

BREAKTHROUGH REPORT

The eccDNA Replicon: A heritable, extra-nuclear vehicle that enables gene amplification and glyphosate resistance in *Amaranthus palmeri*

William T. Molin^a, Allison Yaguchi^b, Mark Blenner^b and Christopher A. Saski^{cd}

^aUSDA-ARS, Crop Protection Systems Research Unit, Stoneville, MS 38776

^bClemson University, Department of Chemical and Biomolecular Engineering, Clemson, SC 29634

^cClemson University, Department of Plant and Environmental Sciences, Clemson, SC 29634

^dCorresponding Author: Christopher A. Saski (saski@clemson.edu)

Short Title: The eccDNA replicon enables glyphosate resistance in *Amaranthus palmeri*

One-sentence summary: The eccDNA replicon is a large extra-nuclear circular DNA that is composed of a sophisticated repetitive structure, harbors the *EPSPS* and several other genes that are transcribed during glyphosate stress.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Christopher A. Saski (saski@clemson.edu)

ABSTRACT

Gene copy number variation is a predominant mechanism by which organisms respond to selective pressures in nature, which often results in unbalanced structural variations that perpetuate as adaptations to sustain life. However, the underlying mechanisms that give rise to gene proliferation are poorly understood. Here, we show a unique result of genomic plasticity in *Amaranthus palmeri*, a massive, ~400kb extrachromosomal circular DNA (eccDNA), that harbors the 5-enoylpyruvylshikimate-3-phosphate synthase (*EPSPS*) gene and 58 other encoded genes whose functions traverse detoxification, replication, recombination, transposition, tethering, and transport. Gene expression analysis under glyphosate stress showed transcription of 41 of the 59 genes, with high expression of *EPSPS*, aminotransferase, zinc-finger, and several uncharacterized proteins. The genomic architecture of the eccDNA replicon is comprised of a complex arrangement of repeat sequences and mobile genetic elements interspersed among arrays of clustered palindromes that may be crucial for stability, DNA duplication and tethering, and/or a means of nuclear integration of the adjacent and intervening sequences. Comparative analysis of orthologous genes in grain amaranth (*Amaranthus hypochondriacus*) and water-hemp (*Amaranthus tuberculatus*) suggest higher order chromatin interactions contribute to the genomic origins of the *Amaranthus palmeri* eccDNA replicon structure.

INTRODUCTION

McClintock (McClintock, 1984) stated “a sensing mechanism must be present in plants when experiencing unfavorable conditions to alert the cell to imminent danger and to set in motion the orderly sequence of events that will mitigate this danger.” Amplification of genes and gene clusters is an example of one such stress avoidance mechanism that leads to altered physiology and an intriguing example of genomic plasticity that imparts genetic diversity which can rapidly lead to instances of adaptation. This phenomenon is conserved across kingdoms, and these amplified genes are often incorporated and maintained as extrachromosomal circular DNAs (eccDNA). EccDNAs are found in human healthy and tumor cells (Kumar et al., 2017; Moller et al., 2018), cancer cell lines (Vanloon et al., 1994), in a variety of human diseases, and with limited reports in plants and other eukaryotic organisms (Lanciano et al., 2017). Recent research has reported that eccDNAs harbor bona fide oncogenes with massive expression levels driven by increased copy numbers (Wu et al., 2019). Furthermore, this study also demonstrated that eccDNAs in cancer contains highly accessible chromatin structure, when compared to chromosomal DNA, and enables ultra-long range chromatin contacts (Wu et al., 2019). Reported eccDNAs vary in size from a few hundred base pairs to megabases, which include double minutes (DM), and ring chromosomes (Storlazzi et al., 2010; Turner et al., 2017; Koo et al., 2018b; Koo et al., 2018a). The genesis and topological architecture of eccDNAs are not well understood, however, the formation of small eccDNAs (<2kb) are thought to be a result of intramolecular recombination between adjacent repeats (telomeric, centromeric, satellite, etc.) initiated by a double-strand break (Mukherjee and Storici, 2012) and propagated using the breakage-fusion-bridge cycle (McClintock, 1941). Larger eccDNAs, more than 100kb, such as those found in human tumors, most likely arise from a random mutational process involving all parts of the genome (Vonhoff et al., 1992). In yeast, some larger eccDNAs may have the capacity to self-replicate (Moller et al., 2015), and up to eighty percent of studied yeast eccDNAs contain consensus sequences for autonomous replication origins, which may partially explain their maintenance (Moller et al., 2015).

In plants, the establishment and persistence of weedy and invasive species has long been under investigation to understand the elements that contribute to their dominance (Thompson and Lumaret, 1992; Chandi et al., 2013). In the species *Amaranthus palmeri*, amplification of the

gene encoding 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*) and its product, EPSP synthase, confers resistance to the herbicide glyphosate. The *EPSPS* gene may become amplified 40 to 100-fold in highly resistant populations (Gaines et al., 2010). Amplification of the *EPSPS* gene and gene product ameliorates the unbalanced or unregulated metabolic changes, such as shikimate accumulation and loss of aromatic amino acids associated with glyphosate activity in sensitive plants (Gaines et al., 2010; Sammons and Gaines, 2014). Previous low-resolution FISH analysis of glyphosate-resistant *A. palmeri*, shows the amplified *EPSPS* gene is distributed among many chromosomes, suggesting a transposon-based mechanism of mobility; while EPSP synthase activity was also elevated (Gaines et al., 2010). Partial sequencing of the genomic landscape flanking the *EPSPS* gene revealed a large contiguous sequence of 297-kb, that was termed the *EPSPS* cassette (Molin et al., 2017), and amplification of the *EPSPS* gene correlated with amplification of flanking genes and sequence. Flow cytometry revealed significant genome expansion (eg. 11% increase in genome size with ~100 extra copies of the cassette), suggesting that the amplification unit was large (Molin et al., 2017). A follow up study reported that this unit was an intact eccDNA, and through high resolution fiber extension microscopy, various structural polymorphisms were identified that include circular, dimerized circular, and linear forms (Koo et al., 2018a). A comprehensive analysis of eccDNA distribution in meiotic pachytene chromosomes also revealed a chromosome tethering mechanism for inclusion in daughter cells during cell division, rather than complete genome integration. Here, we present the reference sequence of the eccDNA replicon, its unique genomic content and structural organization, experimental verification of autonomous replication, and a putative mechanism of genome persistence. We anticipate these findings will lead to a better understanding of adaptive evolution, and perhaps, toward the development of stable artificial plant chromosomes carrying agronomically useful traits.

