the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2025) **121,** e70027 doi: 10.1111/tpj.70027

RESOURCE

Chromosome-level assemblies of *Amaranthus palmeri*, *Amaranthus retroflexus*, and *Amaranthus hybridus* allow for genomic comparisons and identification of a sex-determining region

Damilola A. Raiyemo[1,†], Jacob S. Montgomery[2,3,†], Luan Cutti[3], Fatemeh Abdollahi[2], Victor Llaca[4], Kevin Fengler[4], Alexander J. Lopez[1], Sarah Morran[2], Christopher A. Saski[5], David R. Nelson[6], Eric L. Patterson[3], Todd A. Gaines[2] and Patrick J. Tranel[1,*]

[1]*Department of Crop Sciences, University of Illinois, Urbana, Illinois, USA,*
[2]*Department of Agricultural Biology, Colorado State University, Fort Collins, Colorado, USA,*
[3]*Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, Michigan, USA,*
[4]*Genome Center of Excellence, Corteva Agriscience, Johnston, Iowa, USA,*
[5]*Department of Plant and Environmental Sciences, Clemson University, Clemson, South Carolina, USA, and*
[6]*Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, Memphis, Tennessee, USA*

Received 1 October 2024; revised 17 December 2024; accepted 24 January 2025.
*For correspondence (e-mail tranel@illinois.edu).
†These authors contributed equally to this work.

SUMMARY

Amaranthus palmeri **(Palmer amaranth),** ***Amaranthus retroflexus*** **(redroot pigweed), and** ***Amaranthus hybridus*** **(smooth pigweed) are troublesome weeds that are economically damaging to several cropping systems. Collectively referred to as "pigweeds," these species are incredibly adaptive and have become successful competitors in diverse agricultural settings. The development of genomic resources for these species promises to facilitate the elucidation of the genetic basis of traits such as biotic and abiotic stress tolerance (e.g., herbicide resistance) and sex determination. Here, we sequenced and assembled chromosome-level genomes of these three pigweeds. By combining the haplotype-resolved assembly of** ***A. palmeri*** **with existing restriction site-associated DNA sequencing data, we identified an approximately 2.84 Mb region on chromosome 3 of Hap1 that is male-specific and contains 37 genes. Transcriptomic analysis revealed that two genes,** ***RESTORER OF FERTILITY 1*** **(*RF1*) and** ***TLC DOMAIN-CONTAINING PROTEIN*** **(*TLC*), within the male-specific region were upregulated in male individuals across the shoot apical meristem, the floral meristem, and mature flowers, indicating their potential involvement in sex determination in** ***A. palmeri*****. In addition, we rigorously classified cytochrome P450 genes in all three pigweeds due to their involvement in non-target-site herbicide resistance. Finally, we identified contiguous extrachromosomal circular DNA (eccDNA) in** ***A. palmeri*****, a critical component of glyphosate resistance in this species. The findings of this study advance our understanding of sex determination in** ***A. palmeri*** **and provide genomic resources for elucidating the genetic basis and evolutionary origins of adaptive traits within the genus.**

Keywords: ***Amaranthus palmeri*****,** ***Amaranthus retroflexus*****,** ***Amaranthus hybridus*****, genomics, dioecy, weed science, herbicide resistance, 5-enolpyruvylshikimate 3-phosphate synthase, cytochrome P450.**

INTRODUCTION

Amaranthus palmeri S. Watson (Palmer amaranth), *Amaranthus retroflexus* L. (redroot pigweed), and *Amaranthus hybridus* L. (smooth pigweed) are troublesome summer annual weeds of several cropping systems (Sellers et al., 2003). While *A. retroflexus* and *A. hybridus* are monoecious and primarily self-pollinated, *A. palmeri* is a dioecious (i.e., with separate male and female plants),

obligate-outcrossing species that has now spread beyond its native origins in the American Southwest to many parts of the United States and the world (Steckel, 2007; Trucco et al., 2007). *A. palmeri* has now been reported in 45 countries, and climate models indicate that it will expand to most temperate and subtropical areas of the United States, Africa, Australia, South America, Eastern Asia, and Europe by 2050 (Kistner & Hatfield, 2018; Roberts & Florentine, 2022). Several studies have documented significant yield losses that occur from the interference of weedy amaranths (collectively referred to as "pigweeds") with diverse crops such as corn, soybean, and cotton (Bensch et al., 2003; Hager et al., 2002; Manalil et al., 2017; Massinga et al., 2001; Morgan et al., 2001).

The notoriety of the pigweeds as damaging to agricultural crops has spurred the exploration of several options for their control (e.g., cover crops, herbicide rotations, mechanical seed destruction, flood irrigation management, tillage, and grazing). A meiotic gene drive of a sex-determining genetic factor has been proposed as a radically new method for the suppression of dioecious weed species such as *A. palmeri* (Barrett et al., 2019; Legros et al., 2021; Neve, 2018; Rode et al., 2019). To utilize such a weed control strategy, however, requires understanding the mechanism of sex determination in the species. Previous studies on sex determination in amaranths revealed that males are the heterogametic sex with an XY system, which was subsequently confirmed using restriction-site associated DNA (RAD) sequencing (Grant, 1959; Montgomery et al., 2019; Murray, 1940; Neves et al., 2020). An approximately 1.3 Mb male-specific Y region (MSY) that contained 121 gene models was identified in a draft genome assembly of a male *A. palmeri* plant (Montgomery et al., 2020, 2021). The draft assembly was also combined with short reads of *Amaranthus watsonii* to detect the male-specificity of a copy of a gene encoding a pentatricopeptide repeat-containing protein (PPR) within the MSY region on scaffold 20 (Raiyemo et al., 2023). Similarly, transcriptomic analyses between several tissue types of male and female individuals revealed the upregulation of *PPR247* in males, a gene within the MSY region (Bobadilla et al., 2023). The male-specific expression of this gene across three tissue types led Bobadilla et al. (2023) to hypothesize a single-gene model of sex determination for *A. palmeri*. Despite the advances above, the genomic landscape of sex chromosomes and how they evolved for the dioecious *Amaranthus* species are still poorly understood. For example, inversions, centromeric location of sex-determining regions (SDRs), or hemizygosity have been reported as possible routes of recombination suppression that lead to the evolution of sex chromosomes in some dioecious species (Carey et al., 2022). In fact, we recently reported the identification of an SDR in the closely related *Amaranthus tuberculatus*, near a chromosomal fusion and large inversion event (Raiyemo et al., 2024).

The evolution and spread of herbicide resistance in pigweeds is now a major concern for crop production. Non-target-site herbicide resistance (NTSR), specifically metabolism-based resistance, is particularly alarming due to its genetic complexity and the ability of one resistance allele to give cross-resistance to diverse herbicides (Rigon et al., 2020; Yu & Powles, 2014). Cytochrome P450 (CYP) genes are involved in enhanced herbicide metabolism and are known to mediate the detoxification of herbicides by adding a functional group to the herbicide molecule via oxidation, reduction, or hydrolysis, thus rendering the herbicide less effective (Dimaano & Iwakami, 2021; Gaines et al., 2020). In fact, CYPs have been implicated in metabolism-based herbicide resistance of some *A. tuberculatus* and *A. palmeri* biotypes (De Figueiredo et al., 2022; Oliveira et al., 2018; Rigon et al., 2024; Varanasi et al., 2018). Understanding the structure–activity relationship between CYP enzymes and the herbicides they metabolize via biochemical modeling could help predict herbicide classes that could be degraded even before they are synthesized (Bobadilla & Tranel, 2024; Gaines et al., 2020). Development of genome assemblies and careful annotation of these *CYPs* for weed species of interest provides a library of herbicide metabolism genes for systematic approaches to such investigations.

Furthermore, the identification of gene duplication of *5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE* (*EPSPS*) on extrachromosomal circular DNA (eccDNA) conferring resistance to the herbicide glyphosate in *A. palmeri* was a surprising and landmark discovery for weed genomics (Gaines et al., 2010; Koo et al., 2018; Molin, Yaguchi, et al., 2020). Since this discovery in *A. palmeri*, the role of eccDNAs in herbicide resistance, their variation within and among species, and how they evolved have become open areas of research in plants (Camposano et al., 2022; Fu et al., 2023). It has been suggested that the eccDNAs that contain *EPSPS* in *A. palmeri* are diverse with multiple evolutionary events (Camposano et al., 2022), but the source of eccDNAs is not well understood. A better understanding of the evolutionary origins of eccDNAs and how they amplify genes will shed light on this surprising mechanism of gene duplication and determine the necessity of resistance-containment protocols.

Toward furthering our understanding of the biology of weed species including pigweeds, the International Weed Genomics Consortium (IWGC) was initiated to provide genomic resources for the science community (Montgomery et al., 2024). As part of this effort, we produced chromosome-level genome assemblies of *A. palmeri* (male; haplotype-resolved), *A. retroflexus*, and *A. hybridus* individuals utilizing PacBio long-read sequencing, Hi-C scaffolding, and Bionano optical mapping. Utilizing these genome assemblies, we identified a contiguous male-specific region and candidate genes for sex determination

Table 1 Comparison of assembly statistics between the two haplomes of *A. palmeri*, *A. retroflexus*, and *A. hybridus* genome assemblies

	A. palmeri			
Genome characteristics	Haplome 1	Haplome 2	*A. retroflexus*	*A. hybridus*
Contig assembly size (Mbp)	369.1	354.5	439.5	403.0
Contig N50	6.82	5.71	6.65	2.26
Scaffold assembly size (Mbp)	384.1	364.5	446.79	437.99
Scaffold N50 (Mbp)	23.59	22.01	24.31	27.31
Scaffold L50	8	8	8	7
GC content (%)	33.4	33.4	33.0	33.0
Complete BUSCO (%)	97.7	97.3	97.3	97.3
LTR assembly index (LAI)	18.25	21.85	10.78	15.29
Protein-coding genes	24 873	24 791	27 377	22 771
Mean gene length (bp)	4637	4633	4623	4874
Mean CDS length (bp)	1178	1174	1092	1184
Mean exon length (bp)	273	272	298	270
Mean exon per gene	5.5	5.5	5.2	5.5
Number of tRNA	1480	1508	1499	1535
Total annotated genes	26 353	26 299	28 876	24 306

in *A. palmeri*, characterized the CYP gene family in the three pigweeds, and identified the conservation of eccDNA sequence across diverse accessions of *A. palmeri*. The genome assemblies reported here along with existing *Amaranthus* genomes promise to improve weed management through the development of new weed control technologies as well as crop improvement through the identification of the genetic basis of stress tolerance in these prolific species.

RESULTS

Genome assembly, annotation, and repeat analyses

Statistics summarizing the raw sequence data used in genome assembly and annotation are listed in Table S1. Evaluation of assembly completeness for *A. palmeri* haplome 1 (Hap1), *A. palmeri* haplome 2 (Hap2), *A. retroflexus*, and *A. hybridus* genomes using BUSCO and LTR assembly index (LAI) revealed highly complete assemblies for all species (Table 1). The total number of predicted protein-coding genes for the assemblies ranged from 22 771 genes for *A. hybridus* to 27 377 genes for *A. retroflexus* (Table 1; Tables S2 and S3). While the total gene number is lower than that of some other diploids, these annotations are consistent with previous annotations of *Amaranthus* genomes, including *A. palmeri* and *A. hybridus* (Lightfoot et al., 2017; Ma et al., 2021; Montgomery et al., 2021; Wang et al., 2023). Analysis of repetitive elements revealed over half of each genome assembly was made up of transposable elements with LTR/*Copia* elements being the most abundant (approximately 10% of each genome) (Table S4). Analysis of centromeric repeats indicates chromosomes have varying types of centromeres. For example, chromosomes 1 and 5 appear telocentric while chromosomes 2 and 3 appear metacentric or submetacentric for *A. palmeri* (Table S5). A search of simple telomeric repeat against the assemblies revealed telomeric repeat sequences at approximately 75% of chromosome ends for *A. retroflexus* and both haplotypes of *A. palmeri*, but just 25% of chromosome ends for *A. hybridus* (Table S6). Localization of genomic features depicted in Figure 1 indicates that regions abundant in gene content are poor in LTR retrotransposons (LTR-RT), and vice-versa.

