and annotations are also available on CoGe under accession IDs 69059 and 69062. The mitochondrial and chloroplast genomes assemblies in this study were submitted along with the nuclear genomes under BioProject PRJNA1085752. PEBP protein sequences, MAFFT alignment, and RAxML tree are available on figshare https://doi.org/10.6084/m9.figshare.29856191.

Declarations

Ethics approval and consent to participate

The plant material, including seeds, used in this research does not require permission, license, or ethical approval.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 July 2025 / Accepted: 3 October 2025
Published online: 03 November 2025

References

1. Renner SS, Ricklefs RE. Dioecy and its correlates in the flowering plants. Am J Bot. 1995;82:596.
2. Ming R, Bendahmane A, Renner SS. Sex chromosomes in land plants. Annu Rev Plant Biol. 2011;62:485–514.
3. Renner SS. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. Am J Bot. 2014;101:1588–96.
4. Charlesworth B, Charlesworth D. A model for the evolution of dioecy and gynodioecy. Am Nat. 1978;112:975–97.
5. Lloyd DG. The distributions of gender in four angiosperm species illustrating two evolutionary pathways to dioecy. Evolution. 1980;34:123–34.
6. Bawa KS. Evolution of dioecy in flowering plants. Annu Rev Ecol Syst. 1980;11:15–39.
7. Henry IM, Akagi T, Tao R, Comai L. One hundred ways to invent the sexes: theoretical and observed paths to dioecy in plants. Annu Rev Plant Biol. 2018;69:553–75.
8. Akagi T, Charlesworth D. Pleiotropic effects of sex-determining genes in the evolution of dioecy in two plant species. Proceedings of the Royal Society B: Biological Sciences. 2019;286:20191805.
9. Cronk Q, Müller NA. Default sex and single gene sex determination in dioecious plants. Front Plant Sci. 2020;11:1–5.
10. Akagi T, Pilkington SM, Varkonyi-Gasic E, Henry IM, Sugano SS, Sonoda M, et al. Two Y-chromosome-encoded genes determine sex in Kiwifruit. Nat Plants. 2019;5:801–9.
11. Harkess A, Huang K, van der Hulst R, Tissen B, Caplan JL, Koppula A, et al. Sex determination by two Y-linked genes in garden asparagus. Plant Cell. 2020;32:1790–6.
12. Akagi T, Henry IM, Tao R, Comai L. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. Science. 2014;346:646.
13. Müller NA, Kersten B, Leite Montalvão AP, Mähler N, Bernhardsson C, Bräutigam K, et al. A single gene underlies the dynamic evolution of Poplar sex determination. Nat Plants. 2020;6(6):630–7.
14. Massonnet M, Cochetel N, Minio A, Vondras AM, Lin J, Muyle A, et al. The genetic basis of sex determination in grapes. Nat Commun. 2020;11:1–12.
15. Ma X, Yu L, Fatima M, Wadlington WH, Hulse-Kemp AM, Zhang X, et al. The spinach YY genome reveals sex chromosome evolution, domestication, and introgression history of the species. Genome Biol. 2022;23:1–30.
16. Na JK, Wang J, Ming R. Accumulation of interspersed and sex-specific repeats in the non-recombining region of Papaya sex chromosomes. BMC Genomics. 2014;15:1–12.
17. Hobza R, Kubat Z, Cegan R, Jesionek W, Vyskot B, Kejnovsky E. Impact of repetitive DNA on sex chromosome evolution in plants. Chromosom Res. 2015;23:561–70.
18. Charlesworth D. Plant sex chromosomes. Annu Rev Plant Biol. 2016;67:397–420.
19. Charlesworth D. Young sex chromosomes in plants and animals. New Phytol. 2019;224:1095–107.
20. Akagi T, Varkonyi-Gasic E, Shirasawa K, Catanach A, Henry IM, Mertten D, et al. Recurrent neo-sex chromosome evolution in kiwifruit. Nat Plants. 2023;9:393–402.
21. Du Q, Wu Z, Liu P, Qing J, He F, Du L, et al. The chromosome-level genome of *Eucommia ulmoides* provides insights into sex differentiation and α-linolenic acid biosynthesis. Front Plant Sci. 2023;14:1–12.

