

# Sex-specific markers for waterhemp (*Amaranthus tuberculatus*) and Palmer amaranth (*Amaranthus palmeri*)

## Research Article

**Cite this article:** Montgomery JS, Sadeque A, Giacomini DA, Brown PJ, Tranel PJ (2019) Sex-specific markers for waterhemp (*Amaranthus tuberculatus*) and Palmer amaranth (*Amaranthus palmeri*). *Weed Sci.* **67**: 412–418. doi: [10.1017/wsc.2019.27](https://doi.org/10.1017/wsc.2019.27)

Received: 12 March 2019  
Revised: 2 May 2019  
Accepted: 12 May 2019

### Associate Editor:

Ramon G. Leon, North Carolina State University

### Keywords:

Dioecious; gender; restriction site-associated DNA sequencing; sex determination

**Author for correspondence:** Patrick J. Tranel, University of Illinois, 1201 W. Gregory Drive, Urbana, IL, 61801. Email: [tranel@illinois.edu](mailto:tranel@illinois.edu)

\*These authors contributed equally to this work.

Jacob S. Montgomery<sup>1,\*</sup>, Ahmed Sadeque<sup>1,\*</sup>, Darci A. Giacomini<sup>2</sup>, Patrick J. Brown<sup>3</sup> and Patrick J. Tranel<sup>4</sup>

<sup>1</sup>Graduate Student, University of Illinois, Urbana, IL, USA; <sup>2</sup>Research Assistant Professor, University of Illinois, Urbana, IL, USA; <sup>3</sup>Assistant Professor, University of Illinois, Urbana, IL, USA; current: Associate Professor, University of California, Davis, CA, USA and <sup>4</sup>Professor, University of Illinois, Urbana, IL, USA

### Abstract

Waterhemp [*Amaranthus tuberculatus* (Moq.) J. D. Sauer] and Palmer amaranth (*Amaranthus palmeri* S. Watson) are troublesome weeds of row-crop production in the United States. Their dioecious reproductive systems ensure outcrossing, facilitating rapid evolution and distribution of resistances to multiple herbicides. Little is known, however, about the genetic basis of dioecy in *Amaranthus* species. In this work, we use restriction site-associated DNA sequencing (RAD-Seq) to investigate the genetic basis of sex determination in *A. tuberculatus* and *A. palmeri*. For each species, approximately 200 plants of each sex were sampled and used to create RAD-Seq libraries. The resulting libraries were separately bar-coded and then pooled for sequencing with the Illumina platform, yielding millions of 64-bp reads. These reads were analyzed to identify sex-specific and sex-biased sequences. We identified 345 male-specific sequences from the *A. palmeri* data set and 2,754 male-specific sequences in *A. tuberculatus*. An unexpected 723 female-specific sequences were identified in a subset of the *A. tuberculatus* females; subsequent research, however, indicated female specificity of these markers was limited to the population from which they were identified. Primer sets designed to specifically amplify male-specific sequences were tested for accuracy on multiple, geographically distinct populations of *A. tuberculatus* and *A. palmeri*, as well as other *Amaranthus* species. Two primer sets for *A. palmeri* and four primer sets for *A. tuberculatus* were each able to distinguish between male and female plants with at least 95% accuracy. In the near term, sex-specific markers will be useful to the *A. tuberculatus* and *A. palmeri* research communities (e.g., to predict sex for crossing experiments). In the long-term, this research will provide the foundational tools for detailed investigations into the molecular biology and evolution of dioecy in weedy *Amaranthus* species.

## Introduction

The evolution of herbicide resistance in weed species presents a global threat to the sustainability of crop production (Cummins et al. 2013; Yu and Powles 2014). One particularly problematic plant family, Amaranthaceae, includes many agronomically important weeds that have evolved multiple herbicide resistances. Perhaps the most problematic weeds in this genus are Palmer amaranth (*Amaranthus palmeri* S. Watson) and waterhemp [*Amaranthus tuberculatus* (Moq.) J. D. Sauer].

*Amaranthus tuberculatus* and *A. palmeri* are herbaceous, outcrossing annual plants native to the American Midwest and Southwest, respectively (Sauer 1957). Both species have successfully infested a wide range of cropping systems across the United States, and their increasing ability to survive herbicidal treatments makes them two of the main targets for weed management programs in row-crop systems. Whereas most of the species of genus *Amaranthus* are monoecious, *A. palmeri* and *A. tuberculatus* are dioecious (Murray 1940; Waselkov and Olsen 2014). Their dioecy ensures outcrossing to exchange genetic information and maintain high genetic diversity (Adhikary and Pratt 2015), thereby increasing the potential for the evolution and spread of herbicide resistance. Unfortunately, the genetic components of dioecy and its evolution are still not well understood in *Amaranthus* (Gaines et al. 2012; Tranel et al. 2011; Trucco et al. 2009), although some recent studies have had success investigating dioecy in other plant species (Harkess et al. 2017; Henry et al. 2018). In *A. palmeri* and *A. tuberculatus*, it is thought that males are the heterogametic sex (Murray 1940; Trucco et al. 2009); however, cytological evaluation failed to identify heteromorphic sex chromosomes (Grant 1959).