Intraspecific comparisons of genome structure

Syntenic analysis of homoeologous relationships among the 16 or 17 chromosomes for each of the three *Amaranthus* genome assemblies was carried out using MCScanX. Our analysis revealed collinear blocks of genes in one half of chromosome 1 were syntenic to collinear blocks of genes in the other half of the chromosome (Figure 1), consistent with previous observations in *Amaranthus hypochondriacus* (Lightfoot et al., 2017) and *Amaranthus cruentus* (Ma et al., 2021), and thus seems to be conserved across species in the subgenera *Acnida* and *Amaranthus*. Chromosome 1 has been suggested to have evolved from the fusion of two homoeologs following the most recent whole-genome duplication that occurred in the amaranths between 18 and 34 million years ago (MYA) (Lightfoot et al., 2017; Ma et al., 2021). Our analysis also revealed very minimal synteny between other chromosomes and chromosome 7 in *A. palmeri* and *A. retroflexus* (or chromosome 5 in *A. hybridus*, *A. cruentus*, and *A. hypochondriacus*), supporting previous evidence that the chromosome had lost its homoeologous copy during evolution (Ma et al., 2021). Six chromosomes in *A. palmeri* (Chr2, Chr3, Chr4, Chr5, Chr8, and Chr10), three chromosomes in *A. retroflexus* (Chr2, Chr5, and Chr8) and four

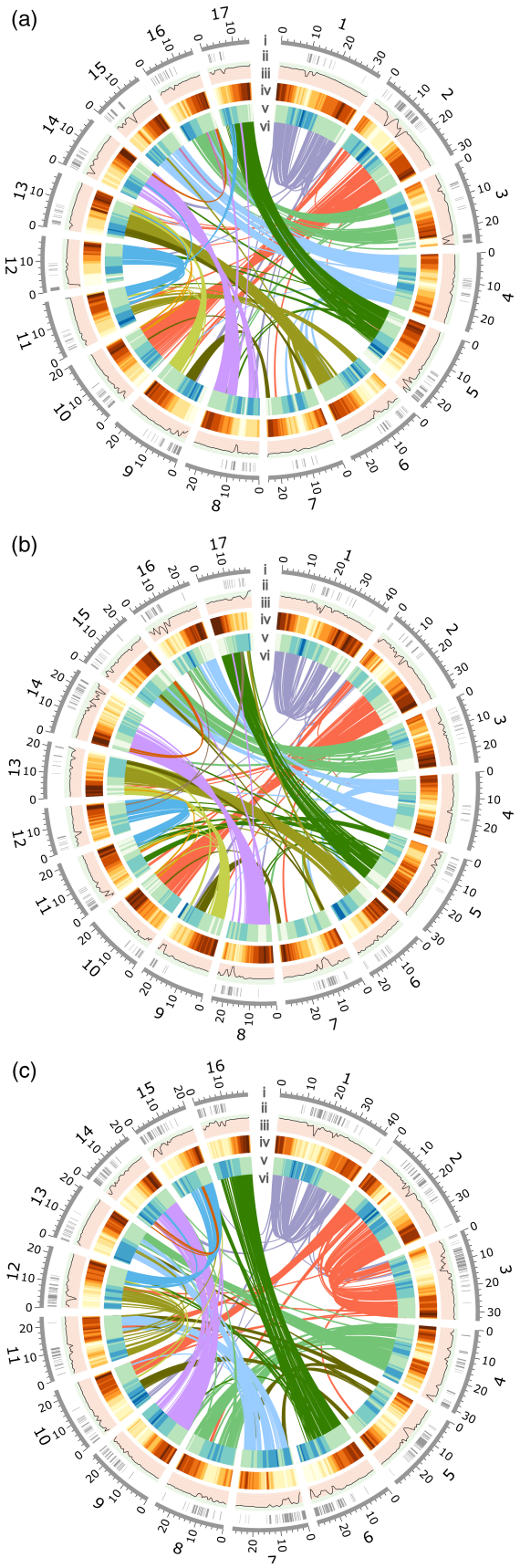

Figure 1. Genomic features.
(a) *Amaranthus palmeri* Hap1.
(b) *A. retroflexus*.
(c) *A. hybridus* assemblies. Circos plot depicts (i) chromosome number and length (Mb), (ii) distribution of gaps in the genomes, (iii) GC content along the chromosomes, with peaks in the light green area representing GC content greater than the median and peaks in the light red area representing GC content less than the median (median GC contents: approximately 0.32), (iv) gene density across the chromosomes, with brown representing gene-rich regions and yellow representing gene-poor regions, (v) Long terminal repeats (LTR) density along chromosomes, with blue representing LTR-rich regions and green representing LTR-poor regions, and (vi) inner ribbons represent duplicated genes, and ribbons are colored to represent syntenic chromosomes across the three genomes. Window size of 1 Mb and step size of 500 Kbp for (ii–iv).

chromosomes in *A. hybridus* (Chr2, Chr4, Chr6, and Chr9) had discernable one-to-one homeologous relationships (i.e., Chr2–Chr10; Chr3–Chr16; Chr4–Chr15; Chr5–Chr17; Chr8–Chr14 for *A. palmeri*, Chr2–Chr10; Chr5–Chr17, Chr8–Chr14 for *A. retroflexus*, and Chr2–Chr3; Chr4–Chr8, Chr6–Chr16, Chr9–Chr14 for *A. hybridus*) (Figure 1).

Analysis of structural rearrangements among the three *Amaranthus* genome assemblies from this study, *A. tuberculatus*, and three monoecious amaranths indicated a highly conserved gene order, except for a few chromosomes (Figure 2). For example, chromosome 4 in *A. hypochondriacus*, *A. hybridus*, and *A. cruentus* appears to have been derived from the fusion of two ancestral chromosomes (Figure 2). Also, chromosome 1 (or chromosome 2 in *A. tuberculatus*) in species within the two subgenera (i.e., *Acnida* and *Amaranthus*) to which the three genomes belong appears to have originated from the fusion of parts of chromosomes 1 and 2 in *Amaranthus tricolor*, which belongs to the subgenus *Albersia* (Figure 2).

Identification of the SDR in *A. palmeri*

Individuals with more than 60% missing data were removed from a previously produced RAD-seq dataset, which retained 52 female samples and 53 male samples from across six geographically diverse accessions. The filtering improved the quality of the data (i.e., 24.57% missing data across the 105 individuals). The individuals were then mapped to Hap1 of the genome assembly, and 234 066 variants retained across the individuals were used for genome-wide association (GWA) analysis. The analysis revealed four significant SNPs above the Bonferroni threshold that are associated with sex; two on chromosome 3 (24 609 596 bp, $P = 1.7481e\text{-}20$; 24 595 358 bp, $P = 6.8453e\text{-}17$), and one each on chromosome 4 (13 423 105 bp, $P = 1.0254e\text{-}08$), and chromosome 16 (7 114 501 bp, $P = 1.0704e\text{-}17$) (Figure 3a). All significant SNPs were within intergenic regions, and there was no evidence of systematic bias in the GWA analysis based on the QQ plot (Figure 3b). Further analysis of genetic differentiation between females and males using fixation index (F_{ST})

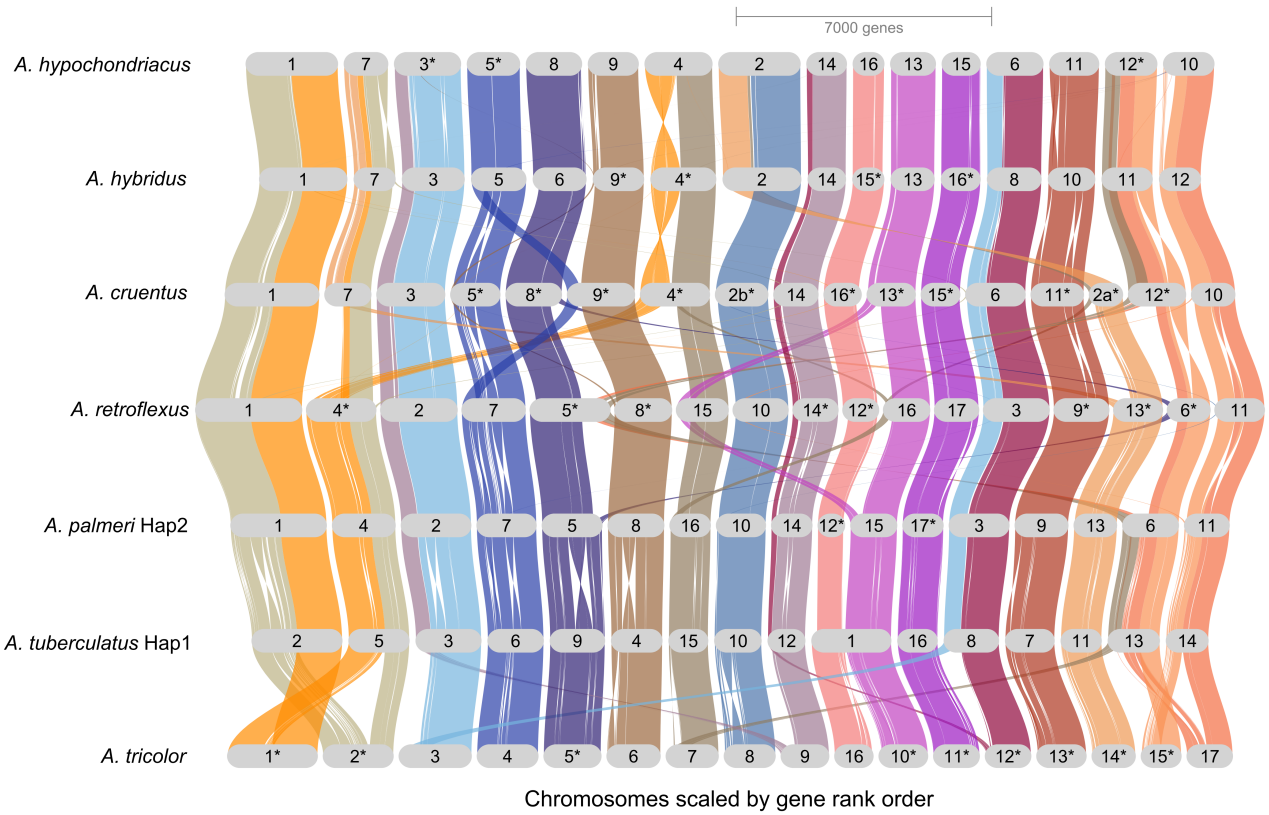

Figure 2. Synteny plot between the haplotype assemblies of four previously reported chromosome-level *Amaranthus* spp. assemblies and three newly assembled ones (*A. palmeri*, *A. retroflexus*, and *A. hybridus*).
Asterisks (*) indicate chromosomes that were manually inverted to keep the gene order consistent with *A. tuberculatus*. Species are ordered to reflect phylogenetic relationships from STAG in OrthoFinder, which also corresponds to the species tree in Wang et al. (2023).

also points to the end of chromosome 3 as being highly differentiated between female and male individuals of *A. palmeri* (Figure 3c,d). While multiple lines of evidence point to the distal end of chromosome 3 as the SDR, we note that the significant association of the SNP on chromosome 16 with sex is due to duplication, which is evident in our synteny analysis indicating that chromosomes 3 and 16 are ancient homoeologs. We consider the significant association of the SNP on chromosome 4 with sex a spurious association. Several factors including linkage disequilibrium with causative SNPs, gene duplication, chromosomal rearrangements, or population stratification have been noted to cause spurious associations whereby SNPs that are not causative SNPs are significantly associated with a trait of interest (Clauw et al., 2024).

Analysis of synteny between the haplotype assemblies indicated a mostly one-to-one relationship in gene content between the haplotypes (Figure S1a,b). Identification of syntenic orthologs revealed 21 421 gene pairs are reciprocal best-hit matches between the two haplotypes. A region on chromosome 3 between 22 994 134 and 25 836 006 (approximately 2.84 Mb) is hemizygous and present near the end of Hap1 but absent from Hap2. This chromosome was assembled into one hybrid scaffold with 18 contigs and 3 Bionano maps in Hap1 and 12 contigs and 3 Bionano maps in Hap2 (Figure S2). In the absence of this approximately 2.84 Mb region, Hap1 and Hap2 were highly syntenic (Figure 3e). Further analysis revealed 1 857 183 bp (65.35%) of this hemizygous region of Hap1 is made up of Ns. While the sequence is not represented, the length of gaps is known because Bionano maps estimate the physical distance between markers at well-assembled loci. This region is well supported by several hundred molecules that are rich with label sites (Figure S3), so the size of gaps is likely accurate. There are 37 genes within this hemizygous region out of 1620 protein-coding genes on chromosome 3 of Hap1 (Table S7). Analysis of structural rearrangements indicated inversions were also present on chromosome 3. They were, however, less than 260 Kbp; shorter than other inversions on chromosomes 4, 7, 14, or 17 that were greater than 1 Mb (Table S8). Given the location of the SDR on chromosome 3, we designate Hap1 as the assembly containing the Y chromosome while Hap2 contains the X chromosome.