RESULTS

Sequence composition and structure of the eccDNA plant replicon

Recent work determined the partial genomic sequence and cytological verification of the presence of a massive eccDNA that tethers to mitotic and meiotic chromosomes for transgenerational maintenance in glyphosate resistant *A. palmeri* (Molin et al., 2017; Koo et al., 2018a). Through single-molecule sequencing of the complete BAC tile path, our sequence assembly further corroborates the circular structure as a massive extrachromosomal DNA element (Koo et al., 2018a). The circular assembly is comprised of 399,435 bp, and is much larger than any eccDNA reported to date in plants (Figure 1A). The internal links connect direct and inverted sequence repeats with their match(es) indicating extensive repetitive sequences within the replicon (Figure 1A).

The eccDNA replicon contains 59 predicted protein coding gene sequences, including the *EPSPS* gene (Supplemental Table 1). The *EPSPS* gene is flanked by 2 putative AC transposase genes (Molin et al., 2017), which belong to a class of transcription factors known to interact with nucleic acid and other protein targets (Figure 1A, Supplemental Table 1). Nearby the *EPSPS* gene are two predicted replication protein A genes (Figure 1A, Supplemental Table 1), which can be involved in repair of double-strand DNA breaks induced by genotoxic stresses (Ishibashi et al., 2005). Downstream of *EPSPS* is a tandem array of genes with unknown functions but contain domains with Ribonuclease H-like activity and a C-terminal dimerization domain common to transposases belonging to the Activator (hAT) superfamily.

Many of the eccDNA replicon encoded genes have predicted functional protein domains that may endow the critical cellular processes necessary for stress avoidance, maintenance, stability, replication, and tethering of the eccDNA replicon. These processes include DNA transport and mobility, molecule sequestration, hormonal control, DNA replication and repair, heat shock, transcription regulation, DNA unwinding, and nuclease activity, (Supplemental Table 1). For example, the eccDNA replicon encodes 6 genes with aminotransferase domains (3 of which are highly expressed in the glyphosate resistant (GR biotype), which have been shown to be crucial for proper cell division and differentiation (Uhlken et al., 2014). The eccDNA replicon encodes a heat shock protein that belongs to the Hsp70 family, which is upregulated by heat stress and toxic chemicals; and may also inhibit apoptosis (Beere et al., 2000). There are 7 genes with plant mobile domains that function in the post-transcriptional gene silencing pathway,

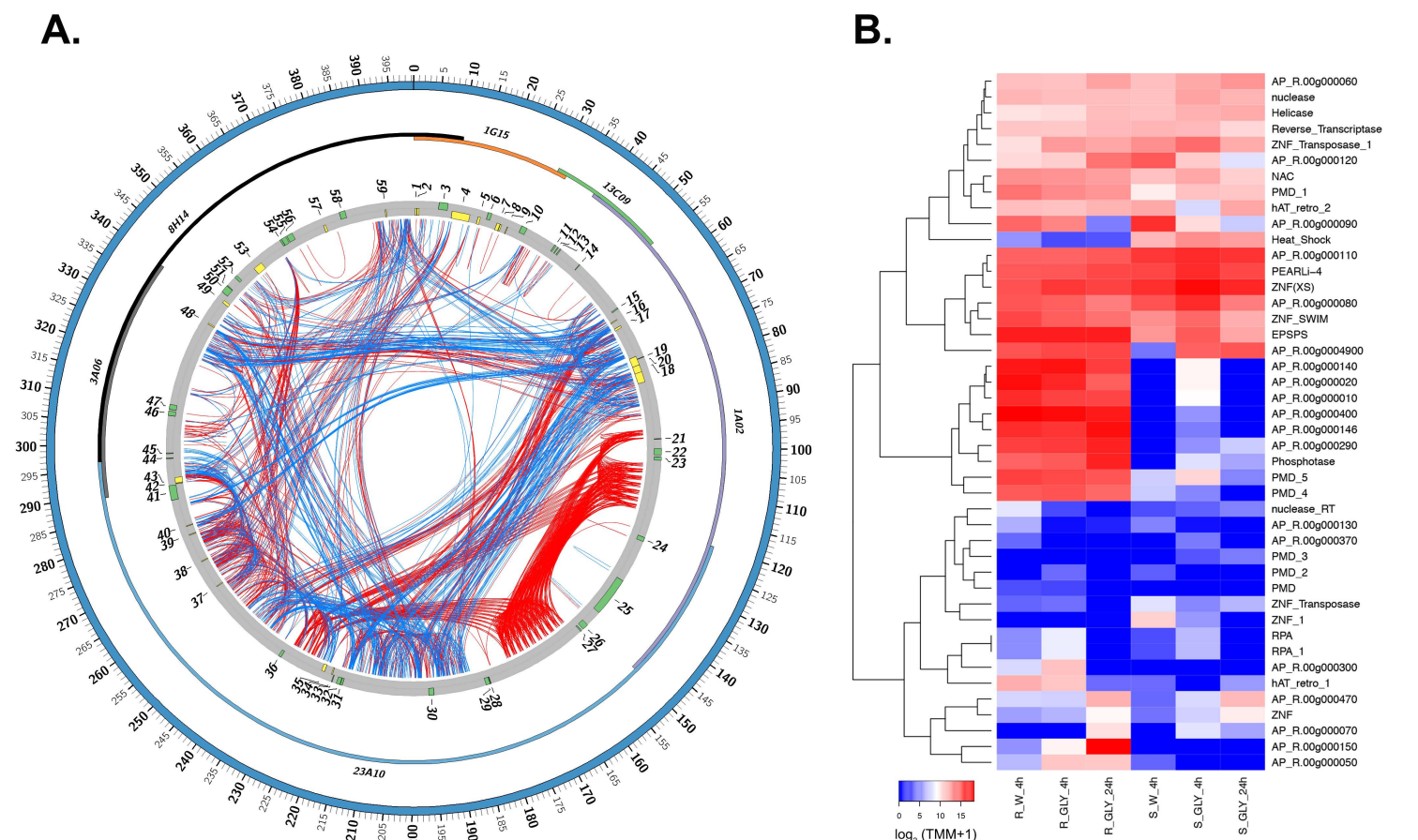


Figure 1. The eccDNA replicon structure and gene expression profiles. **A.** Circos plot of the *Amaranthus palmeri* eccDNA replicon. The solid blue ideogram represents the circular reference assembly of 399,435bp. The multicolored inner ideogram displays the overlapping alignment of sequenced BAC clones (1G15, 13C09, 1A02, 23A10, 3A06, and 8H14) composing the 399k assembly. The next track (gray) depicts the 59 predicted genes with their direction of transcription designated (green is clockwise, yellow is counterclockwise). The internal links connect direct (red) and inverted (blue) sequence repeats with their respective internal match(es). **B.** A heat map of gene expression of key genes under glyphosate (GLY) and water controls in a glyphosate resistant (R) and sensitive (S) biotype at 4 hours and 24 hours after treatment.