22. Healey AL, Piatkowski B, Lovell JT, Sreedasyam A, Carey SB, Mamidi S, et al. Newly identified sex chromosomes in the *Sphagnum* (peat moss) genome alter carbon sequestration and ecosystem dynamics. Nat Plants. 2023;9:238–54.
23. Kafkas S, Ma X, Zhang X, Topçu H, Navajas-Pérez R, Wai CM et al. Pistachio genomes provide insights into nut tree domestication and ZW sex chromosome evolution. Plant Commun. 2023;4:100497. https://doi.org/10.1016/j.xplc.2022.100497.
24. Montgomery J, Morran S, MacGregor DR, McElroy JS, Neve P, Neto C, et al. Current status of community resources and priorities for weed genomics research. Genome Biol. 2024;25:139.
25. Sauer J. Recent migration and evolution of the dioecious amaranths. Evolution. 1957;11:11–31.
26. Steckel LE. The dioecious *Amaranthus* spp.: here to stay. Weed Technol. 2007;21:567–70.
27. Sauer J. The dioecious amaranths: A new species name and major range extensions. Madrono. 1972;21:426.
28. Mosyakin SL, Robertson KR. New infrageneric taxa and combinations in *Amaranthus* (Amaranthaceae). Ann Bot Fennici. 1996;33:275–81.
29. Murray MJ. The genetics of sex determination in the family Amaranthaceae. Genetics. 1940;25:409–31.
30. Montgomery JS, Sadeque A, Giacomini DA, Brown PJ, Tranel PJ. Sex-specific markers for waterhemp (*Amaranthus tuberculatus*) and Palmer amaranth (*Amaranthus palmeri*). Weed Sci. 2019;67:412–8.
31. Grant WF. Cytogenetic studies in Amaranthus. Can J Bot. 1959;37:413–7.
32. Montgomery JS, Giacomini DA, Weigel D, Tranel PJ. Male-specific Y-chromosomal regions in waterhemp (*Amaranthus tuberculatus*) and Palmer amaranth (*Amaranthus palmeri*). New Phytol. 2021;229:3522–33.
33. Montgomery JS, Giacomini D, Waithaka B, Lanz C, Murphy BP, Campe R, et al. Draft genomes of *Amaranthus tuberculatus*, *Amaranthus hybridus*, and *Amaranthus palmeri*. Genome Biol Evol. 2020;12:1988–93.
34. Raiyemo DA, Bobadilla LK, Tranel PJ. Genomic profiling of dioecious *Amaranthus* species provides novel insights into species relatedness and sex genes. BMC Biol. 2023;21:1–18.
35. Harig L, Beinecke FA, Oltmanns J, Muth J, Müller O, Rüping B, et al. Proteins from the FLOWERING LOCUS T-like subclade of the PEBP family act antagonistically to regulate floral initiation in tobacco. Plant J. 2012;72:908–21.
36. Freytes SN, Canelo M, Cerdán PD. Regulation of flowering time: when and where? Curr Opin Plant Biol. 2021;63:102049.
37. Chardon F, Damerval C. Phylogenomic analysis of the PEBP gene family in cereals. J Mol Evol. 2005;61:579–90.
38. Karlgren A, Gyllenstrand N, Källman T, Sundström JF, Moore D, Lascoux M, et al. Evolution of the PEBP gene family in plants: functional diversification in seed plant evolution. Plant Physiol. 2011;156:1967–77.
39. Wang M, Tan Y, Cai C, Zhang B. Identification and expression analysis of phosphatidy ethanolamine-binding protein (PEBP) gene family in cotton. Genomics. 2019;111:1373–80.
40. Danilevskaya ON, Meng X, Hou Z, Ananiev EV, Simmons CR. A genomic and expression compendium of the expanded PEBP gene family from maize. Plant Physiol. 2008;146:250–64.
41. Dong L, Lu Y, Liu S. Genome-wide member identification, phylogeny and expression analysis of PEBP gene family in wheat and its progenitors. PeerJ. 2020;8:1–22.
42. Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT et al. Activation tagging of the floral inducer FT. Science. 1999;286:1962–5.