Query of a primer set (PAMS-940), used in Montgomery et al. (2019) as a male-specific marker, against both

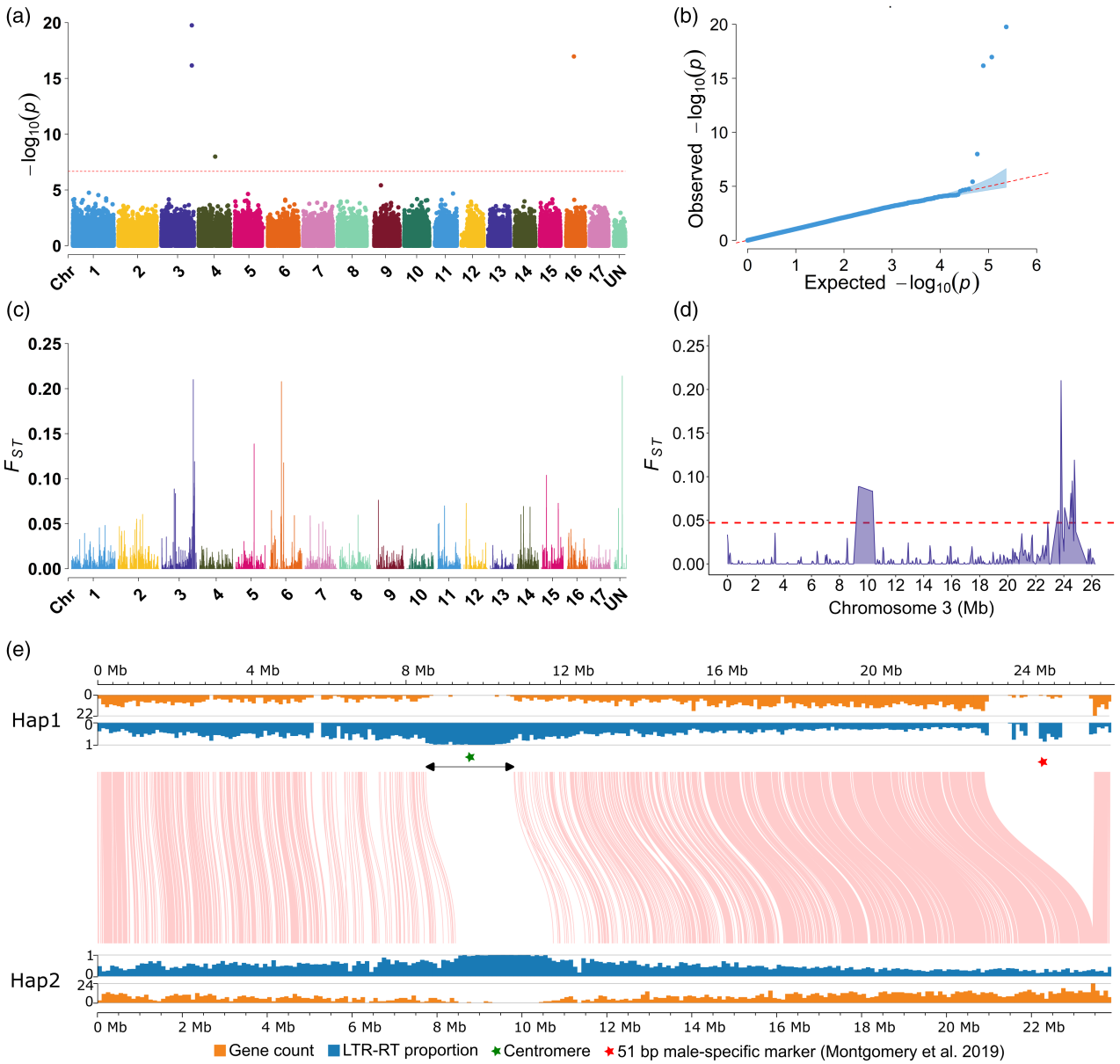

Figure 3. Identification of *Amaranthus palmeri* sex-determining region on chromosome 3.
(a) Manhattan plot of GWA analysis using RAD-seq data from 52 females and 53 males. The dashed red line indicates the Bonferroni threshold of $-\log_{10}(P) = 6.6704$.
(b) Quantile-quantile (QQ) plot of the GWA analysis.
(c) Fixation index (F_{ST}) between females and males across all 17 chromosomes (window size 100 Kbp; step size 50 Kbp).
(d) F_{ST} between females and males for chromosome 3 with a dashed red line representing the top 1% threshold at 0.0472 (window size 100 Kbp; step size 50 Kbp).
(e) Synteny plot showing collinear regions on chromosome 3. The green asterisk represents the centromere while the red asterisk represents a previously reported 51 bp male-specific marker (Montgomery et al., 2019) that matched a region on Hap1. Gene count and LTR-RT proportion were calculated based on 100 Kbp non-overlapping windows.

haplotype assemblies of *A. palmeri* revealed a perfect match to chromosome 3 of Hap1 with a predicted amplicon size of 51 bp (24 464 780–24 464 831 bp) (Figure 3e). This size is consistent with the amplified product size reported in Montgomery et al. (2019). The forward and reverse primer sequence for JM940 from Montgomery et al. (2021) also matched chromosome 3 (24 755 051–24 755 052), but two single nucleotide polymorphisms exist between the forward primer and the genomic sequence. Sequence alignment of the previous draft *A. palmeri* male assembly to both haplotype assemblies identified large structural differences between

assemblies (Figure S4a,b). This variation could be due to real differences in genome structure but also could be due to errors in assembly or scaffolding. Scaffolds 5 (previously proposed as containing the pseudoautosomal region) and 20 (containing a previously proposed male-specific region) are both aligned to chromosome 3 in the two haplotype assemblies. As expected, the region identified as male-specific on scaffold 20 only aligned to the hemizygous region on chromosome 3 of Hap1 (Figure S4a–f). Reciprocal best hit searches between genes on scaffold 20 of the draft assembly and those in the haplotype assemblies revealed 41 and 37 genes had matches on chromosome 3 of Hap1 and Hap2 assemblies, respectively (Table S9). However, only four genes (coding for three unknown proteins and HEADING DATE 3A) that matched between the draft assembly and Hap1 were within the male-specific region in Hap1 (Table S9). Taken together, we consider chromosome 3 as the likely sex chromosome candidate in our assembly and the approximately 2.84 Mb hemizygous region as the likely SDR.

Transcriptome profiling, GO term enrichment, and coverage analyses

A previous mRNA dataset reported by Bobadilla et al. (2023) was mapped to the Hap1 assembly of *A. palmeri*. Uniquely mapped reads ranged from 76.40 to 91.61% across male and female samples for shoot apical meristem, floral meristem, and mature flower tissues. Out of the 24 873 annotated Hap1 protein-coding genes, 20 046 genes were retained for differential expression (DE) analysis after filtering and trimmed mean of *M* values (TMM) normalization (Figure S5a,b). Eight genes were differentially expressed between males and females in the shoot apical meristem, while 29 genes were differentially expressed for floral meristem, and 2595 genes for mature flower (Figure S5a,b). Only five genes (encoding a serine/threonine-protein kinase EDR1-like protein, two RF1 proteins, an unknown protein, and a TLC domain-containing protein) were differentially expressed between male and female individuals across the three tissue types, and all mapped to chromosome 3 (Tables S10–S12). A search of the unknown protein in the NCBI nonredundant protein database using BLASTP revealed 71% homology to a wall-associated receptor kinase-like protein in several species. Two of these five genes, encoding RF1 (a pentatricopeptide repeat-containing restorer of male fertility gene, AmaPaChr03Ag063970) or *TLC* (AmaPaChr03Ag063980), are within the MSY region on chromosome 3 (Tables S10–S12). A further inspection of the raw counts of mRNA reads mapped to the 37 genes within the MSY region revealed over 60% of the genes had no transcripts mapped to them (Table S7). Gene ontology (GO) term enrichment analysis, with mature flower DEGs selected based on a false discovery rate threshold of $P < 0.05$ and $Log_2FoldChange > 1.2$, revealed the list of DEGs was predictably enriched for genes involved in pollen and anther development (Figure S5c; Tables S13–S16).

To determine the conservation of these two genes with male-specific expression in our predicted SDR, we mapped genomic sequence data of *A. watsonii*, a species closely related to *A. palmeri* (Raiyemo et al., 2023; Raiyemo & Tranel, 2023), to the Hap1 genome assembly of *A. palmeri*. Coverage analysis revealed that while the *TLC* gene that was upregulated in males and within the MSY region has male-biased coverage in *A. watsonii* (Figure 4a), the *RF1* gene that was also upregulated in males and within the MSY region appears to be male-specific in *A. watsonii* (Figure 4b). After filtering to retain only uniquely mapped reads, the *TLC* gene no longer had any reads aligning from male or female *A. watsonii* plants while the *RF1* gene still had male-specific coverage for the longest exon (Figure S6). Similarly, a gene encoding a different pentatricopeptide repeat-containing protein (AmaPaChr03gA063950), although not differentially expressed between males and females, was male-specific in *A. watsonii* even after filtering for only uniquely mapped reads (Figure S7). Other genes of unknown function within the MSY of *A. palmeri* also showed male-enriched coverages in *A. watsonii* (Figures S7 and S8). Given their location in the SDR region, male-specific expression, and conservation across species, we hypothesize that *RF1* and *TLC* are necessary for maleness and/or suppression of femaleness in *A. palmeri* and *A. watsonii*.

Identification and comparative analyses of *Amaranthus* cytochrome P450 gene superfamily

A total of 184, 214, and 228 cytochrome P450 (CYP) gene models were identified in *A. palmeri*, *A. retroflexus*, and *A. hybridus*, respectively, and named according to the Standardized Cytochrome Nomenclature Committee (http://drnelson.uthsc.edu/CytochromeP450.html) (Tables S17–S19). Among these gene models, 130 from *A. palmeri*, 163 from *A. retroflexus*, and 152 from *A. hybridus* were determined to be full-length (ranging from 350 to 700 amino acids) while the rest (less than 350 amino acids) were designated as fragments and discarded from the analysis (Tables S17–S19). The full-length protein sequences for all species are provided in Table S20.

A neighbor-joining (NJ) tree was constructed from protein sequence to investigate the evolutionary relationship among the full-length CYP genes (Figure S9a,b). Genes were thus categorized into two groups: A-type, with a single clan (71) and non-A-type with multiple clans including 710, 85, 711, 86, 97, 72, 727, and 74. Some clans (74, 97, 710, 711, and 727) contained only one gene family each. In contrast, other clans, namely 71, 72, 85, and 86, included multiple families (Figure S9a,b). In total, 50% of CYP genes in *A. palmeri* (65/130), 45.3% in *A. retroflexus*

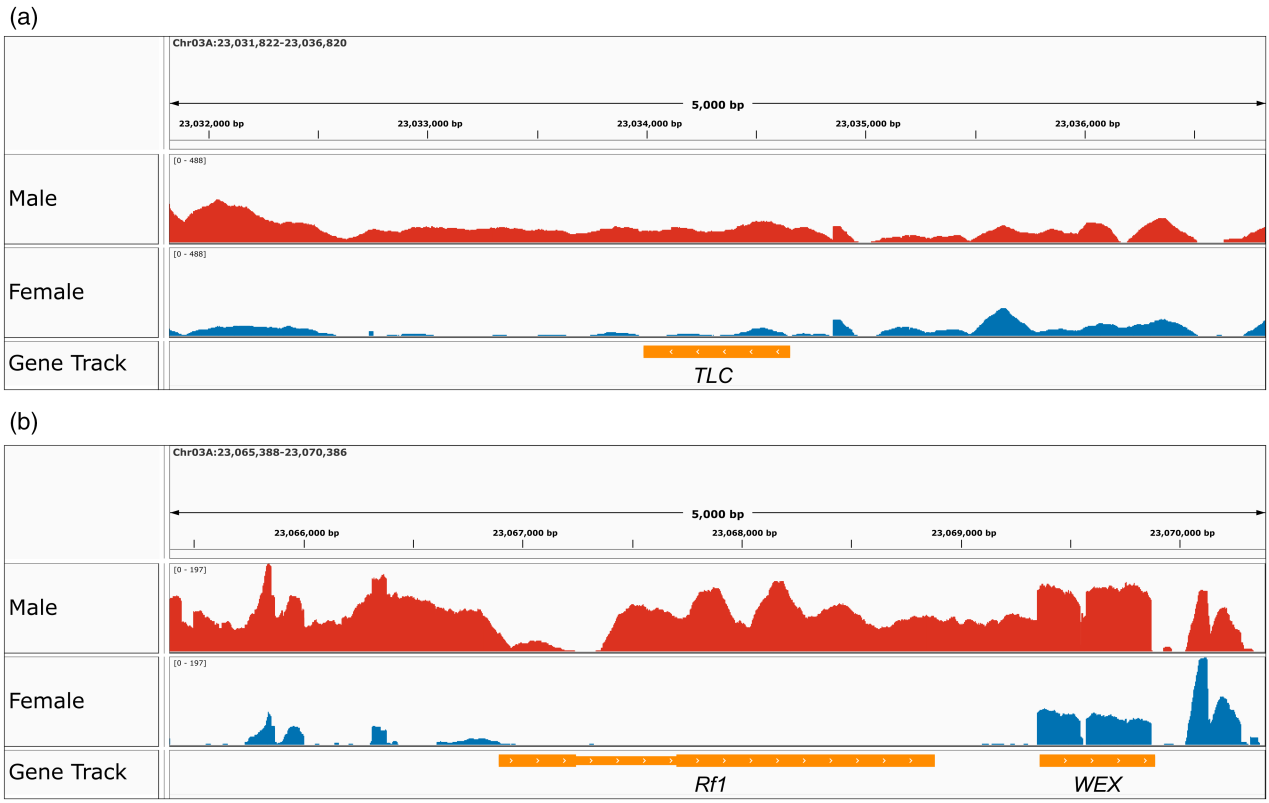

Figure 4. Male-specific gene coverage analysis. Reads of male and female *Amaranthus watsonii* mapped to *A. palmeri* Hap1 assembly and visualized in IGV.
(a) TLC protein (AmaPaChr03Ag063980).
(b) RF1 protein (AmaPaChr03Ag063970).