132 natural virus resistance, and the production of trans-acting siRNAs (Mourrain et al., 2000;
 133 Peragine et al., 2004) (RNAi machinery). A gene with a NAC domain is also predicted and
 134 expressed, which represents a class of transcription factors that regulate plant defense(Xie et al.,
 135 1999) and abiotic stress responses (Hegedus et al., 2003). There are 6 genes with zinc finger
 136 (ZF) domains, which are known to associate with DNA and protein, perhaps forming DNA-
 137 protein-DNA associations, that may function in chromosomal tethering. AP_R.00g000430
 138 harbors a ZNF-SWIM domain, which has been shown to associate with *RAD51* paralogs to
 139 promote homologous recombination (Makarova et al., 2002; Durrant et al., 2007). Furthermore,
 140 a recent study by Patterson et. al., 2019 has shown a *RAD51* ortholog was co-duplicated with the

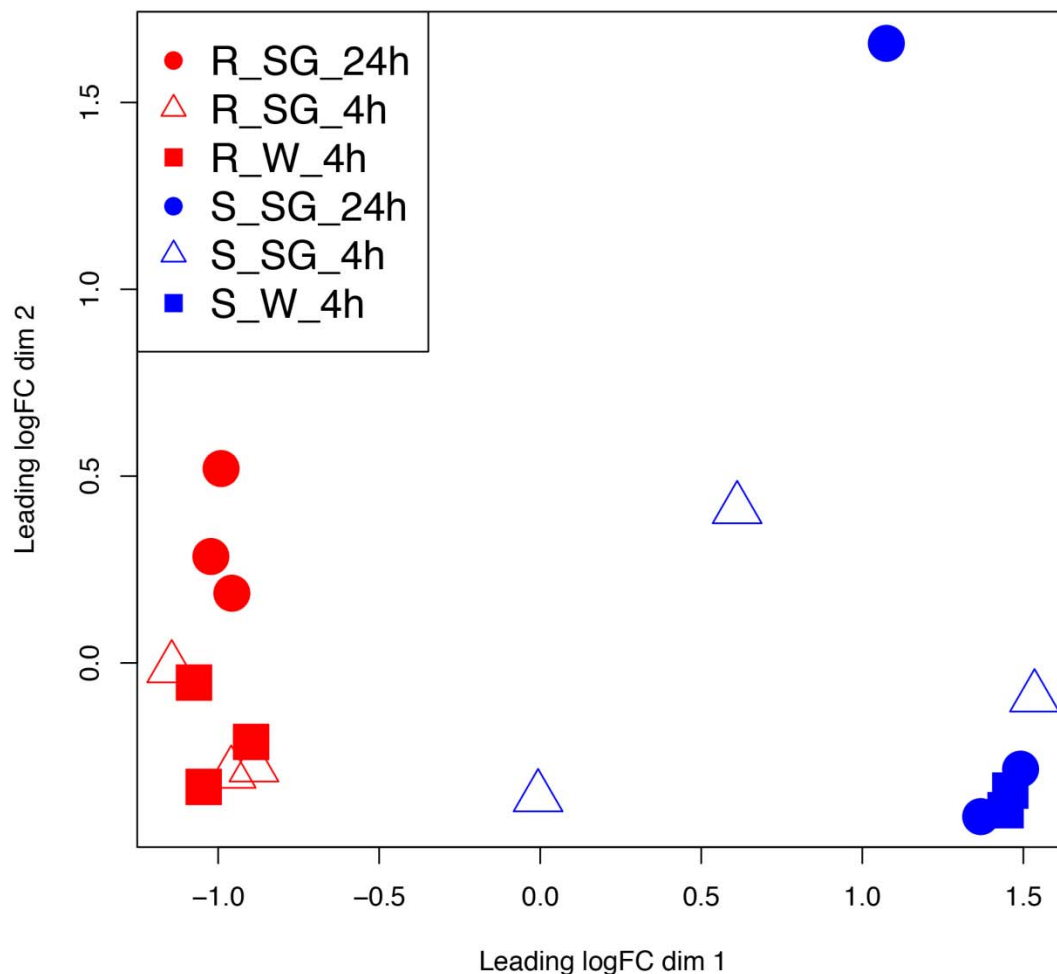
EPSPS locus that confers glyphosate resistance in *Kochia scoparia* (Patterson et al., 2019); which may be indicative of parallel genomic recombination features. AP_R.00g000496 encodes a helicase domain, which is implicated in DNA replication and unwinding. AP_R.00g000450 contains a zinc binding reverse transcriptase domain and an integrase catalytic core, which are characteristic of a retroviral mechanism to integrate viral DNA into the host. This integrase is also found in various transposase proteins and is a member of the ribonuclease H-like superfamily involved in replication, homologous recombination, DNA repair, transposition, and RNA interference (Supplemental Table 1). In the GR biotype with an assembled eccDNA replicon, the most transcriptionally active genes at 24 hours post glyphosate treatment are a number of uncharacterized proteins, the *EPSPS* synthase (AP_R.00g000210), zinc finger nuclease (AP_R.00g000492), and several genes with plant mobile domains (Figure 1B).

Transcriptional activity of the eccDNA replicon under glyphosate stress.

Forty-one of the 59 genes encoded in the eccDNA replicon are transcriptionally active during at either 4 hours or 24 hours under glyphosate or water exposure in replicon containing glyphosate-resistant plants (Figure 1B and Supplemental Table 2). The most transcriptionally active genes include 2 uncharacterized proteins, followed by the *EPSPS* gene (Supplemental Table 2). Additional genes include 6 other uncharacterized protein sequences and 2 aminotransferase like genes that are mostly expressed in the GR biotype. Interestingly, at 4 hours after glyphosate and water control treatments, the resistant and sensitive plants gene expression profiles were indistinguishable from their water controls (Supplemental Figure 1 and Supplemental Table 2). At 24 hours after glyphosate exposure, 11 genes were detected to be upregulated in the GR biotype with at least a 2-fold increase when compared to the glyphosate sensitive biotype (Supplemental Table 3). These genes include 7 uncharacterized proteins, 2 aminotransferase-like genes, *EPSPS*, and a protein with a zinc finger SWIM domain (Supplemental Table 3).