43. Bellinazzo F, Nadal Bigas J, Hogers RAH, Kodde J, van der Wal F, Kokkinopoulou P, et al. Evolutionary origin and functional investigation of the widely conserved plant PEBP gene stepmother of FT and TFL1 (SMFT). Plant J. 2024;120:1410–20.
44. Liao Z, Zhang T, Lei W, Wang Y, Yu J, Wang Y et al. A telomere-to-telomere reference genome of ficus (*Ficus hispida*) provides new insights into sex determination. Hortic Res. 2024;11:uhad257. https://doi.org/10.1093/hr/uhad257.
45. Dowling CA, Michael TP, Mccabe PF, Schilling S, Melzer R. FT-like genes in cannabis and hops: sex specific expression and copy-number variation may explain flowering time variation. BioRxiv. 2024;v1:1–34. https://doi.org/10.1101/2024.10.04.616617.
46. Okada S, Sone T, Fujisawa M, Nakayama S, Takenaka M, Ishizaki K, et al. The Y chromosome in the liverwort *Marchantia polymorpha* has accumulated unique repeat sequences harboring a male-specific gene. Proc Natl Acad Sci U S A. 2001;98:9454–9.
47. Wang J, Na JK, Yu Q, Gschwend AR, Han J, Zeng F, et al. Sequencing papaya X and Y h chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc Natl Acad Sci U S A. 2012;109:13710–5.
48. Kejnovsky E, Kubat Z, Hobza R, Lengerova M, Sato S, Tabata S, et al. Accumulation of chloroplast DNA sequences on the Y chromosome of *Silene latifolia*. Genetica. 2006;128:167–75.
49. Yamato KT, Ishizaki K, Fujisawa M, Okada S, Nakayama S, Fujishita M, et al. Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proc Natl Acad Sci U S A. 2007;104:6472–7.
50. Vanburen R, Ming R. Organelle DNA accumulation in the recently evolved papaya sex chromosomes. Mol Genet Genomics. 2013;288:277–84.
51. Yue J, VanBuren R, Liu J, Fang J, Zhang X, Liao Z, et al. Sunup and sunset genomes revealed impact of particle bombardment mediated transformation and domestication history in papaya. Nat Genet. 2022;54:715–24.
52. Sousa A, Bellot S, Fuchs J, Houben A, Renner SS. Analysis of transposable elements and organellar DNA in male and female genomes of a species with a huge Y chromosome reveals distinct Y centromeres. Plant J. 2016;88:387–96.
53. Kozik A, Rowan BA, Lavelle D, Berke L, Eric Schranz M, Michelmore RW, et al. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. PLoS Genet. 2019;15:1–30.
54. Stetter MG, Schmid KJ. Analysis of phylogenetic relationships and genome size evolution of the *Amaranthus* genus using GBS indicates the ancestors of an ancient crop. Mol Phylogenet Evol. 2017;109:80–92.
55. Wang H, Xu D, Wang S, Wang A, Lei L, Jiang F, et al. Chromosome-scale *Amaranthus tricolor* genome provides insights into the evolution of the genus *Amaranthus* and the mechanism of betalain biosynthesis. DNA Res. 2023;30:1–15.
56. Bobadilla LK, Baek Y, Tranel PJ. Comparative transcriptomic analysis of male and females in the dioecious weeds *Amaranthus palmeri* and *Amaranthus tuberculatus*. BMC Plant Biol. 2023;23:1–26.
57. Xu H, Xiang N, Du W, Zhang J, Zhang Y. Genetic variation and structure of complete chloroplast genome in alien monoecious and dioecious *Amaranthus* weeds. Sci Rep. 2022;12:1–9.
58. Raiyemo DA, Tranel PJ. Comparative analysis of dioecious *Amaranthus* plastomes and phylogenomic implications within Amaranthaceae s.s. BMC Ecol Evol. 2023;23:15.
59. Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol. 2013;14:1–20.
60. Natri HM, Merilä J, Shikano T. The evolution of sex determination associated with a chromosomal inversion. Nat Commun. 2019;10:145. https://doi.org/10.1038/s41467-018-08014-y.
61. Hearn KE, Koch EL, Stankowski S, Butlin RK, Faria R, Johannesson K, et al. Differing associations between sex determination and sex-linked inversions in two ecotypes of *Littorina saxatilis*. Evol Lett. 2022;6:358–74.