(74/163), and 44.5% in *A. hybridus* (81/182) belonged to clan 71, which emerged as the largest clan and includes 17 families (CYP71, CYP81-82, CYP92-93, CYP75-79, CYP84, CYP701, CYP703, CYP706, CYP89, CYP98, and CYP736). Here, clans are higher-level groupings of CYP genes that contain several families, whereas a CYP family is a more specific classification within a clan that is based on higher sequence identity (at least 40%) and closer evolutionary relationships.

Collinearity analysis was also conducted to explore the evolutionary relationships among the CYP gene family of the three species. We found a total of 106 collinear gene pairs between *A. palmeri* and *A. hybridus* (82% of *A. palmeri* CYPs), as well as 95 collinear gene pairs between *A. palmeri* and *A. retrofloxus* (73% of *A. palmeri* CYPs) (Figure S9c). These results highlight the significant evolutionary conservation present within the CYP genes among *A. palmeri*, *A. retroflexus*, and *A. hybridus*.

Origin of extrachromosomal circular DNA (eccDNA) and gene amplification in *A. palmeri*

To understand the genomic origins of eccDNA containing the gene *EPSPS* and conferring glyphosate resistance, the complete assembly of a 399 Kbp eccDNA molecule reported by Molin, Patterson, and Saski (2020) and Molin, Yaguchi, et al. (2020) was aligned to the reference genome (Hap1) of *A. palmeri*, which was obtained from a glyphosate-susceptible individual. Relatively stringent filtering (map quality = 60 and alignment length >1000 bp) revealed few syntenic regions between the canonical *EPSPS*-containing eccDNA molecule and the genome (Figure S10a). The longest stretch of sequence similarity was the region containing *EPSPS* (approximately 10 Kbp). However, relaxing the filtration of alignments (95% similarity, minimum alignment length >30 bp) resulted in many matches across the genome (Figure S10b).

To understand the evolutionary history of *EPSPS* amplified via eccDNAs, geographically diverse glyphosate-susceptible and glyphosate-resistant samples of *A. palmeri* were sequenced with PacBio HiFi and their genomes were assembled. When the assemblies of the glyphosate-sensitive individuals were aligned to the canonical eccDNA molecule, the assemblies had few, short alignments compared to glyphosate-resistant individuals (Tables S21–S28). This suggests the glyphosate-sensitive samples do not have an "empty" version of the eccDNA molecule that is simply missing *EPSPS*. In contrast, glyphosate-resistant samples had contigs that aligned

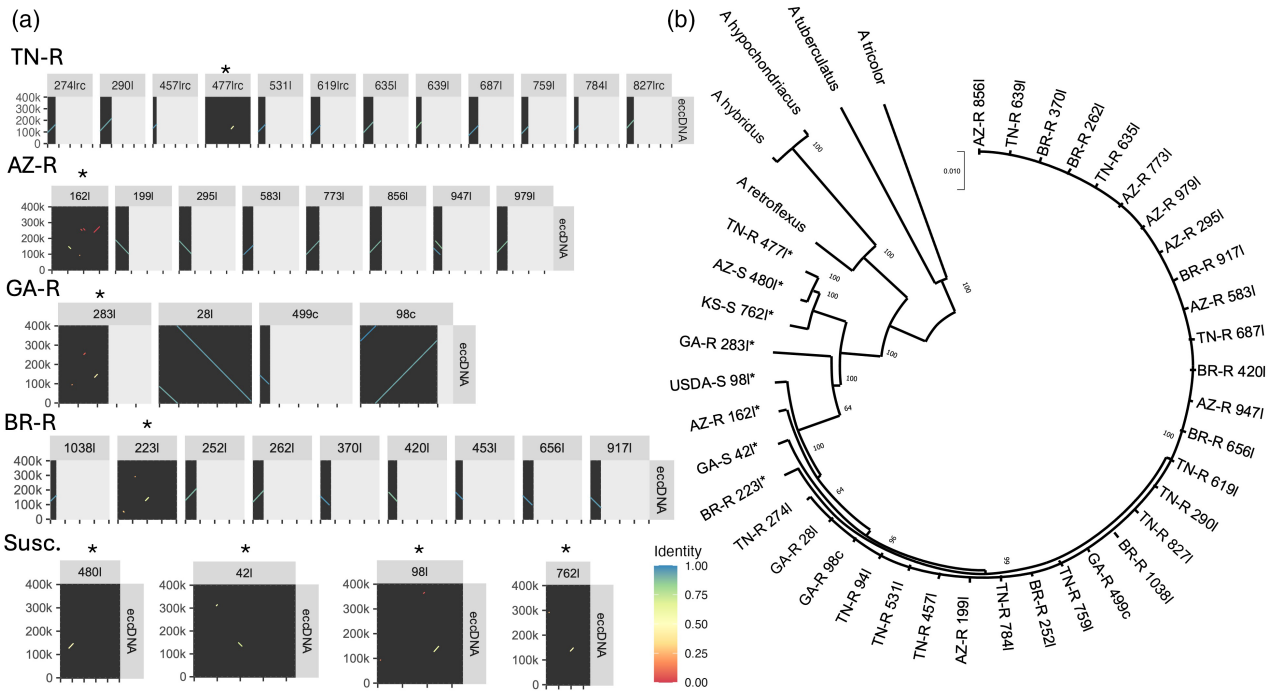

Figure 5. Comparison of extra-chromosomal circular DNA (eccDNA) containing *EPSPS* gene models from *Amaranthus palmeri*.
(a) Dotplots showing synteny between the canonical eccDNA molecule reported by Molin, Patterson, and Saski (2020) and Molin, Yaguchi, et al. (2020) and assembled contigs containing *EPSPS* gene models (located near 140Kbp on the eccDNA) from diverse *A. palmeri* accessions. Each row contains contigs from a different individual, except the last row contains the contigs from all four glyphosate-susceptible (Susc.) samples.
(b) Maximum likelihood phylogenetic tree estimating the evolution of *EPSPS* genes within contigs of *de novo* assemblies of each resequenced *A. palmeri* accession. *EPSPS* sequences from other *Amaranthus* species were used as out groups. Values at each branch indicate the number of bootstrap iterations that generated that branch (n = 100). Names of each branch indicate species or *A. palmeri* accession, contig number, and whether that contig was circularized or linear (l = linear, c = circular, rc = reverse complement). "*" Indicates contigs that are part of the nuclear genome. Arizona-resistant (AZ-R), Arizona-susceptible (AZ-S), Brazil-resistant (BR-R), Georgia-resistant (GA-R), Georgia-susceptible (GA-S), USDA-susceptible (USDA-S), Tennessee-resistant (TN-R), and Kansas-susceptible (KS-S).

across the entirety of the canonical eccDNA molecule, in line with findings from Molin, Patterson, and Saski (2020).

After collapsing redundant contigs, full *EPSPS* gene models were only found on one contig for each glyphosate-susceptible sample. Assemblies of glyphosate-resistant samples retained between 4 and 13 contigs containing *EPSPS* gene models. The alignment of these *EPSPS*-containing contigs to the canonical *EPSPS*-containing eccDNA sequence showed complete synteny between the contigs and the eccDNA sequence (Figure 5a). While most contigs were only fragments of the canonical eccDNA, a 395 Kbp circular contig containing *EPSPS* was recovered from the assembly of a glyphosate-resistant plant from Georgia. Though its length is similar to the eccDNA molecule reported by Molin, Patterson, and Saski (2020) and Molin, Yaguchi, et al. (2020), there are several large insertions (up to 3 Kbp) and deletions (up to 8 Kbp) in the *de novo* assembled version and very few single nucleotide polymorphisms. These insertions and deletions within the eccDNA sequence are mainly found in regions of simple repeats and could be due to errors in either our assembly or the original assembly reported by Molin, Yaguchi, et al. (2020). However, the insertions and deletions are conserved across the four diverse glyphosate-resistant samples, suggesting that they are likely not sequencing or assembly errors, but widespread among resistant samples. Genomic versions of *EPSPS* were found in all samples on contigs that did not have synteny with the canonical eccDNA molecule, but these were determined to be assemblies of the genomic locus containing *EPSPS* on chromosome 7.

A phylogeny of *EPSPS* gene sequences from all sequenced individuals and several outgroup species clustered all *A. palmeri* sequences together, indicating *EPSPS* from an *A. palmeri* plant was inserted into the eccDNA molecule (Figure 5b). Koo et al. (2023) showed that interspecific hybridization can transfer this eccDNA to sexually compatible species of *Amaranthus*, but this result confirms that it originally evolved in *A. palmeri*. Further, Figure 5(b) shows that all eccDNA versions of *EPSPS* group very tightly together, suggesting the insertion of *EPSPS* into the eccDNA occurred only once, evolutionarily recently, followed by rapid geographic spreading through gene flow.

DISCUSSION

We present the first haplotype-resolved chromosome-level assemblies of *A. palmeri*, *A. retroflexus*, and *A. hybridus*; three species that have become troublesome weeds in numerous agricultural systems around the world. Genome assembly identified chromosomal rearrangements that led to diversity in chromosome numbers across the genus. Our analysis identified an approximately 2.84 Mb region near the end of chromosome 3 on *A. palmeri* Hap1 that is male-specific, which is longer than the 1.3 Mb region that was previously identified as male-specific for the species (Montgomery et al., 2021). The difference in size is likely due to the inclusion of optical mapping in this study, which allowed us to represent unassembled regions as stretches of *N*s. With 65.35% of the male-specific region on Hap1 made up of N-gaps, our assembly is still missing male-specific sequences within the region, but we are confident that these gaps are appropriately sized (Figure S3). Considering that some of the gaps are flanked by AT repeats or other combinations of simple sequence repeats, it is possible that the gaps contain difficult-to-assemble repeats that pose assembly challenges but may also house non-repetitive islands with important genes. Future work to assemble these regions with longer reads (e.g., those generated through Oxford Nanopore sequencing) may allow for more contiguous assembly of this region. Nevertheless, we identified key genes (encoding RF1 and TLC) that are likely involved in sex determination. While the function(s) of TLC proteins are unclear in plants, RF1 proteins are well-documented as restorers of fertility in some crops including sorghum and wheat (Klein et al., 2005; Melonek et al., 2021). More so, over half of the cloned restorer genes encode pentatricopeptide repeat-containing (PPR) proteins with roles in restoring normal pollen production to plants carrying a cytoplasmic male sterility locus (Chen & Liu, 2014; Gaborieau et al., 2016; Uyttewaal et al., 2008).

Bobadilla et al. (2023) reported three genes, *PPR247*, *ACD6*, and *WEX*, as likely candidates involved in sex determination in *A. palmeri*. They proposed that the presence of *PPR247* within the MSY region of *A. palmeri* results in the post-transcriptional silencing of *ACD6* and *WEX*, thus enabling the formation or development of male reproductive organs. In our study, we identified an *RF1* gene that shares homology with their identified *PPR* gene, but phylogenetic clustering places these genes on different branches (Figure S11). The other genes proposed by Bobadilla et al. (2023) fall outside of the male-specific region and are therefore not likely to be male determinants, although they could play a role in sex expression.

Recently, Wu et al. (2023) showed that while staminate flowers in *A. palmeri* initiate both carpel and stamen primordia at an early stage, the carpel remains undeveloped, and individuals become functionally male, whereas pistillate flowers initiate and develop only carpel primordium. Given that both androecium and gynoecium are initiated in staminate flowers, but only functional androecium develops, it is plausible that the expression of a dominant activator of maleness (*M*) at stage 4/5 of floral organogenesis (Wu et al., 2023) could result in the initiation and establishment of stamen primordia, whereas the expression of a dominant suppressor of female organs could result in the arrest of the initial gynoecium primordia at stage 7. It is therefore possible that the *RF1* gene within the MSY region of *A. palmeri* is the male-promoting factor and its presence and expression result in the formation of male reproductive organs. However, whether this *RF1* gene also acts as the suppressor of female function is not clear. Knocking out this *RF1* gene should result in male-to-female conversion if it acts as both the male activator and female suppressor but should cause a male-to-sterile conversion if a separate gene acts as the suppressor of female function.