Repetitive elements and structural organization

The eccDNA replicon sequence is composed of a complex arrangement of direct and indirect repeat motifs of variable lengths dispersed among retroelements composed of SINEs, LINEs, and LTR elements, in addition to DNA transposons interspersed by predicted MITE and



Supplemental Figure 1. Multidimensional scaling plot of RNAseq data aligned to the eccDNA replicon reference derived from glyphosate resistant and sensitive biotypes sampled at 4 hours and 24 hours after exposure to glyphosate and water only controls. These data support Figure 1B.

172 Helitron elements (Figure 2A and 2B). Flanking the *EPSPS* gene is an asymmetric set of direct
 173 repeats that are composed of arrays of clustered long and short interspersed palindromic repeat
 174 sequences (CLiSPRs), separated by identical MITES (Figure 2A). The complete palindromic
 175 array is bordered by LTR/ERV, DNA/MULE MUDR and DNA/TdMar Stowaway elements
 176 (not shown). The CLiSPR block regions are also composed of elevated A+T segments (up to
 177 80%), which may serve as a mechanism for stability or nuclear recognition sites for tethering,
 178 integration into open chromatin, or transcriptional hotspots. Downstream of the CLiSPR arrays
 179 are repetitive triplicate clusters of LTR-Cassandra and DNA hAT-Ac elements, each bordered by
 180 MITE elements. Still further downstream are clustered A-rich and LTR/Gypsy elements. These
 181 clustered arrangements may indicate functional relationships. The LTR/Cassandra and

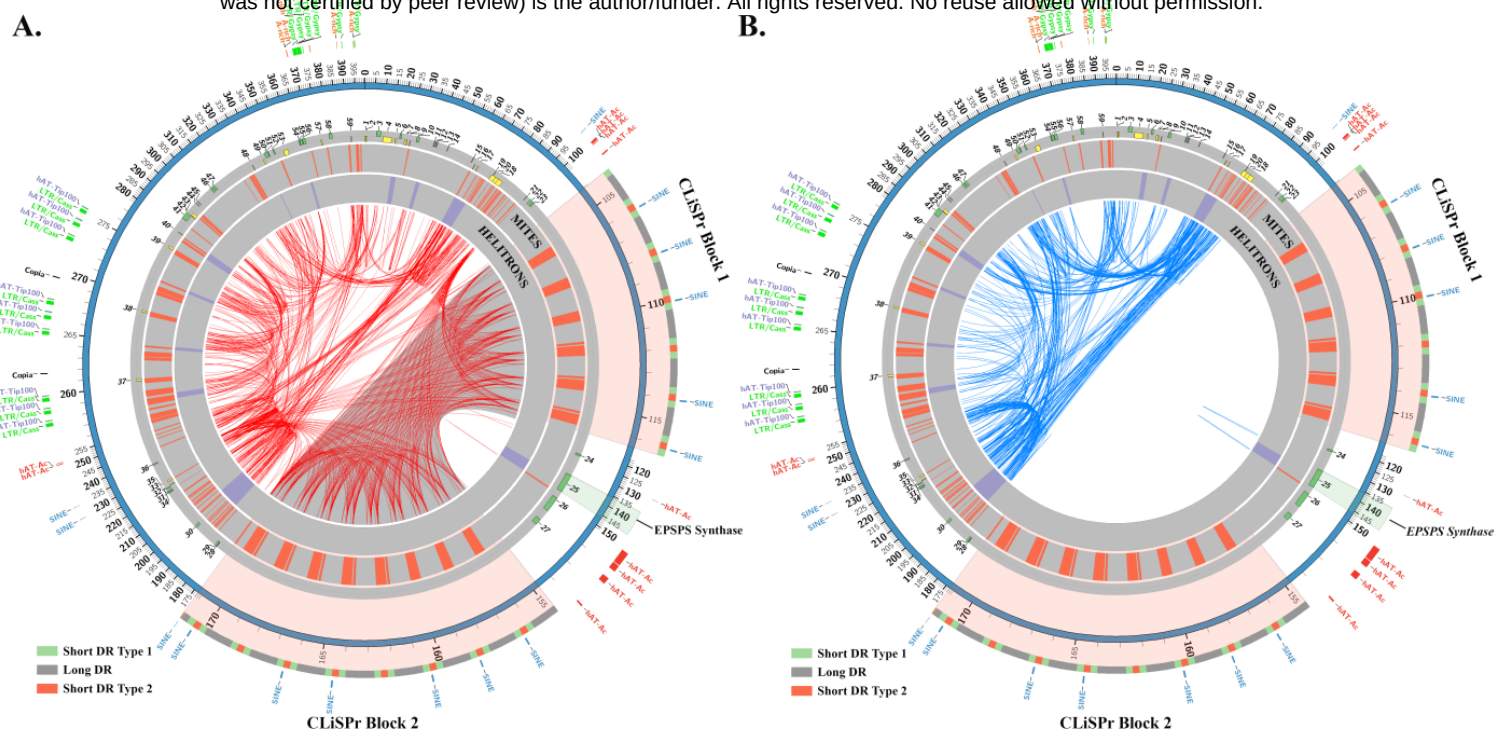


Figure 2. The eccDNA replicon repeat structure. **A.** Circos plot of the *Amaranthus palmeri* eccDNA replicon and key repeat content. The outer colored histograms with labels are predicted repetitive elements with organized arrangements. The highlighted repeat arrays are the clustered long and short interspersed palindromic repeats (CLISPr) arrays that flank the EPSPS synthase gene. The two repeat blocks are large and asymmetric direct repeats. The innertracks depict predicted MITE and helitron repetitive elements. The internal red links are direct repeats and their relationships within the replicon. **B.** In this figure, the direct repeats are hidden and the indirect repeats are shown as blue links.

182 LTR/Gypsy elements divide the coding regions other than that contained within the direct
 183 repeats into 3 segments from 195 to 250 kb, from 287 to 355 kb and from 395 to 90 kb (Figure
 184 3A and 3B).