62. Kreiner JM, Montgomery JS, Todesco M, Bercovich N, Gong Y, Elphinstone C, Tranel PJ, Rieseberg LH, Wright SI. The evolution of separate sexes in waterhemp is associated with surprising chromosomal diversity and complexity. PLOS Biology. 2025;23:e3003254. https://doi.org/10.1371/journal.pbio.3003254.
63. Gerton JL. A working model for the formation of Robertsonian chromosomes. J Cell Sci. 2024;137:jcs261912. https://doi.org/10.1242/jcs.261912.
64. Harkess A, Mercati F, Abbate L, McKain M, Pires JC, Sala T et al. Retrotransposon proliferation coincident with the evolution of dioecy in asparagus. G3: Genes, Genomes, Genetics. 2016;6:2679–85.
65. Takashi Akagi A, Fujita N, Masuda K, Shirasawa K, Nagaki K, Horiuchi A, et al. Rapid and dynamic evolution of a giant Y chromosome in *Silene latifolia*. Science. 2025;387:637–43.
66. Moraga C, Branco C, Rougemont Q, Veltsos P, Jedlička P, Muyle A, et al. The *Silene latifolia* genome and its giant Y chromosome. Science. 2025;387:630–6.
67. Waselkov KE, Boleda AS, Olsen KM. A phylogeny of the genus *Amaranthus* (Amaranthaceae) based on several low-copy nuclear loci and Chloroplast regions. Syst Bot. 2018;43:439–58.
68. Raiyemo DA, Bobadilla LK, Tranel PJ. Genomic profiling of dioecious *Amaranthus* species provides novel insights into species relatedness and sex genes. BMC Biol. 2023;21:1–18. https://doi.org/10.1186/s12915-023-01539-9
69. Chen X, Baker D, Dolzhenko E, Devaney JM, Noya J, Berlyoung AS, et al. Genome-wide profiling of highly similar paralogous genes using HiFi sequencing. Nat Commun. 2025;16:1–13.
70. Porubsky D, Vollger MR, Harvey WT, Rozanski AN, Ebert P, Hickey G, et al. Gaps and complex structurally variant loci in phased genome assemblies. Genome Res. 2023;33:496–510.

71. Mimida N, Goto K, Kobayashi Y, Araki T, Ahn JH, Weigel D, et al. Functional divergence of the TFL1-like gene family in *Arabidopsis* revealed by characterization of a novel homologue. Genes Cells. 2001;6:327–36.
72. Odipio J, Getu B, Chauhan RD, Alicai T, Bart R, Nusinow DA, et al. Transgenic overexpression of endogenous FLOWERING LOCUS T-like gene MeFT1 produces early flowering in cassava. PLoS ONE. 2020;15:1–16.
73. Xu H, Guo X, Hao Y, Lu G, Li D, Lu J, et al. Genome-wide characterization of PEBP gene family in *Perilla frutescens* and PfFT1 promotes flowering time in *Arabidopsis thaliana*. Front Plant Sci. 2022;13:1–12.
74. Pin PA, Benlloch R, Bonnet D, Wremerth-Weich E, Kraft T, Gielen JJL et al. An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. Science (1979). 2010;330:1397–400.
75. Wu Q, Bai X, Nie M, Li L, Luo Y, Fan Y et al. Genome-wide identification and expression analysis disclose the pivotal PHOSPHATIDYLETHANOLAMINE BINDING PROTEIN members that may be utilized for yield improvement of *Chenopodium quinoa*. Front Plant Sci. 2023;13:1119049. https://doi.org/10.3389/fpls.2022.1119049.
76. Drabešová J, Cerná L, Masterová H, Kolousková P, Potocky M, Štorchová H. The evolution of the FT/TFL1 genes in amaranthaceae and their expression patterns in the course of vegetative growth and flowering in Chenopodium rubrum. G3: Genes Genomes Genet. 2016;6:3065–76.
77. Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJA, et al. The genome of *Chenopodium quinoa*. Nature. 2017;542:307–12.