The similarity between chromosome 3 of Hap1 and Hap2 assemblies outside the MSY region in Hap1 suggests that hemizygosity (i.e., the presence of the MSY region near the end of chromosome 3 in Hap1 but absence in Hap2) is responsible for the recombination suppression between the two haplotypes. In contrast, large inversions within the SDR on chromosome 1 of *A. tuberculatus* were identified and might mediate recombination suppression in that species (Raiyemo et al., 2024). Similar to our findings for *A. palmeri*, reduced recombination due to hemizygosity has been reported in garden asparagus (Harkess et al., 2020). The lack of one-to-one synteny between genes in the SDRs of *A. palmeri* and *A. tuberculatus* further supports our previous conclusions of the independent evolution of SDRs in both species (Figure 2). However, the similarity between the GO terms enriched for floral development in both species suggests that the downstream genes or pathways recruited in sex expression are similar. For instance, the most significantly overrepresented terms for male versus female differentially expressed genes in both species were pollen tube growth and regulation of pollen tube growth.

Beyond sex determination, comparative analyses also provided insights into cytochrome P450s and integration of *EPSPS* into eccDNA within *Amaranthus* species. We identified a total of 445 CYP genes in three *Amaranthus* species, which were grouped into A-types and non-A-types, and further divided into nine clans and 39 families. The numbers of CYP clans and families vary among different plant species; for example, Arabidopsis has 9 clans and 47 families (Bak et al., 2011), while soybean has 10 clans and 48 families (Khatri et al., 2022). Among the identified clans, clan 71 contains most of the *Amaranthus* CYP genes, whereas clans 74, 97, 710, 711, and 727 have only a few genes, reflecting their ancient origin and conserved nature

(Bak et al., 2011). The presence of single-family clans 74, 97, and 710 in chlorophytes and charophytes indicates their deep evolutionary roots. Additionally, the emergence of the multi-family clan 71 and the single-family clan 711 is linked to the origin of mosses, with clan 71 expanding across various extant plant genomes (Durst & Nelson, 1995; Hansen et al., 2021; Nelson & Werck-Reichhart, 2011; Werck-Reichhart et al., 2024). The critical role of clan 71 in plant metabolism, stress adaptation, defense, and development may explain its large number of CYP genes.

Findings from this study revealed significant gene expansion (i.e., the presence of multiple copies of the same gene) and conservation (i.e., the same gene appearing across multiple species) at the family and subfamily levels. Notably, CYP76, CYP96, and CYP81 families show expansion across the three *Amaranthus* species. Furthermore, in *A. hybridus* and *A. retroflexus*, CYP71 and CYP72 families were also expanded. Some CYP families were unique to certain species; for example, CYP703 was found exclusively in *A. retroflexus*, and CYP712 was specific to *A. palmeri* (Figure S9a,b). The expansion and conservation of CYPs is a common phenomenon and is likely influenced by the species' habitat and lifestyle (Padayachee et al., 2020). Gain and loss of clans and families in *Amaranthus* species, consistent with findings from several other studies, were also observed (Chhun et al., 2007; Helliwell et al., 2001; Li & Wei, 2020; Nelson et al., 2004). It should be noted that among the discarded CYP gene models, there may be fragmented annotations that, when combined, may represent full gene models. Further in-depth analysis of CYP genes within these species should consider this possibility to maximize their inclusion of CYP diversity.

Molin, Patterson, and Saski (2020) and Molin, Yaguchi, et al. (2020) reported the first complete assembly of a 399 Kbp eccDNA molecule that included *EPSPS* in *A. palmeri*. This eccDNA had little synteny when compared to reference genomes of closely related *Amaranthus* species that were available at the time. Similar approaches using the new Hap1 genome assembly of *A. palmeri* indicate similarly limited synteny between the canonical *EPSPS*-containing eccDNA and the Hap1 genome (Figure S10a). However, when alignment filters are relaxed, most of the eccDNA sequence does find a match in the genome (Figure S10b). Therefore, we hypothesize that the eccDNA is built from a highly repetitive sequence from the *A. palmeri* genome that has undergone significant duplication and rearrangement. Regions of the eccDNA that lack matches even under lax filtration may be regions that were not assembled and included in the reference genome.

Here, we report a contiguously assembled circular sequence containing *EPSPS* that is highly syntenic to the canonical *EPSPS*-containing eccDNA molecule from a glyphosate-resistant plant from Georgia. Structural variants in this new sequence compared to the previously reported sequence are conserved across newly sequenced individuals in our study, suggesting our sequence to be more representative of the eccDNA molecule that has proliferated throughout the world. Our analysis finds no evidence of *EPSPS* being inserted into diverse eccDNA molecules (Figure 5a). In fact, the lack of sequence divergence in *EPSPS* sequence contained within eccDNAs between geographically diverse accessions suggests a single duplication event of *EPSPS* into an eccDNA that has since proliferated throughout the world while generating almost no sequence diversity (Figure 5b; Molin, Patterson, & Saski, 2020). These results highlight the rarity of the generation of *de novo* insertions of herbicide target-site genes into eccDNAs and the need for containment of recently identified herbicide-resistant weed populations.

In sum, our study highlights genomic features in *Amaranthus* species. We revealed candidate genes that are likely involved in sex determination in *A. palmeri*. The development of a working transformation protocol for *A. palmeri* would facilitate the functional validation of these candidates. While gene order was fairly conserved across the assemblies of *Amaranthus* species, chromosomal fusions and other structural rearrangement events contribute to the evolution of chromosomes in the genus. Our study also identified and classified cytochrome P450 genes in each studied species, facilitating investigations of the role of these genes in NTSR and other phenotypes. Finally, we explore the genomic origins of eccDNA in *A. palmeri* and note the conservation of amplified *EPSPS* sequence across diverse accessions. Moving forward, the genomic resources provided here will be valuable for understanding the mechanistic processes involved in herbicide resistance and in furthering the study of adaptive trait evolution within the genus.

MATERIALS AND METHODS

Plant material and growth conditions

The plant materials sequenced in this study are publicly available with the USDA Germplasm Resources Information Network (GRIN). *Amaranthus palmeri* and *A. retroflexus* seeds are available under the accession numbers PI 632235 and PI 572263, respectively, while the seeds of *A. hybridus* are from a population (FT-21605-14) that was sequenced and assembled in Montgomery et al. (2020). Seeds from the three species were sown separately in pots filled with premoistened soil (Lambert LM-GPS), and bottom irrigated. Seedlings were transplanted at approximately 5 cm height into 16 cm pots (America Clay Works I-A650MP) filled with the same soil. Plants were grown at a temperature of 25/20°C and a photoperiod of 16/8 h (light/dark) regimes in greenhouses at Colorado State University. Tissue collection, DNA and RNA extractions, and library preparation have been previously described (Raiyemo et al., 2024). A male *A. palmeri* plant was selected for HiFi and Hi-C sequencing, while a mix of male and female plants were used to generate transcript libraries. Library preparation and

sequencing were performed at the Genome Center of Excellence at Corteva Agriscience.

Genome sequencing, assembly, and annotation

The protocols adopted in the sequencing, assembly, and annotation of *A. palmeri*, *A. retroflexus*, and *A. hybridus* genomes were fully described in Raiyemo et al. (2024). Briefly, high molecular weight DNA was isolated from frozen leaf tissue of *A. palmeri* and *A. retroflexus* using a Nucleobond HMW DNA kit. The isolated DNA was sheared and used to construct an SMRTBell HiFi library, followed by size selection (target 15–20 Kbp). The size selected fragments were then sequenced on PacBio Sequel IIe with a target depth of approximately 50× the expected diploid genome size. Nuclei were extracted from the young leaf tissue of each of the three species for Bionano optical mapping. Additionally, DNA was isolated from the frozen leaf tissue of each of the three species, and chromatin was crosslinked with formaldehyde for Hi-C library preparation and sequencing on Illumina NovaSeq 6000.

Amaranthus palmeri and *A. retroflexus* genomes were phased and assembled independently using HiFi and Hi-C data with Hifiasm v0.16.1 (Cheng et al., 2021). Unless otherwise noted, commands were run with default parameters. The assembly of the Bionano maps, and filtering and alignment of resulting phased contig assemblies to the maps were performed using Bionano Access v1.7 and Bionano Solve software v3.7. Trio-binned *A. hybridus* contigs reported by Montgomery et al. (2020) were also aligned to newly generated Bionano maps. Hybrid scaffolds were constructed from the alignment of contigs to the maps using Bionano Access v1.7 and Bionano Solve v3.7. Gaps of either 13 or 23 bp were inserted between two consecutive or overlapping contigs. The resulting hybrid scaffolds were manually curated and polished by aligning PacBio HiFi corrected reads to the assemblies using minimap2 v2-2.24-r1122 (Li, 2018). Hi-C data were used to validate phasing and place hybrid scaffolds into pseudomolecules using Juicer v0.7.0 and Juicebox (Durand, Robinson, et al., 2016; Durand, Shamim, et al., 2016). Gaps of 100 bp were placed between hybrid scaffolds within a pseudomolecule. Hybrid scaffolds not assigned to any of the pseudomolecules were assigned to Chr00 in each species/haplotype again with 100 bp gaps placed between hybrid scaffolds. Only one haplotype representative was reported for *A. retroflexus* because the second haplotype was substantially shorter (75% of the length of haplotype 1) and less contiguous.

Annotation was performed in two steps: gene models were predicted, and functions were attributed. RepeatModeler v2.0.2, RepeatMasker v4.1.2 (Flynn et al., 2020), and BEDTools v2.30.0 (Quinlan & Hall, 2010) were employed to generate softmasked genomes. Iso-seq PacBio reads were then aligned to the softmasked genomes using pbmm2 v1.10.0 (https://github.com/PacificBiosciences/pbmm2), and transcripts with minimum alignment coverage of 90% and minimum alignment identity of 95% were collapsed to unique isoforms utilizing IsoSeq3 v3.8.2 (https://github.com/PacificBiosciences/IsoSeq). The final gene model predictions were performed with MakerP v1.0 (Campbell et al., 2014) using the softmasked genome, the collapsed gene predictions from IsoSeq3, the classified repeat consensus libraries from RepeatModeler, and protein sequences from *Amaranthus* species available on NCBI. Filtering steps were performed with custom scripts to remove small proteins (<35 amino acids). Functional annotation was performed on the longest isoform of each gene model using several strategies to predict protein localization, function, and homology, through the use of tools including MultiLoc2 v1.0 (Blum et al., 2009), InterProScan5 v5.47 (Jones et al., 2014) and Uniprot, MMSeqs2 v4.1 (Steinegger & Söding, 2017), and databases Uniref50, KEGG Orthology (KO), and NCBI. The genome annotation file includes relevant functional information in the notes column for each gene including the InterPro IDs, GO IDs, and the closest known annotated protein. Further details about genome annotation are described in Raiyemo et al. (2024). Transfer RNA (tRNA) genes were predicted using tRNAscan-SE v2.0.12 (Chan et al., 2021), which is part of the Maker pipeline.

Assembly quality and completeness were determined using the embryophyta_odb10 library within Benchmarking Universal Single-Copy Orthologs (BUSCOs) v4.0.2 (Simão et al., 2015). Additional assembly characteristics were computed using "agat_sp_statistics.pl" from AGAT Toolkit v1.0.0 (Dainat, 2022) and "bbstats.sh" from BBTools (Bushnell, 2014). The quality of intergenic and repetitive sequence space of all assemblies was also accessed using the LTR assembly index (LAI) from the LTR retriever pipeline (Ou et al., 2018; Ou & Jiang, 2018). Putative centromeric repeats in all assemblies were identified by searching the assemblies for centromeric repeats using CentroMiner (Lin et al., 2023). Telomeric repeats were identified by searching the telomere repeat motif $(TTTAGGG)_4$ against all assemblies using BLASTN (Camacho et al., 2009).

Sex determination GWAS in *A. palmeri*

An *A. palmeri* RAD-seq dataset (192 females and 192 males) that was previously used to develop sex-specific markers by Montgomery et al. (2019) was used for GWA analysis. The single-end raw reads data were demultiplexed and cleaned using the "process_radtags" command in Stacks version 2.65 (Rochette et al., 2019), followed by mapping each sample to the Hap1 assembly of *A. palmeri* using BWA-MEM v0.7.17 (Li, 2013). The "gstacks" command from Stacks was then used to build loci from the aligned reads. We grouped the individuals by sex and then ran the "populations" command to filter out rare alleles (--min-maf 0.05). We then accessed the level of missingness in the two data groups using VCFtools v0.1.16 (--missing-indv) following a previous approach (Cerca et al., 2021). We removed individuals that had greater than 60% missing data and then ran the "populations" command once again on the filtered samples utilizing additional filtering criteria (--min-maf 0.05, -p 2, -r 0.4). Using the set of variants obtained, we performed a principal component analysis with PLINK v1.90b7 (Chang et al., 2015) and then carried out a GWA analysis with BLINK-C (Huang et al., 2019) using the first five principal components from PLINK to account for population stratification and familial relatedness. Sex of the individuals were used as binary input (phenotypes) in the analysis. Calculation of *F*-statistics (F_{ST}) was carried out using VCFtools. Manhattan and F_{ST} plots were generated using the CMplot package (Yin et al., 2021) in R.