185 186 *Genome tethering*

187 The first cytological evidence of an eccDNA tethering to chromosomes in plants as a
 188 mechanism of genome persistence was recently reported (Koo et al., 2018b). EccDNA
 189 maintenance has been extensively studied in DNA viruses that maintain their genomes as
 190 extrachromosomal circular DNAs, such as Epstein-Barr, Rhadinovirus, papillomavirus, and
 191 others (Feeney and Parish, 2009). A commonality among these viruses is genome tethering,
 192 which is facilitated by virus encoded DNA-binding proteins that associate with repeated
 193 sequences in the viral genome that also have an affinity with host cell proteins that associate with
 194 mitotic chromatin to ensure nuclear retention (Feeney and Parish, 2009). For example, the
 195 Epstein-barr virus has been reported to anchor to the host genome through interaction of encoded
 196 Epstein-Barr nuclear antigen 1 (EBNA1) and the cellular protein Eukaryotic rRNA processing
 197 protein (EBP2) forming a protein-protein interaction or by directly associating to chromatin via
 198 an AT-hook motif that binds to A/T rich sequences on metaphase chromosomes (Wu et al., 2000;
 199 Sears et al., 2004). In Rhadinoviruses, a role has been suggested for the latency-associated

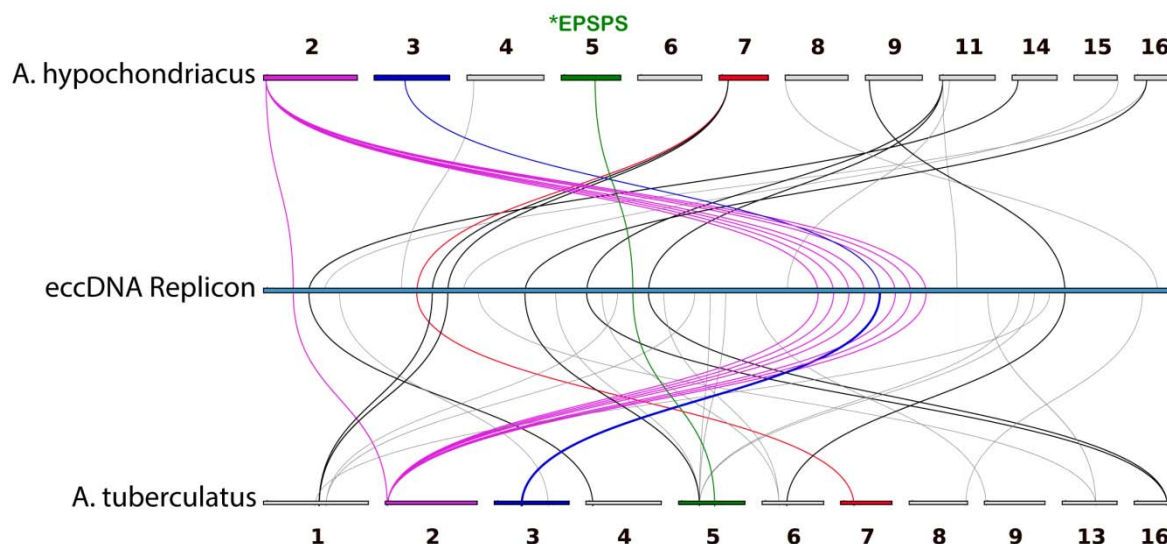


Figure 3. Collinear and syntenic arrangements in the *A. hypochondriacus* and *A. tuberculatus* genomes and their relationship to the eccDNA replicon. Gray lines indicate syntenic gene arrangements between one or the other genomes and colored lines and ideograms indicate the gene is collinear in both genomes.

nuclear antigen gene (*LANA*), which is thought to interact with terminal repeat regions (TR) in the virus (comprised of long tandem repeats) (Russo et al., 1996), where the C-terminus of the protein attaches the TR and the N-terminus tethers the episome to the chromosome (Piolot et al., 2001). In Papillomaviruses, the regulatory protein gene (*E2*) is a multifunctional DNA-binding protein that interacts with the helicase protein (*E1* helicase) for replication (Masterson et al., 1998) and facilitates genome association through interaction within the N-terminal transactivation domain (Skiadopoulos and McBride, 1998). Computational analysis of the eccDNA replicon revealed several genes which may function in the tethering mechanism. AP_R.00g000496 contains 2 core AT-hook motifs (GRP) and also encodes a zinc finger SWIM domain which is recognized to bind DNA, proteins, and/or lipid structures (Supplemental Table 1). The optimal binding sequences of the core AT-hook are AAAT and AATT, which when bound together forms a concave DNA conformation for tight binding (Reeves, 2000). In the eccDNA replicon, there are 143 and 186 (AAAT)₂ or (AATT)₂ motifs, respectively (Figure 2A). There are consistent clusters of tandemly repeated motifs in the CLiSPr repeats. These AT-hook gene products may interact with the A/T rich regions of the eccDNA replicon and other nuclear scaffold proteins. Furthermore, the helicase domain has recently been demonstrated to be an important regulator of the chromatin association, establishment, and maintenance of the Herpes

virus (Harris et al., 2017). AP_R.00g000496 is predicted to encode a helicase motif which may also have a role in tethering of the eccDNA to nuclear chromatin.

*Synteny and collinearity with *Amaranthus hypochondriacus* and *Amaranthus tuberculatus**

It is unknown whether the genesis of the eccDNA replicon was a result of a focal amplification surrounding the *EPSPS* gene or through recombination of distal genomic segments that combined multiple genes for simultaneous amplification. The high degree of synteny and collinearity preserved among the Angiosperms can offer insights into the genomic origin(s) of the eccDNA replicon. The eccDNA replicon was aligned to two the closely related species grain amaranth (*Amaranthus hypochondriacus*) and water-hemp (*Amaranthus tuberculatus*), each with a haploid chromosome number (n=16) and pseudochromosome-scale reference assemblies, (Lightfoot et al., 2017; Kreiner et al., 2019). Both comparator genome assemblies have a single copy of the *EPSPS* gene located near the middle of scaffold 5 (Figure 3). Out of the 59 eccDNA replicon genes, a total of 24 and 34 had reciprocal best hits (RBH) to the grain amaranth and water-hemp pseudochromosomes, respectively. The spatial topology of these ortholog matches was distributed across most of the pseudochromosomes in the assemblies (Figure 4 and Supplemental Table 4). Eleven of the RBH's between the eccDNA replicon and the two *Amaranthus* genomes were located in a collinear fashion on the same scaffolds in the same order (Figure 4 colored ideograms and links). Interestingly, the first orthologous eccDNA gene (AP_R.00g000030 – aminotransferase-like) with synteny to scaffold 2 in both of the genomes is approximately 254 kb from the next cluster of collinear genes (Figure 4 – purple links) which are annotated in *A. hypochondriacus* as F-box family proteins with domain of unknown function (Supplemental Table 4). Moreover, this cluster of genes is separated in the middle by a gene that resides on a completely different pseudochromosome (scaffold 3) and is annotated as a NAC (No Apical Meristem) (Figure 4 blue link and Supplemental Table 4). These results suggest that the eccDNA replicon coding sequences likely originate from multiple genomic regions in *Amaranthus palmeri*, which supports the hypothesis of intra-genomic recombination during the formation of the eccDNA replicon.