78. Kleine T, Maier UG, Leister D. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol. 2009;60:115–38.
79. Maréchal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. New Phytol. 2010;186:299–317.
80. Noutsos C, Richly E, Leister D. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Res. 2005;15:616–28.
81. Zhang GJ, Dong R, Lan LN, Li SF, Gao WJ, Niu HX. Nuclear integrants of organellar DNA contribute to genome structure and evolution in plants. Int J Mol Sci. 2020;21:707. https://doi.org/10.3390/ijms21030707.
82. Kejnovsky E, Hobza R, Cermak T, Kubat Z, Vyskot B. The role of repetitive DNA in structure and evolution of sex chromosomes in plants. Heredity. 2009;102:533–41.
83. Li SF, Li JR, Wang J, Dong R, Jia KL, Zhu HW, et al. Cytogenetic and genomic organization analyses of Chloroplast DNA invasions in the nuclear genome of *Asparagus officinalis* L. provides signatures of evolutionary complexity and informativity in sex chromosome evolution. BMC Plant Biol. 2019;19:1–13.
84. Urasaki N, Tarora K, Shudo A, Ueno H, Tamaki M, Miyagi N et al. Digital transcriptome analysis of putative sex-determination genes in Papaya (*Carica papaya*). PLoS ONE. 2012;7:e40904. https://doi.org/10.1371/journal.pone.0040904.
85. Maughan PJ, Chaney L, Lightfoot DJ, Cox BJ, Tester M, Jellen EN, et al. Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (*Chenopodium quinoa* Willd). Sci Rep. 2019;9:1–11.
86. Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). Nucleic Acids Res. 2000;28:2571–6.
87. Cai X, Jiao C, Sun H, Wang X, Xu C, Fei Z, et al. The complete mitochondrial genome sequence of spinach, *Spinacia Oleracea* L. Mitochondrial DNA B Resour. 2017;2:339–40.
88. Smith DR. Mitochondrion-to-plastid DNA transfer: it happens. New Phytol. 2014;202:736–8.
89. Pinard D, Myburg AA, Mizrachi E. The plastid and mitochondrial genomes of *Eucalyptus grandis*. BMC Genomics. 2019;20:1–14.
90. Ye N, Wang X, Li J, Bi C, Xu Y, Wu D et al. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*. PeerJ. 2017;5:e3148. https://doi.org/10.7717/peerj.3148.
91. Gandini CL, Sanchez-Puerta MV. Foreign plastid sequences in plant mitochondria are frequently acquired via mitochondrion-to-mitochondrion horizontal transfer. Sci Rep. 2017;7:1–8.
92. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
93. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18:170–5. https://doi.org/10.1038/s41592-020-01056-5
94. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3:95–8.
95. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016;3:99–101.
96. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and samtools. Bioinformatics. 2009;25:2078–9.
97. Lemas JM, Patterson EL, Cutti L, Morran S, Johnson NA, Montgomery J, et al. Assembly and annotation of the tetraploid *Salsola tragus* (Russian Thistle) genome. Genome Biol Evol. 2025;17:1–6.
98. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117:9451–7.
99. Quinlan AR, Hall IM, BEDTools:. A flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
100. Lightfoot DJ, Jarvis DE, Ramaraj T, Lee R, Jellen EN, Maughan PJ. Single-molecule sequencing and Hi-C-based proximity-guided assembly of Amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. BMC Biol. 2017;15:1–15.
101. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
102. Consortium TU. UniProt: the universal protein knowledgebase in 2023 - Google scholar. Nucleic Acids Res. 2023;51:523–31.
103. Chan PP, Lin BY, Mak AJ, Lowe TM. TRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021;49:9077–96.
104. Tian F, Yang DC, Meng YQ, Jin J, Gao G, PlantRegMap. Charting functional regulatory maps in plants. Nucleic Acids Res. 2020;48:D1104–13.
105. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
106. Dainat J. Another Gtf/Gff Analysis Toolkit (AGAT): Resolve interoperability issues and accomplish more with your annotations. Plant and Animal Genome XXIX Conference. 2022.
107. Bushnell B. BBTools software package. 2014.
108. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:4–9.
109. Ellinghaus D, Kurtz S, Willhoeft U. *LTRharvest*, an efficient and flexible software for *de novo* detection of LTR retrotransposons. BMC Bioinformatics. 2008;9:18. https://doi.org/10.1186/1471-2105-9-18.