Identification of previous sex markers and genes in the *A. palmeri* assembly

A search of the previously reported primer sets, PAMS-940 (Montgomery et al., 2019) and JM940 (Montgomery et al., 2021), that amplified male-specific regions of *A. palmeri* was carried out against both haplotype assemblies with BLASTN (Camacho et al., 2009) using the -task blastn-short parameter. Reciprocal best hit (RBH) searches between genes on the previously identified scaffold 20 and the haplotype assemblies, as well as between the two haplotype assemblies, were carried out using MMseqs2 (Steinegger & Söding, 2017).

Synteny and intragenomic analyses

Relationships between putative homoeologs or paralogs within each of the *A. palmeri* haplotypes and in *A. retroflexus* and *A. hybridus* genome assemblies were investigated using MCScanX (Wang et al., 2012). The output collinearity files, which contain the putative homoeologs in anchor pairs were formatted using "jcvi.compara.synteny" from JCVI utility libraries v1.1.22 and plotted together with other annotations (i.e., distribution of gaps, GC, genes, and LTR transposons) using Circos (Krzywinski et al., 2009). To identify orthologous genes among available chromosome-level genome assemblies of *Amaranthus* species; the three assemblies from this study and four (*A. hypochondriacus*, *A. cruentus*, *A. tricolor*, and *A. tuberculatus*) from previous studies were extracted, clustered to identify single-copy genes with one-to-one relationships among species, and the relationships visualized using GENESPACE v1.3.1 (Lovell et al., 2022). MCscan-python version (Tang et al., 2008) from JCVI was also used to identify reciprocal best hit (RBH) and collinear gene blocks between haplotype 1 and 2 of *A. palmeri* genome assemblies. The anchors' file containing collinear gene blocks was visualized and Chr01 plotted using JBrowse 2 (Deish et al., 2023). Sequence alignment between the two haplotypes of *A. palmeri* assemblies was further carried out using Minimap2 v2.24-r1122 (Li, 2021), and structural rearrangements between the two haplotypes were refined using SyRI (Goel et al., 2019). To determine the position of previously identified male-specific regions from a male *A. palmeri* draft assembly in the current assembly, alignment between Hap1 and the draft assembly was carried out using Minimap2. The alignments were then visualized as dotplots using D-Genies (Cabanettes & Klopp, 2018).

Transcriptome profiling and GO enrichment analysis

Quality control (QC) assessed and adapter trimmed mRNA reads for three tissue types (mature flowers, shoot apical meristem, and floral meristems) from four biological replicates of each sex from Bobadilla et al. (2023) were mapped to Hap1 of the *A. palmeri* genome assembly using STAR v2.7.10b (Dobin et al., 2013). Two replicates from the mature female flower category were removed from downstream analyses as in the previous study due to low mapping quality (61.89 and 31.45% uniquely mapped reads for replicate 2 and 4, respectively). Read counting was then carried out using featureCounts v2.0.6 from the subread package (Liao et al., 2014). Genes with a CPM value of less than 1 in at least two samples were filtered out and the counts were normalized using the TMM normalization in edgeR (Robinson et al., 2009). The count normalization was then followed by the analyses of differential expression (DE) between male and female individuals for each tissue type in edgeR. Genes were assessed as differentially expressed based on FDR <0.05 and Log_2FC >1.2 thresholds using the "*glmTreat*" function within edgeR. Heatmaps were constructed with pheatmap package after a log2-transformed normalization of read counts in DESeq2 (Love et al., 2014). The counts were filtered to retain only genes that were differentially expressed from the previous analysis prior to the heatmap construction.

The translated protein-coding sequences (CDS) of Hap1 were assigned GO annotations using eggNOG-mapper v2.1.12 (Cantalapiedra et al., 2021). GO term enrichment analysis was then carried out using topGO with nodeSize = 10. The enrichment test was performed using Fisher's exact test and the "elim" algorithm. GO terms were assessed as significantly enriched at the default $P < 0.01$ threshold. The enrichment plot was generated using ggplot2 (Wickham et al., 2019) in R.

Coverage analysis

To determine the conservation and specificity of genes within the MSY region of the *A. palmeri* Hap1 genome assembly, short reads of male and female *A. watsonii* individuals from previous studies (Raiyemo et al., 2023; Raiyemo & Tranel, 2023) were mapped to the *A. palmeri* Hap1 genome using BWA-MEM v0.7.5 (Li, 2013). SAMtools v1.14 (Danecek et al., 2021) fixmate was used to fill mate coordinates and insert size fields, and duplicates were marked with Picard v2.26.9 (https://broadinstitute.github.io/picard/). The alignment files were then filtered for uniquely mapped reads using SAMtools and "grep" to remove unmapped reads (-F 4), reads with MAPQ <5, alternative hits (tag XA:Z), and supplementary alignments (tag SA:Z). Both the unfiltered and filtered reads for the male and female individuals were visualized in IGV (Thorvaldsdóttir et al., 2013).

Cytochrome P450 identification, naming, and comparative analyses

To identify CYP genes in the annotated protein sequences of *A. palmeri*, *A. hybridus*, and *A. retroflexus*, "grep" was used to identify genes in the annotation file associated with the InterPro codes "IPR001128" and "IPR036396." All identified protein candidates were blasted against all named plant CYPs (http://drnelson.uthsc.edu/CytochromeP450.html) and were submitted to the Standardized Cytochrome Nomenclature Committee (https://drnelson.uthsc.edu/CytochromeP450.html) to ensure consistent naming conventions. Briefly, sequences demonstrating >40% identity were grouped in the corresponding family as the named homologous proteins, while sequences displaying >55% identity were categorized into the same subfamily. Proteins containing less than 350 amino acids were considered fragmented annotations and were not included in further analysis or Table S20.

The amino acid sequences of all full-length CYP proteins from *A. palmeri*, *A. retroflexus*, and *A. hybridus* were aligned using Clustal W (Chenna et al., 2003). A NJ phylogenetic tree was constructed using MEGAX with 1000 bootstrap replications (Kumar et al., 2018). The resulting phylogenetic tree was visualized using the ITOL9 web server (https://itol.embl.de/). MCScanX (Wang et al., 2012) was used to identify the collinearity of CYP genes in the three species. TBtools (Chen et al., 2020) was then employed to map the collinear gene pairs of the CYPs.

Origin of extrachromosomal DNA (eccDNA)

Palmer amaranth seeds were grown under the same conditions outlined for whole genome sequencing. Accessions included glyphosate-resistant populations from Georgia, Tennessee, Arizona, and Brazil as well as glyphosate-susceptible populations from Georgia, Arizona, Kansas, and USDA-S, a GRIN accession (PI 632235) originally collected in 1986 from Arizona that was also used to build the reference genome reported here. One plant per accession of unknown sex was selected for whole genome sequencing with PacBio HiFi technology as outlined above for genome assembly. Each sample was sequenced to a depth of at least 35× according to the haploid genome size of 410 Mbp.

Resequencing of eight plants yielded between 15 and 24 Gbp of PacBio HiFi sequence per individual. The reads of each sequenced sample were assembled *de novo* using Hifiasm v0.19.8 (Cheng et al., 2021). Assemblies were reported for both haplotypes, but only the Hap1 assembly was retained per individual. The number of contigs obtained for each assembly ranged from 833 to 1299 contigs with N50 values approximately 2 Mbp. CD-HIT

v4.8.1 (Fu et al., 2012) was used to collapse contigs with >95% similarity. BLAST v2.10.1+ (Camacho et al., 2009) was used to identify the remaining contigs from each assembly with all eight exons of the *A. palmeri EPSPS* gene model (AmaPaChr07Ag122820 in the Hap1 annotation). The *EPSPS* gene models (including introns) were manually extracted from these contigs for all individuals. Contigs containing the genomic version of *EPSPS* were identified by aligning the sequence of *UDP-glucuronate* (AmaPaChr07Ag122840 in the Hap1 annotation), a single-copy gene near *EPSPS* on chromosome 7. This gene was not co-duplicated with *EPSPS* in the eccDNA molecule, so it was only found co-assembled with the genomic version of *EPSPS*. A phylogenetic tree was generated to infer the evolutionary history of the extracted *EPSPS* sequences along with *EPSPS* sequences from several closely related species, using several *Amaranthus* species as outgroups. The Maximum Likelihood method and Tamura 3-parameter model (Tamura, 1992) were used in MEGA X (Kumar et al., 2018) for tree construction.

Minimap2 v2.26 (Li, 2018) was used to identify regions of synteny (map quality = 60, minimum alignment length = 1000) between the canonical eccDNA harboring *EPSPS* (Molin, Yaguchi, et al., 2020) and the Hap1 assembly of *A. palmeri*. Surviving alignments were plotted with Circos (Krzywinski et al., 2009). Nucmer, part of the MUMmer v4.0.0rc1 package (Marçais et al., 2018), was also used to identify regions of similarity between the eccDNA and Hap1 (minimum cluster length = 25, using all anchor matches regardless of their uniqueness). Surviving alignments were plotted with mummerplot. Minimap2 (Li, 2018) was also used to identify regions of synteny (map quality = 60, minimum alignment length = 1000) between *EPSPS*-containing contigs from the resequenced assemblies and the canonical eccDNA (Molin, Yaguchi, et al., 2020). These alignments were plotted using minidot (https://github.com/thackl/minidot).

AUTHOR CONTRIBUTIONS

PJT, TAG, ELP, CAS, and DRN conceived the research study. JSM and SM grew plants, harvested, and prepared samples for shipping to the sequencing center. VL and KF performed DNA and RNA extraction, PacBio HiFi, Bionano DLS, Hi-C, and Iso-Seq sequencing, as well as genome assembly. ELP and LC annotated the genomes and provided genome assembly metrics. DAR carried out genome-wide association analysis, transposable elements analysis, and comparative genomics analyses. JSM carried out the eccDNA analysis. FA and DRN carried out the cytochrome P450 analysis. AJL carried out the coverage analysis. DAR, JSM, LC, and FA wrote the manuscript. All authors read, revised, and approved the manuscript.

ACKNOWLEDGMENTS

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

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Sequencing data analyzed in this study are available through the National Center for Biotechnology Information (NCBI) (*A. palmeri*, PRJNA1120405; *A. retroflexus*, PRJNA1120413; *A. hybridus*, PRJNA1125363; and *A. palmeri* resequencing, SRA accession SRR30167488–SRR30167495). Genome assemblies including FASTA files and accompanying GFF annotation files have been deposited on NCBI (*A. palmeri* Hap1; JBEFMX000000000, *A. palmeri* Hap2; JBEFMY000000000, *A. retroflexus*; JBEFNP000000000, and *A. hybridus*; JBELOC000000000), CoGe (*A. palmeri* Hap1 [68302], *A. palmeri* Hap2 [68303], *A. retroflexus* [68315], and *A. hybridus* [68316]), and the International Weed Genomics Consortium online database WeedPedia (https://www.weedgenomics.org/species/amaranthus-palmeri/, https://www.weedgenomics.org/species/amaranthus-retroflexus/, https://www.weedgenomics.org/species/amaranthus-hybridus/). A 395 Kbp circular contig containing *EPSPS* from the assembly of a glyphosate-resistant plant from Georgia has been deposited on NCBI (GenBank accession number: PQ096843).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Genomic features of the two haplotype assemblies of *A. palmeri*.

Figure S2. Schematic showing assembly of HiFi + Hi-C generated contigs along with Bionano optical mapping-derived maps to produce hybrid scaffolds for *A. palmeri* chromosome 3.

Figure S3. Coverage of bionano optical map molecules at the proposed sex-determining region on chromosome 3 of haplotype 1 in *A. palmeri*.

Figure S4. Dotplots of sequence alignment between the haplotype assemblies and previous draft assembly.

Figure S5. Differential gene expression analysis between male and female individuals across three tissue types.

Figure S6. Visualization of coverage of *A. watsonii* male and female reads mapped to *A. palmeri* Hap1 genome assembly in IGV after filtering for only uniquely mapped reads.

Figure S7. Visualization of coverage of *A. watsonii* male and female reads mapped to *A. palmeri* Hap1 genome assembly in IGV.

Figure S8. Visualization of coverage of *A. watsonii* male and female reads mapped to *A. palmeri* Hap1 genome assembly in IGV.

Figure S9. Neighbor-joining trees of (a) A-type, and (b) non-A-type.