The eccDNA replicon is a massive eccDNA vehicle for gene amplification, trait expression, maintenance and transfer of genomic information. The genomic origin is unknown, but likely a result of mobile element activation and extensive genome shuffling that may involve

higher order chromatin interactions that may be influenced by xenobiotic pressures. It has various functional modalities for integration, stability and maintenance to ensure genomic persistence. Furthermore, because of the functional implications of the putative genes in the eccDNA replicon, the presence of this unit could cause a general increase in abiotic stress resilience, or perhaps, an increased disposition to adapt. This vehicle may afford new directions in breeding and biotechnology through deeper understandings of its origin and function. Further investigations will be required to ascertain the essential components of eccDNA replicon proliferation and its compatibility in other species.

Methods

Plant Material

The glyphosate-resistant (GR) *A. palmeri* plant used herein was collected in 2013 by W. T. M. from a soybean field in Washington County, Mississippi that was exposed to extensive application of glyphosate during the decade prior. Sixteen plants were collected and seeds collected from each plant were kept separate. For initial resistance assessment, seeds from each plant were sown in pots containing potting mix (Metro-Mix 360, Sun Gro Horticulture, Bellevue, WA) and lightly covered with 2 mm of mix and subirrigated. Pots were transferred to a greenhouse at the Jamie Whitten Delta States Research Center of USDA-ARS in Stoneville, Mississippi set to 25/20°C ± 3°C day/night temperature and a 15-h photoperiod under natural sunlight conditions supplemented with high-pressure sodium lights providing 400 μmol m⁻² s⁻¹. At the 2 leaf stage the seedlings were sprayed with glyphosate at 0.84 kg·ai·ha⁻¹ using an air-pressurized indoor spray chamber (DeVries Manufacturing Co., Hollandale, MN) equipped with a nozzle mounted with 8002E flat-fan tip (Spraying Systems Co., Wheaton, IL) delivering 190 L·ha⁻¹ at 220 kPa. All seedlings survived from a plant designated #13 and seedlings from this seed lot were used for all experiments. #13 had a copy number as determined by qPCR of 34 (Molin et al., 2018).

BAC Isolation, Sequencing, and Analysis

BAC library construction, partial tile path isolation, sequencing and analysis were described previously (Molin et al., 2017). Two additional BAC clones, 08H14 and 01G15, were determined by chromosome walking by hybridization with overgo probes designed from unique distal sequence on the terminal ends of the *EPSPS* cassette (clones 03A06 and 13C09). These two BAC clones were harvested and sequenced using Pacific Biosciences RSII sequencing to a depth greater than 100X, as described in (Molin et al., 2017). Raw single molecule sequence was self-corrected using the CANU Celera assembler (Koren et al., 2017) with the corOutCoverage=1000 to increase the output of corrected sequences. BAC end sequences were determined using standard Sanger sequencing methods and aligned to the reference assemblies with Phrap and opened in Consed (Gordon et al., 1998) for editing. BAC overlaps were identified using CrossMatch (Gordon et al., 1998) and ends joined manually to form a circular structure. The consensus eccDNA replicon was annotated using a combination of the MakerP

pipeline (Campbell et al., 2014) with RNAseq (below) used as evidence with final manual
curation. Functional domain scans and homology based annotations were determined by
BLAST, InterproScan, and HMM using the SwissProt, non-redundant, and PfamA databases,
respectively. Repeat characterization and masking were conducted with the RepeatMasker
software (<http://www.repeatmasker.org>). MITE and helitron sequences were predicted with the
detectMITE (Ye et al., 2016) and HelitronScanner (Xiong et al., 2014) tools. Circular figures
were prepared using the Circos plotting toolset (Krzywinski et al., 2009).

RNAseq

Plants for RNAseq were grown as previously to the two-true leaf stage under previously
described greenhouse conditions at which time they were transplanted into 8 cm × 8 cm × 7 cm
pots containing the same potting mix. Thereafter, plants were watered as needed and fertilized
once two weeks after transplanting with a water-soluble fertilizer (Miracle-Gro, Scotts Miracle-
Gro Products, Inc., Marysville, OH). When seedlings reached the six-leaf stage they were
sprayed with either water, or water plus surfactant, or water plus surfactant plus glyphosate using
the spray chamber. The surfactant was 0.5 % v/v Tween 20 and glyphosate was applied at 0.42
kg·ai·ha⁻¹ after neutralization with 0.1 N KOH solution. Leaves from the third and fourth nodes
were harvested for RNA extraction at 0, 4 and 24 hours after treatment. Plants were held for two
weeks post leaf harvest to verify survival following glyphosate treatment.

Total RNA was harvested at 4 and 24 hours in biological triplicates using the RNeasy plant
mini kit (Qiagen). Purified RNA was verified for intactness on a Bioanalyzer 2100 (Agilent) and
subject to stranded mRNA-seq using standard TruSeq procedures and sequenced to a target
depth of at least 15M reads per sample. Raw sequence data was preprocessed for adapter and
low-quality bases with the Trimmomatic tool (Krzywinski et al., 2009) and cleaned reads aligned
to the eccDNA replicon consensus assembly with Bowtie2 v.2.3.4.1 (Langmead and Salzberg,
2012) and the following arguments: --no-mixed --no-discordant --gbar 1000 --end-to-end -k
200 -q -X 800. TMM and FPKM transcript quantification was determined with
RSEM v1.3.0 (Li and Dewey, 2011).

Synteny and collinearity

Coding sequences (cds) and annotations (gff3) files for *Amaranthus hypochondricus* (v1.0)
were downloaded from Phytozome (<https://phytozome.jgi.doe.gov>). CDS and annotations for

Amaranthus tuberculatus (v2, id54057) were downloaded from from CoGe (<https://genomevolution.org/coge/SearchResults.pl?s=amaranthus&p=genome>). Reciprocal best hits (RBH) were determined using the Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) considering high identity paralogous matches. The linear riparian plot (Figure 3) was created with custom modifications to the python version of MCscan ([https://github.com/tanghaibao/jcvi/wiki/MCscan-\(Python-version\)](https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version))).

Accession Numbers

Sequence data from this article can be found in the EMBL/GenBank data under BioProject ID PRJNA413471; Submission MT025716.

Author Contributions and Acknowledgements

Author contributions

W.M. and C.S. designed, conducted, analyzed, and interpreted the eccDNA replicon capture, sequencing, and analysis components of the study, which include BAC screening, isolation, sequencing, assembly, annotation, and data analysis. W.M. and C.S. also prepared the initial draft of the manuscript and edited subsequent versions. A.Y. and M. B interpreted data and edited the manuscript.