110. Ou S, Jiang N, LTR_FINDER_parallel. Parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. Mob DNA. 2019;10:1–3.
111. Ou S, Jiang N, LTR_retriever. A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. 2018;176:1410–22.
112. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.
113. Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR assembly index (LAI). Nucleic Acids Res. 2018;46:e126.
114. Vollger MR, Kerpedjiev P, Phillippy AM, Eichler EE. StainedGlass: interactive visualization of massive tandem repeat structures with identity heatmaps. Bioinformatics. 2022;38:2049–51.
115. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:1–9.
116. Wu C, Davis AS, Tranel PJ. Limited fitness costs of herbicide-resistance traits in *Amaranthus tuberculatus* facilitate resistance evolution. Pest Manag Sci. 2018;74:293–301.
117. Rochette NC, Rivera-Colón AG, Catchen JM. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol Ecol. 2019;28:4737–54.
118. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997v2. 2013;00:1–3.
119. Cerca J, Maurstad MF, Rochette NC, Rivera-Colón AG, Rayamajhi N, Catchen JM, et al. Removing the bad apples: A simple bioinformatic method to improve loci-recovery in *de novo* RADseq data for non-model organisms. Methods Ecol Evol. 2021;12:805–17.
120. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:1–16.
121. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. Gigascience. 2019;8:1–12.

122. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and vcftools. Bioinformatics. 2011;27:2156–8.
123. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35:1026–8.
124. Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res. 2008;18:1944–54.
125. Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A et al. JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome Biol. 2023;24:74. https://doi.org/10.1186/s13059-023-02914-z.
126. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:1–14.
127. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.
128. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37:4572–4.
129. Goel M, Sun H, Jiao WB, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 2019;20:1–13.
130. Cabanettes F, Klopp C. D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. PeerJ. 2018;2018.
131. Ma X, Vaistij FE, Li Y, Jansen van Rensburg WS, Harvey S, Bairu MW, et al. A chromosome-level *Amaranthus cruentus* genome assembly highlights gene family evolution and biosynthetic gene clusters that May underpin the nutritional value of this traditional crop. Plant J. 2021;107:613–28.
132. Raiyemo DA, Montgomery JS, Cutti L, Abdollahi F, Llaca V, Fengler K, et al. Chromosome-level assemblies of *Amaranthus palmeri*, *Amaranthus retroflexus*, and *Amaranthus hybridus* allow for genomic comparisons and identification of a sex-determining region. Plant J. 2025;121:1–17.
133. Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, et al. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. Elife. 2022;11:1–20.
134. Jeffares DC, Bartlomiej T, Sojo V, dos Reis M. A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. In: Peacock C, editor. Parasite genomics protocols methods in molecular biology. New York: Humana; 2014. pp. 1–365.
135. Álvarez-Carretero S, Kapli P, Yang Z. Beginner's guide on the use of PAML to detect positive selection. Mol Biol Evol. 2023;40:1–18.
136. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.
137. Yang Z, Nielsen R. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. Mol Biol Evol. 2008;25:568–79.
138. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J Roy Stat Soc. 1995;57:289–300.
139. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
140. Liao Y, Smyth GK, Shi W, FeatureCounts. An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
141. Robinson MD, McCarthy DJ, Smyth GK, edgeR. A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26:139–40.
142. Love MI, Huber W, Anders S. Moderated Estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:1–21.
143. Cantalapiedra CP, Herŋandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38:5825–9.
144. McGrath JM, Funk A, Galewski P, Ou S, Townsend B, Davenport K et al. A contiguous *de Novo* genome assembly of sugar beet EL10 (*Beta vulgaris* L). DNA Res. 2023;30.
145. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:29–37.
146. Hulse-Kemp AM, Bostan H, Chen S, Ashrafi H, Stoffel K, Sanseverino W et al. An anchored chromosome-scale genome assembly of spinach improves annotation and reveals extensive gene rearrangements in euasterids. Plant Genome. 2021;14.
147. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J et al. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. 2012;40.
148. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
149. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52:W78–82.
150. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31:1296–7.
151. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq - Versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45:W6–11.
152. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59–64.
153. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: A web server for microsatellite prediction. Bioinformatics. 2017;33:2583–5.
154. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R, REPuter. The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29:4633–42.
155. R Core Team. R: A Language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.