Figure S10. Comparison of eccDNA containing the gene *EPSPS* in *A. palmeri* to the 17 chromosomes of the nuclear genome.

Figure S11. Phylogenetic tree of coding sequences of *PPRs* and *Rf1* on chromosome 3 of both haplotype assemblies.

Table S1. Annotation statistics for protein coding genes.

Table S2. Number of protein-coding genes in *A. palmeri* haplome 1 and 2 assemblies.

Table S3. Repeat composition of *A. palmeri* haplome 1 and 2, *A. retroflexus*, and *A. hybridus* assemblies.

Table S4. Candidate centromeric repeats for haplome 1 from CentroMiner analysis.

Table S5. Summary of BLAST query of telomeric repeat against *A. palmeri* haplome 1 assembly.

Table S6. Genes within the male-specific Y region (MSY) on chromosome 3 of haplome 1.

Table S7. Inversions between haplome 1 and haplome 2 across the genome. Red color indicates the two largest inversions.

Table S8. Reciprocal best hit (RBH) search of genes from the male-specific scaffold 20 from Montgomery et al. (2020) to *A. palmeri* haplome 1.

Table S9. Differentially expressed genes in male vs female comparison of shoot apical meristem (SAM).

Table S10. Differentially expressed genes for male versus female comparison of floral meristem (FM).

Table S11. Differentially expressed genes for male vs female comparison of mature flower.

Table S12. GO enrichment analysis (Biological process) for the DEGs between male and female comparison for mature flower.

Table S13. Significant genes within enriched GO terms (Biological process).

Table S14. GO enrichment analysis (Molecular function) for the DEGs between male and female comparison for mature flower.

Table S15. GO enrichment analysis (Cellular component) for DEGs between male and female comparison for mature flower.

Table S16. Information on *Amaranthus palmeri* CYPs.

Table S17. Information on *Amaranthus retroflexus* CYPs.

Table S18. Information on *Amaranthus hybridus* CYPs.

Table S19. *A.palmeri, A. retroflexus*, and *A.hybridus* CYP protein sequences.

Table S20. Alignment of *de novo* assembled contigs of glyphosate-resistant *A. palmeri* from Tennessee to EPSPS-containing eccDNA sequence reported by Molin et al. (2020).

Table S21. Alignment of *de novo* assembled contigs of glyphosate-sensitive *A. palmeri* from Arizona to EPSPS-containing eccDNA sequence reported by Molin et al. (2020).

Table S22. Alignment of *de novo* assembled contigs of glyphosate-resistant *A. palmeri* from Arizona to EPSPS-containing eccDNA sequence reported by Molin et al. (2020).

Table S23. Alignment of *de novo* assembled contigs of glyphosate-resistant *A. palmeri* from Brazil to EPSPS-containing eccDNA sequence reported by Molin et al. (2020).

Table S24. Alignment of *de novo* assembled contigs of glyphosate-resistant *A. palmeri* from Georgia to EPSPS-containing eccDNA sequence reported by Molin et al. (2020).

Table S25. Alignment of *de novo* assembled contigs of glyphosate-sensitive *A. palmeri* from Georgia to EPSPS-containing eccDNA sequence reported by Molin et al. (2020).

Table S26. Alignment of *de novo* assembled contigs of glyphosate-sensitive *A. palmeri* (USDA-S) to EPSPS-containing eccDNA sequence reported by Molin et al. (2020).

Table S27. Alignment of *de novo* assembled contigs of glyphosate-sensitive *A. palmeri* from Kansas to EPSPS-containing eccDNA sequence reported by Molin et al. (2020).

Table S28. Summary statistics of raw data used in genome assembly and annotation.

REFERENCES

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S. *et al.* (2011) Cytochromes P450. *Arabidopsis Book*, **9**, e0144.

Barrett, L.G., Legros, M., Kumaran, N., Glassop, D., Raghu, S. & Gardiner, D.M. (2019) Gene drives in plants: opportunities and challenges for weed control and engineered resilience. *Proceedings of the Royal Society B: Biological Sciences*, **286**, 1–9.

Bensch, C.N., Horak, M.J. & Peterson, D. (2003) Interference of redroot pigweed (*Amaranthus retroflexus*), palmer amaranth (*A. palmeri*), and common waterhemp (*A. rudis*) in soybean. *Weed Science*, **51**, 37–43.

Blum, T., Briesemeister, S. & Kohlbacher, O. (2009) MultiLoc2: integrating phylogeny and gene ontology terms improves subcellular protein localization prediction. *BMC Bioinformatics*, **10**, 274.

Bobadilla, L.K., Baek, Y. & Tranel, P.J. (2023) Comparative transcriptomic analysis of male and females in the dioecious weeds *Amaranthus palmeri* and *Amaranthus tuberculatus*. *BMC Plant Biology*, **23**, 1–26.

Bobadilla, L.K. & Tranel, P.J. (2024) Predicting the unpredictable: the regulatory nature and promiscuity of herbicide cross resistance. *Pest Management Science*, **80**, 235–244.

Bushnell, B. (2014) *BBTools software package*. Available from: https://sourceforge.net/projects/bbmap [Accessed May 15, 2024]

Cabanettes, F. & Klopp, C. (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, **6**, e4958.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 1–9.

Campbell, M.S., Law, M.Y., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E. *et al.* (2014) MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiology*, **164**, 513–524.

Camposano, H.S., Molin, W.T. & Saski, C.A. (2022) Sequence characterization of eccDNA content in glyphosate sensitive and resistant palmer amaranth from geographically distant populations. *PLoS One*, **17**, 1–20.

Cantalapiedra, C.P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, **38**, 5825–5829.

Carey, S.B., Lovell, J.T., Jenkins, J., Leebens-Mack, J., Schmutz, J., Wilson, M.A. *et al.* (2022) Representing sex chromosomes in genome assemblies. *Cell Genomics*, **2**, 100132.

Cerca, J., Maurstad, M.F., Rochette, N.C., Rivera-Colón, A.G., Rayamajhi, N., Catchen, J.M. *et al.* (2021) Removing the bad apples: a simple bioinformatic method to improve loci-recovery in *de novo* RADseq data for non-model organisms. *Methods in Ecology and Evolution*, **12**, 805–817.

Chan, P.P., Lin, B.Y., Mak, A.J. & Lowe, T.M. (2021) TRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research*, **49**, 9077–9096.

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 1–16.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y. *et al.* (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, **13**, 1194–1202.

Chen, L. & Liu, Y.G. (2014) Male sterility and fertility restoration in crops. *Annual Review of Plant Biology*, **65**, 579–606.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. & Li, H. (2021) Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods*, **18**, 170–175.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G. *et al.* (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, **31**, 3497–3500.

Chhun, T., Aya, K., Asano, K., Yamamoto, E., Morinaka, Y., Watanabe, M. *et al.* (2007) Gibberellin regulates pollen viability and pollen tube growth in rice. *Plant Cell*, **19**, 3876–3888.

Clauw, P., Ellis, T.J., Liu, H. & Sasaki, E. (2024) Beyond the standard GWAS–A guide for plant biologists. *Plant and Cell Physiology*, **00**, 1–13.

Dainat, J. (2022) Another Gtf/Gff analysis toolkit (AGAT): resolve interoperability issues and accomplish more with your annotations. *Plant and Animal Genome XXIX Conference*. Available from: https://github.com/NBISweden/AGAT.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O. *et al.* (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, **10**, 1–4.

De Figueiredo, M.R.A., Barnes, H., Boot, C.M., De Figueiredo, A.B.T.B., Nissen, S.J., Dayan, F.E. *et al.* (2022) Identification of a novel 2,4-D

metabolic detoxification pathway in 2,4-D-resistant waterhemp (*Amaranthus tuberculatus*). *Journal of Agricultural and Food Chemistry*, **70**, 15380–15389.

Deish, C., **Stevens, G.J.**, **Xie, P.**, **De Jesus, M.T.**, **Hershberg, E.A.**, **Leung, A.** ***et al.*** (2023) JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biology*, **24**, 74.

Dimaano, N.G. & **Iwakami, S.** (2021) Cytochrome P450-mediated herbicide metabolism in plants: current understanding and prospects. *Pest Management Science*, **77**, 22–32.

Dobin, A., **Davis, C.A.**, **Schlesinger, F.**, **Drenkow, J.**, **Zaleski, C.**, **Jha, S.** ***et al.*** (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

Durand, N.C., **Robinson, J.T.**, **Shamim, M.S.**, **Machol, I.**, **Mesirov, J.P.**, **Lander, E.S.** ***et al.*** (2016) Juicebox provides a visualization system for hi-C contact maps with unlimited zoom. *Cell Systems*, **3**, 99–101.

Durand, N.C., **Shamim, M.S.**, **Machol, I.**, **Rao, S.S.P.**, **Huntley, M.H.**, **Lander, E.S.** ***et al.*** (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems*, **3**, 95–98.

Durst, F. & **Nelson, D.R.** (1995) Diversity and evolution of plant P450 and P450-reductases. *Drug Metabolism and Drug Interactions*, **12**, 189–206.

Flynn, J.M., **Hubley, R.**, **Goubert, C.**, **Rosen, J.**, **Clark, A.G.**, **Feschotte, C.** ***et al.*** (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 9451–9457.

Fu, L., **Niu, B.**, **Zhu, Z.**, **Wu, S.** & **Li, W.** (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 3150–3152.

Fu, W., **MacGregor, D.R.**, **Comont, D.** & **Saski, C.A.** (2023) Sequence characterization of extra-chromosomal circular DNA content in multiple blackgrass (*Alopecurus myosuroides*) populations. *Genes*, **14**, 1905.

Gaborieau, L., **Brown, G.G.** & **Mireau, H.** (2016) The propensity of pentatricopeptide repeat genes to evolve into restorers of cytoplasmic male sterility. *Frontiers in Plant Science*, **7**, 1816.

Gaines, T.A., **Duke, S.O.**, **Morran, S.**, **Rigon, C.A.G.**, **Tranel, P.J.**, **Küpper, A.** ***et al.*** (2020) Mechanisms of evolved herbicide resistance. *The Journal of Biological Chemistry*, **295**, 10307–10330. Available from: https://doi.org/10.1074/jbc.REV120.013572

Gaines, T.A., **Zhang, W.**, **Wang, D.**, **Bukun, B.**, **Chisholm, S.T.**, **Shaner, D.L.** ***et al.*** (2010) Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 1029–1034.

Goel, M., **Sun, H.**, **Jiao, W.B.** & **Schneeberger, K.** (2019) SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, **20**, 1–13.

Grant, W.F. (1959) Cytogenetic studies in *Amaranthus*. *Canadian Journal of Botany*, **37**, 413–417.

Hager, A.G., **Wax, L.M.**, **Stoller, E.W.**, **Bollero, G.A.**, **Hager, A.G.** & **Bollero, G.A.** (2002) Common waterhemp (*Amaranthus rudis*) interference in soybean. *Weed Science*, **50**, 607–610.

Hansen, C.C., **Nelson, D.R.**, **Møller, B.L.** & **Werck-Reichhart, D.** (2021) Plant cytochrome P450 plasticity and evolution. *Molecular Plant*, **14**, 1244–1265.

Harkess, A., **Huang, K.**, **van der Hulst, R.**, **Tissen, B.**, **Caplan, J.L.**, **Koppula, A.** ***et al.*** (2020) Sex determination by two Y-linked genes in garden asparagus. *Plant Cell*, **32**, 1790–1796.

Helliwell, C.A., **Chandler, P.M.**, **Poole, A.**, **Dennis, E.S.** & **Peacock, W.J.** (2001) The CYP88A cytochrome P450, *ent*-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 2065–2070.

Huang, M., **Liu, X.**, **Zhou, Y.**, **Summers, R.M.** & **Zhang, Z.** (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience*, **8**, 1–12.

Jones, P., **Binns, D.**, **Chang, H.Y.**, **Fraser, M.**, **Li, W.**, **McAnulla, C.** ***et al.*** (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.

Khatri, P., **Wally, O.**, **Rajcan, I.** & **Dhaubhadel, S.** (2022) Comprehensive analysis of cytochrome P450 monooxygenases reveals insight into their role in partial resistance against *Phytophthora sojae* in soybean. *Frontiers in Plant Science*, **13**, 1–20.

Kistner, E.J. & **Hatfield, J.L.** (2018) Potential geographic distribution of palmer amaranth under current and future climates. *Agricultural & Environmental Letters*, **3**, 1–5.

Klein, R.R., **Klein, P.E.**, **Mullet, J.E.**, **Minx, P.**, **Rooney, W.L.** & **Schertz, K.F.** (2005) Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theoretical and Applied Genetics*, **111**, 994–1012.

Koo, D.H., **Molin, W.T.**, **Saski, C.A.**, **Jiang, J.**, **Putta, K.**, **Jugulam, M.** ***et al.*** (2018) Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed *Amaranthus palmeri*. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, 3332–3337.