Supplemental Data

Supplemental Figure 1. Multidimensional scaling plot of RNAseq data aligned to the eccDNA replicon reference derived from glyphosate resistant and sensitive biotypes sampled at 4 hours and 24 hours after exposure to glyphosate and water only controls.

Supplemental Table 1. Predicted coding genes and functional annotation of the eccDNA replicon.

Supplemental Table 2. Transcriptional abundance (log2 TMM+1) of eccDNA replicon genes under water and glyphosate exposure 4 hour and 24 hours after treatment in a glyphosate sensitive and resistant biotype.

Supplemental Table 3. Genes upregulated in glyphosate resistant biotype compared to glyphosate sensitive biotype 24 hours after glyphosate exposure. Numerical values included are normalized counts.

349 Supplemental Table 4. Replicon gene ortholog matches and locations in *A. hypochondriacus* and
350 *A. tuberculatus* pseudochromosome assemblies.

351

Parsed Citations

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Beere, H.M., Wolf, B.B., Cain, K., Mosser, D.D., Mahboubi, A., Kuwana, T., Tailor, P., Morimoto, R.I., Cohen, G.M., and Green, D.R. (2000). Heat-shock protein 70 inhibits apoptosis by preventing recruitment of procaspase-9 to the Apaf-1 apoptosome. *Nat Cell Biol* 2, 469-475.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Campbell, M.S., Law, M.Y., Holt, C., Stein, J.C., Moghe, G.D., Hufnagel, D.E., Lei, J.K., Achawanantakun, R., Jiao, D., Lawrence, C.J., Ware, D., Shiu, S.H., Childs, K.L., Sun, Y.N., Jiang, N., and Yandell, M. (2014). MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. *Plant Physiol* 164, 513-524.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Chandi, A., Milla-Lewis, S.R., Jordan, D.L., York, A.C., Burton, J.D., Zuleta, M.C., Whitaker, J.R., and Culpepper, A.S. (2013). Use of AFLP Markers to Assess Genetic Diversity in Palmer Amaranth (*Amaranthus palmeri*) Populations from North Carolina and Georgia. *Weed Sci* 61, 136-145.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Durrant, W.E., Wang, S., and Dong, X.N. (2007). Arabidopsis SNI1 and RAD51D regulate both gene transcription and DNA recombination during the defense response (vol 104, pg 4223, 2007). *P Natl Acad Sci USA* 104, 7307-7307.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Feeney, K.M., and Parish, J.L. (2009). Targeting mitotic chromosomes: a conserved mechanism to ensure viral genome persistence. *Proc Biol Sci* 276, 1535-1544.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Gaines, T.A., Zhang, W., Wang, D., Bukun, B., Chisholm, S.T., Shaner, D.L., Nissen, S.J., Patzoldt, W.L., Tranel, P.J., Culpepper, A.S., Grey, T.L., Webster, T.M., Vencill, W.K., Sammons, R.D., Jiang, J., Preston, C., Leach, J.E., and Westra, P. (2010). Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. *P Natl Acad Sci USA* 107, 1029-1034.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res* 8, 195-202.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Harris, L., McFarlane-Majeed, L., Campos-Leon, K., Roberts, S., and Parish, J.L. (2017). The Cellular DNA Helicase ChlR1 Regulates Chromatin and Nuclear Matrix Attachment of the Human Papillomavirus 16 E2 Protein and High-Copy-Number Viral Genome Establishment. *J Virol* 91.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Hegedus, D., Yu, M., Baldwin, D., Gruber, M., Sharpe, A., Parkin, I., Whitwill, S., and Lydiate, D. (2003). Molecular characterization of *Brassica napus* NAC domain transcriptional activators induced in response to biotic and abiotic stress. *Plant Mol Biol* 53, 383-397.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Ishibashi, T., Koga, A., Yamamoto, T., Uchiyama, Y., Mori, Y., Hashimoto, J., Kimura, S., and Sakaguchi, K. (2005). Two types of replication protein A in seed plants. *FEBS J* 272, 3270-3281.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Koo, D.H., Molin, W.T., Saski, C.A., Jiang, J., Putta, K., Jugulam, M., Friebe, B., and Gill, B.S. (2018a). Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed *Amaranthus palmeri*. *P Natl Acad Sci USA* 115, 3332-3337.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Koo, D.H., Jugulam, M., Putta, K., Cuvaca, I.B., Peterson, D.E., Currie, R.S., Friebe, B., and Gill, B.S. (2018b). Gene Duplication and Aneuploidy Trigger Rapid Evolution of Herbicide Resistance in Common Waterhemp. *Plant Physiol* 176, 1932-1938.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27, 722-736.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Kreiner, J.M., Giacomini, D.A., Bemm, F., Waithaka, B., Regalado, J., Lanz, C., Hildebrandt, J., Sikkema, P.H., Tranel, P.J., Weigel, D., Stinchcombe, J.R., and Wright, S.I. (2019). Multiple modes of convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*. *PNatl Acad Sci U S A* 116, 21076-21084.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research* 19, 1639-1645.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kumar, P., Dillon, L.W., Shibata, Y., Jazaeri, A.A., Jones, D.R., and Dutta, A. (2017). Normal and Cancerous Tissues Release Extrachromosomal Circular DNA(eccDNA) into the Circulation. *Molecular Cancer Research* 15, 1197-1205.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Lanciano, S., Carpentier, M.C., Llauro, C., Jobet, E., Robakowska-Hyzorek, D., Lasserre, E., Ghesquiere, A., Panaud, O., and Mirouze, M. (2017). Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants. *PLoS Genet* 13, e1006630.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Lightfoot, D.J., Jarvis, D.E., Ramaraj, T., Lee, R., Jellen, E.N., and 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 Biol* 15.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Makarova, K.S., Aravind, L., and Koonin, E.V. (2002). SWIM, a novel Zn-chelating domain present in bacteria, archaea and eukaryotes. *Trends Biochem Sci* 27, 384-386.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Masterson, P.J., Stanley, M.A., Lewis, A.P., and Romanos, M.A. (1998). A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase alpha-primase p68 subunit. *J Virol* 72, 7407-7419.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

McClintock, B. (1941). The Stability of Broken Ends of Chromosomes in *Zea Mays*. *Genetics* 26, 234-282.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792-801.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Molin, W.T., Wright, A.A., Lawton-Rauh, A., and Saski, C.A. (2017). The unique genomic landscape surrounding the EPSPS gene in glyphosate resistant *Amaranthus palmeri*: a repetitive path to resistance. *BMC Genomics* 18, 91.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Molin, W.T., Wright, A.A., VanGessel, M.J., McCloskey, W.B., Jugulam, M., and Hoagland, R.E. (2018). Survey of the genomic landscape surrounding the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene in glyphosate-resistant *Amaranthus palmeri* from geographically distant populations in the USA. *Pest Manag Scie* 74, 1109-1117.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Moller, H.D., Parsons, L., Jorgensen, T.S., Botstein, D., and Regenberg, B. (2015). Extrachromosomal circular DNA is common in yeast. *P Natl Acad Sci USA* 112, E3114-E3122.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Moller, H.D., Mohiyuddin, M., Prada-Luengo, I., Sailani, M.R., Halling, J.F., Plomgaard, P., Maretty, L., Hansen, A.J., Snyder, M.P., Pilegaard, H., Lam, H.Y.K., and Regenberg, B. (2018). Circular DNA elements of chromosomal origin are common in healthy human somatic tissue. *Nat Commun* 9, 1069.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J.B., Jouette, D., Lacombe, A.M., Nikic, S., Picault, N., Remoue, K., Sanial, M., Vo, T.A., and Vaucheret, H. (2000). Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance. *Cell* 101, 533-542.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Mukherjee, K., and Storici, F. (2012). A mechanism of gene amplification driven by small DNA fragments. PLoS Genet 8, e1003119.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Patterson, E.L., Saski, C.A., Sloan, D.B., Tranel, P.J., Westra, P., and Gaines, T.A (2019). The Draft Genome of *Kochia scoparia* and the Mechanism of Glyphosate Resistance via Transposon-Mediated EPSPS Tandem Gene Duplication. *Genome Biol Evol* 11, 2927-2940.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev* 18, 2368-2379.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Piolo, T., Tramier, M., Coppey, M., Nicolas, J.C., and Marechal, V. (2001). Close but distinct regions of human herpesvirus 8 latency-associated nuclear antigen 1 are responsible for nuclear targeting and binding to human mitotic chromosomes. *J Virol* 75, 3948-3959.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Reeves, R. (2000). Structure and function of the HMGI(Y) family of architectural transcription factors. *Environ Health Persp* 108, 803-809.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Russo, J.J., Bohenzky, R.A., Chien, M.C., Chen, J., Yan, M., Maddalena, D., Parry, J.P., Peruzzi, D., Edelman, I.S., Chang, Y., and Moore, P.S. (1996). Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *P Natl Acad Sci U S A* 93, 14862-14867.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Sammons, R.D., and Gaines, T.A (2014). Glyphosate resistance: state of knowledge. *Pest Manag Sci* 70, 1367-1377.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Sears, J., Ujihara, M., Wong, S., Ott, C., Middeldorp, J., and Aiyar, A (2004). The amino terminus of Epstein-Barr virus (EBV) nuclear antigen 1 contains AT hooks that facilitate the replication and partitioning of latent EBV genomes by tethering them to cellular chromosomes. *J Virol* 78, 11487-11505.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Skiadopoulos, M.H., and McBride, A.A (1998). Bovine papillomavirus type 1 genomes and the E2 transactivator protein are closely associated with mitotic chromatin. *J Virol* 72, 2079-2088.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Storlazzi, C.T., Lonoce, A., Guastadisegni, M.C., Trombetta, D., D'Addabbo, P., Daniele, G., L'Abbate, A., Macchia, G., Surace, C., Kok, K., Ullmann, R., Purgato, S., Palumbo, O., Carella, M., Ambros, P.F., and Rocchi, M. (2010). Gene amplification as double minutes or homogeneously staining regions in solid tumors: Origin and structure. *Genome Research* 20, 1198-1206.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Thompson, J.D., and Lumaret, R. (1992). The evolutionary dynamics of polyploid plants: origins, establishment and persistence. *Trends Ecol Evol* 7, 302-307.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Turner, K.M., Deshpande, V., Beyter, D., Koga, T., Rusert, J., Lee, C., Li, B., Arden, K., Ren, B., Nathanson, D.A., Kornblum, H.I., Taylor, M.D., Kaushal, S., Cavenee, W.K., Wechsler-Reya, R., Furnari, F.B., Vandenberg, S.R., Rao, P.N., Wahl, G.M., Bafna, V., and Mischel, P.S. (2017). Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* 543, 122-125.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Uhlken, C., Horvath, B., Stadler, R., Sauer, N., and Weingartner, M. (2014). MAIN-LIKE1 is a crucial factor for correct cell division and differentiation in *Arabidopsis thaliana*. *Plant J* 78, 107-120.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Vanloon, N., Miller, D., and Murnane, J.P. (1994). Formation of Extrachromosomal Circular DNA in HeLa-Cells by Nonhomologous Recombination. *Nucleic Acids Res* 22, 2447-2452.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Vonhoff, D.D., McGill, J.R., Forseth, B.J., Davidson, K.K., Bradley, T.P., Vandevanter, D.R., and Wahl, G.M. (1992). Elimination of Extrachromosomally Amplified Myc Genes from Human Tumor-Cells Reduces Their Tumorigenicity. *P Natl Acad Sci USA* 89, 8165-8169.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only Title Only Author and Title](#)

Wu, H., Ceccarelli, D.F., and Frappier, L. (2000). The DNA segregation mechanism of Epstein-Barr virus nuclear antigen 1. *EMBO Rep* 1, 140-144.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wu, S., Turner, K.M., Nguyen, N., Raviram, R., Erb, M., Santini, J., Luebeck, J., Rajkumar, U., Diao, Y., Li, B., Zhang, W., Jameson, N., Corces, M.R., Granja, J.M., Chen, X., Coruh, C., Abnoui, A., Houston, J., Ye, Z., Hu, R., Yu, M., Kim, H., Law, J.A., Verhaak, R.G.W., Hu, M., Furnari, F.B., Chang, H.Y., Ren, B., Bafna, V., and Mischel, P.S. (2019). Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* 575, 699-703.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Xie, Q., Sanz-Burgos, A.P., Guo, H.S., Garcia, J.A., and Gutierrez, C. (1999). GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. *Plant Mol Biol* 39, 647-656.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Xiong, W.W., He, L.M., Lai, J.S., Dooner, H.K., and Du, C.G. (2014). HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *P Natl Acad Sci USA* 111, 10263-10268.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ye, C.T., Ji, G.L., and Liang, C. (2016). detectMITE: A novel approach to detect miniature inverted repeat transposable elements in genomes. *Sci Rep-Uk* 6.

Pubmed: [Author and Title](#)

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)