Koo, D.H., **Sathishraj, R.**, **Nakka, S.**, **Ju, Y.**, **Nandula, V.K.**, **Jugulam, M.** ***et al.*** (2023) Extrachromosomal circular DNA-mediated spread of herbicide resistance in interspecific hybrids of pigweed. *Plant Physiology*, **193**, 229–233.

Krzywinski, M., **Schein, J.**, **Birol, I.**, **Connors, J.**, **Gascoyne, R.**, **Horsman, D.** ***et al.*** (2009) Circos: an information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.

Kumar, S., **Stecher, G.**, **Li, M.**, **Knyaz, C.** & **Tamura, K.** (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, **35**, 1547–1549.

Legros, M., **Marshall, J.M.**, **Macfadyen, S.**, **Hayes, K.R.**, **Sheppard, A.** & **Barrett, L.G.** (2021) Gene drive strategies of pest control in agricultural systems: challenges and opportunities. *Evolutionary Applications*, **14**, 2162–2178.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 13033997v2. Available from: http://arxiv.org/abs/1303.3997.

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.

Li, H. (2021) New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, **37**, 4572–4574.

Li, Y. & **Wei, K.** (2020) Comparative functional genomics analysis of cytochrome P450 gene superfamily in wheat and maize. *BMC Plant Biology*, **20**, 1–22.

Liao, Y., **Smyth, G.K.** & **Shi, W.** (2014) FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.

Lightfoot, D.J., **Jarvis, D.E.**, **Ramaraj, T.**, **Lee, R.**, **Jellen, E.N.** & **Maughan, P.J.** (2017) Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biology*, **15**, 1–15.

Lin, Y., **Ye, C.**, **Li, X.**, **Chen, Q.**, **Wu, Y.**, **Zhang, F.** ***et al.*** (2023) QuarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Horticulture Research*, **10**, 1–7.

Love, M.I., **Huber, W.** & **Anders, S.** (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 1–21.

Lovell, J.T., **Sreedasyam, A.**, **Schranz, M.E.**, **Wilson, M.**, **Carlson, J.W.**, **Harkess, A.** ***et al.*** (2022) GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*, **11**, 1–20.

Ma, X., **Vaistij, F.E.**, **Li, Y.**, **Jansen van Rensburg, W.S.**, **Harvey, S.**, **Bairu, M.W.** ***et al.*** (2021) 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. *The Plant Journal*, **107**, 613–628.

Manalil, S., **Coast, O.**, **Werth, J.** & **Chauhan, B.S.** (2017) Weed management in cotton (*Gossypium hirsutum* L.) through weed-crop competition: a review. *Crop Protection*, **95**, 53–59.

Marçais, G., **Delcher, A.L.**, **Phillippy, A.M.**, **Coston, R.**, **Salzberg, S.L.** & **Zimin, A.** (2018) MUMmer4: a fast and versatile genome alignment system. *PLoS Computational Biology*, **14**, 1–14.

Massinga, R.A., **Currie, R.S.**, **Horak, M.J.**, **Boyer, J.**, **Massinga, R.A.** & **Currie, R.S.** (2001) Interference of palmer amaranth in corn. *Weed Science*, **49**, 202–208.

Melonek, J., **Duarte, J.**, **Martin, J.**, **Beuf, L.**, **Murigneux, A.**, **Varenne, P.** ***et al.*** (2021) The genetic basis of cytoplasmic male sterility and fertility restoration in wheat. *Nature Communications*, **12**, 1036.

Molin, T., **Patterson, E.L.** & **Saski, C.A.** (2020) Homogeneity among glyphosate-resistant *Amaranthus palmeri* in geographically distant locations. *PLoS One*, **15**, 1–14.

Molin, W.T., **Yaguchi, A.**, **Blenner, M.** & **Saski, C.A.** (2020) The EccDNA replicon: a heritable, extranuclear vehicle that enables gene amplification and glyphosate resistance in *Amaranthus palmeri*. *Plant Cell*, **32**, 2132–2140.

Montgomery, J., **Morran, S.**, **MacGregor, D.R.**, **McElroy, J.S.**, **Neve, P.**, **Neto, C.** ***et al.*** (2024) Current status of community resources and priorities for weed genomics research. *Genome Biology*, **25**, 139.

Montgomery, J.S., **Giacomini, D.**, **Waithaka, B.**, **Lanz, C.**, **Murphy, B.P.**, **Campe, R.** ***et al.*** (2020) Draft genomes of *Amaranthus tuberculatus*, *Amaranthus hybridus*, and *Amaranthus palmeri*. *Genome Biology and Evolution*, **12**, 1988–1993.

Montgomery, J.S., **Giacomini, D.A.**, **Weigel, D.** & **Tranel, P.J.** (2021) Male-specific Y-chromosomal regions in waterhemp (*Amaranthus tuberculatus*) and palmer amaranth (*Amaranthus palmeri*). *The New Phytologist*, **229**, 3522–3533.

Montgomery, J.S., **Sadeque, A.**, **Giacomini, D.A.**, **Brown, P.J.** & **Tranel, P.J.** (2019) Sex-specific markers for waterhemp (*Amaranthus tuberculatus*) and palmer amaranth (*Amaranthus palmeri*). *Weed Science*, **67**, 412–418.

Morgan, G., **Baumann, P.** & **Chandler, J.** (2001) Competitive impact of palmer amaranth (*Amaranthus palmeri*) on cotton (*Gossypium hirsutum*) development and yield. *Weed Technology*, **15**, 408–412.

Murray, M.J. (1940) The genetics of sex determination in the family Amaranthaceae. *Genetics*, **25**, 409–431.

Nelson, D. & **Werck-Reichhart, D.** (2011) A P450-centric view of plant evolution. *The Plant Journal*, **66**, 194–211.

Nelson, D.R., **Schuler, M.A.**, **Paquette, S.M.**, **Werck-Reichhart, D.** & **Bak, S.** (2004) Comparative genomics of rice and arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiology*, **135**, 756–772.

Neve, P. (2018) Gene drive systems: do they have a place in agricultural weed management? *Pest Management Science*, **74**, 2671–2679.

Neves, C.J., **Matzrafi, M.**, **Thiele, M.**, **Lorant, A.**, **Mesgaran, M.B.** & **Stetter, M.G.** (2020) Male linked genomic region determines sex in dioecious *Amaranthus palmeri*. *The Journal of Heredity*, **111**, 606–612.

Oliveira, M.C., **Gaines, T.A.**, **Dayan, F.E.**, **Patterson, E.L.**, **Jhala, A.J.** & **Knezevic, S.Z.** (2018) Reversing resistance to tembotrione in an *Amaranthus tuberculatus* (var. *rudis*) population from Nebraska, USA with cytochrome P450 inhibitors. *Pest Management Science*, **74**, 2296–2305.

Ou, S., **Chen, J.** & **Jiang, N.** (2018) Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Research*, **46**, e126.

Ou, S. & **Jiang, N.** (2018) LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, **176**, 1410–1422.

Padayachee, T., **Nzuza, N.**, **Chen, W.**, **Nelson, D.R.** & **Syed, K.** (2020) Impact of lifestyle on cytochrome P450 monooxygenase repertoire is clearly evident in the bacterial phylum Firmicutes. *Scientific Reports*, **10**, 1–12.

Quinlan, A.R. & **Hall, I.M.** (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Raiyemo, D.A., **Bobadilla, L.K.** & **Tranel, P.J.** (2023) Genomic profiling of dioecious *Amaranthus* species provides novel insights into species relatedness and sex genes. *BMC Biology*, **21**, 1–18.

Raiyemo, D.A., **Cutti, L.**, **Patterson, E.L.**, **Llaca, V.**, **Fengler, K.**, **Montgomery, J.S.** ***et al.*** (2024) A phased chromosome-level genome assembly provides insights into the evolution of sex chromosomes in *Amaranthus tuberculatus*. *bioRxiv*. Available from: https://doi.org/10.1101/2024.05.30.596720

Raiyemo, D.A. & **Tranel, P.J.** (2023) Comparative analysis of dioecious *Amaranthus* plastomes and phylogenomic implications within Amaranthaceae s.s. *BMC Ecology and Evolution*, **23**, 15.

Rigon, C.A.G., **Gaines, T.A.**, **Küpper, A.** & **Dayan, F.E.** (2020) Metabolism-based herbicide resistance, the major threat among the non-target site resistance mechanisms. *Outlooks on Pest Management*, **31**, 162–168.

Rigon, C.A.G., **Küpper, A.**, **Sparks, C.**, **Montgomery, J.S.**, **Peter, F.**, **Schepp, S.** ***et al.*** (2024) Function of cytochrome P450 *CYP72A1182* in metabolic herbicide resistance evolution in *Amaranthus palmeri* populations. *bioRxiv*. Available from: https://doi.org/10.1101/2023.12.13.571468

Roberts, J. & **Florentine, S.** (2022) A review of the biology, distribution patterns and management of the invasive species *Amaranthus palmeri* S. Watson (palmer amaranth): current and future management challenges. *Weed Research*, **62**, 113–122.

Robinson, M.D., **McCarthy, D.J.** & **Smyth, G.K.** (2009) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Rochette, N.C., **Rivera-Colón, A.G.** & **Catchen, J.M.** (2019) Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, **28**, 4737–4754.

Rode, N.O., **Estoup, A.**, **Bourguet, D.**, **Courtier-Orgogozo, V.** & **Débarre, F.** (2019) Population management using gene drive: molecular design, models of spread dynamics and assessment of ecological risks. *Conservation Genetics*, **20**, 671–690.

Sellers, B.A., **Smeda, R.J.**, **Johnson, W.G.**, **Kendig, J.A.**, **Ellersieck, M.R.**, **Science, S.W.** ***et al.*** (2003) Comparative growth of six *Amaranthus* species in Missouri. *Weed Science*, **51**, 329–333.

Simão, F.A., **Waterhouse, R.M.**, **Ioannidis, P.**, **Kriventseva, E.V.** & **Zdobnov, E.M.** (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.

Steckel, L.E. (2007) The dioecious *Amaranthus* spp.: here to stay. *Weed Technology*, **21**, 567–570.

Steinegger, M. & **Söding, J.** (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, **35**, 1026–1028.

Tamura, K. (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution*, **9**, 678–687.

Tang, H., **Wang, X.**, **Bowers, J.E.**, **Ming, R.**, **Alam, M.** & **Paterson, A.H.** (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research*, **18**, 1944–1954.

Thorvaldsdóttir, H., **Robinson, J.T.** & **Mesirov, J.P.** (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, **14**, 178–192.

Trucco, F., **Zheng, D.**, **Woodyard, A.J.**, **Walter, J.R.**, **Tatum, T.C.**, **Lane Rayburn, A.** ***et al.*** (2007) Nonhybrid progeny from crosses of dioecious amaranths: implications for gene-flow research. *Weed Science*, **55**, 119–122.

Uyttewaal, M., **Arnal, N.**, **Quadrado, M.**, **Martin-Canadell, A.**, **Vrielynck, N.**, **Hiard, S.** ***et al.*** (2008) Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *Plant Cell*, **20**, 3331–3345.

Varanasi, V.K., **Brabham, C.** & **Norsworthy, J.K.** (2018) Confirmation and characterization of non-target site resistance to fomesafen in palmer amaranth (*Amaranthus palmeri*). *Weed Science*, **66**, 702–709.

Wang, H., **Xu, D.**, **Wang, S.**, **Wang, A.**, **Lei, L.**, **Jiang, F.** ***et al.*** (2023) Chromosome-scale *Amaranthus tricolor* genome provides insights into the evolution of the genus *Amaranthus* and the mechanism of betalain biosynthesis. *DNA Research*, **30**, 1–15.

Wang, Y., **Tang, H.**, **Debarry, J.D.**, **Tan, X.**, **Li, J.**, **Wang, X.** ***et al.*** (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**, 1–14.

Werck-Reichhart, D., **Nelson, D.R.** & **Renault, H.** (2024) Cytochromes P450 evolution in the plant terrestrialization context. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **379**, 20230363.

Wickham, H., **Averick, M.**, **Bryan, J.**, **Chang, W.**, **McGowan, L.**, **François, R.** ***et al.*** (2019) Welcome to the Tidyverse. *Journal of Open Source Software*, **4**, 1686.

Wu, W., **Jernstedt, J.** & **Mesgaran, M.B.** (2023) Comparative floral development in male and female plants of palmer amaranth (*Amaranthus palmeri*). *American Journal of Botany*, **110**, 1–10.

Yin, L., **Zhang, H.**, **Tang, Z.**, **Xu, J.**, **Yin, D.**, **Zhang, Z.** ***et al.*** (2021) rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics & Bioinformatics*, **19**, 619–628.

Yu, Q. & **Powles, S.** (2014) Metabolism-based herbicide resistance and cross-resistance in crop weeds: a threat to herbicide sustainability and global crop production. *Plant Physiology*, **166**, 1106–1118.