Ultimately, a deeper understanding of dioecy in *A. tuberculatus* and *A. palmeri* could provide for novel strategies to manage them. For example, a gene (or genes) controlling sex could be used in a gene-drive system to manipulate sex ratios, potentially to the point of causing local population collapse (National Academies of Sciences, Engineering, and Medicine 2016; Neve 2018;

Tranel and Trucco 2009). Recent advancements in CRISPR-Cas9 technology have opened up opportunities to create such a gene drive (Chen et al. 2019). To be sure, there are significant technical and regulatory hurdles to controlling *A. tuberculatus* and *A. palmeri* via genetic manipulation of sex ratios. Nevertheless, a first step in this process is to gain a better understanding of the underlying genetic factors involved.

Over the last decade, various studies that used cytogenetic techniques, breeding experiments, and/or linkage analysis with genetic markers have been conducted to identify sex-determining regions in non-model species, including plants and animals (Charlesworth and Mank 2010). The use of linkage analysis to separate out sex-linked or sex-specific markers has recently gained popularity in relation to cytogenetic and breeding techniques, due in part to the ever-decreasing cost of sequencing technologies. Restriction site-associated DNA sequencing (RAD-Seq) is one such next-generation sequencing technique that can aid in carrying out high-throughput linkage analysis (Davey and Blaxter 2010). RAD-Seq allows one to generate a massive number of short reads through parallel sequencing, wherein the regions of genomic DNA that flank restriction sites throughout the genome are selectively sequenced. These sequences can then be compared across hundreds of samples to identify abundantly placed polymorphic markers that can be associated with the trait being investigated. This strategy can be used on pooled populations to screen thousands of markers in a bulk segregant analysis. If a good reference genome is available, the markers can also be used to fine-map regions of interest. RAD-Seq has previously been used successfully in several studies of sex determination to identify genetic markers (Kafkas et al. 2015; Palaiokostas et al. 2013).

In this work, we used RAD-Seq to identify genomic segments that contribute or are linked to sex determination in *A. tuberculatus* and *A. palmeri*. The primary goals of this study were to verify that males are the heterogametic sex in both species and to develop reliable genetic markers that are linked to sex-determining regions. To address these goals, this paper focuses on the following research objectives: (1) identify the presence of male-specific regions of the genome in both species (2) design reliable assays to genotypically predict the sex of these two species based on male-specific sequences, and (3) validate the assays across many populations and individuals to confirm the conservation of these markers.

## Material and Methods

### Sampling and DNA Extraction

For *A. tuberculatus*, 192 males and 192 females were randomly selected from the G<sub>1</sub> generation of plants described by Wu et al. (2018). In brief, this population was obtained from intermating plants from four field collections (three in Illinois and one in Missouri) with or without various herbicide-resistance traits. Two or three leaves were taken from each plant following flower formation, and sex was noted for each individual sampled. The tissues were harvested, placed into 96-well, deep-well (1.2-ml) plates containing metal balls, and stored at -80 C for at least 24 h. DNA was extracted using a protocol described by Xin and Chen (2012). Following the extraction of DNA, the nucleic acid concentration was quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA), and samples were diluted in biology-grade water to a concentration of 50 ng  $\mu\text{l}^{-1}$ .

For *A. palmeri*, 180 males and 175 females were grown from six geographically distinct populations from throughout the United

States (Illinois, Georgia, Kansas, Missouri, Arkansas, and New Mexico; Davis et al. 2015). The plants were grown in the same greenhouse conditions as the *A. tuberculatus* plants and sampled in the same manner. Genomic DNA was extracted, quantified, and diluted as previously described.

### Illumina Library Preparation and RAD-Sequencing

The library preparation and RAD-Seq protocols used in this work are explained in multiple literature sources (Baird et al. 2008; Poland et al. 2012). In summary, a combination of two restriction enzymes, *Hind*III (a 6-base cutter) and *Mse*I (a 4-base cutter), was used to double digest the diluted DNA samples at 37 C for 4 h. The restriction digestion reactions were prepared using 0.5  $\mu\text{l}$  of DNA (50 ng  $\mu\text{l}^{-1}$ ), 1.5  $\mu\text{l}$  of 10X CutSmart Buffer (New England Biolabs, Ipswich, MA), 0.1  $\mu\text{l}$  of rare cutter (*Hind*III HF; New England Biolabs), 0.2  $\mu\text{l}$  of common cutter (*Mse*I; New England Biolabs), and 8.5  $\mu\text{l}$  of molecular biology-grade water. Following digestion, custom adapters (Integrated DNA Technologies, Skokie, IL) were ligated to the cut ends of the DNA fragments. The adapters contained multiplex identifier bar codes, which were used to label each sample, and complementary ends, which paired to enzyme cut sites for *Hind*III and *Mse*I. Ligation reactions were performed at 25C for 2 h, followed by 20 min at 65C for ligation inactivation. Ligation reactions were prepared with 10  $\mu\text{l}$  of digested DNA solution mix, 1  $\mu\text{l}$  of 10X T4 DNA Ligase Reaction Buffer (New England Biolabs), 1.5  $\mu\text{l}$  of *Hind*III adapter (0.1  $\mu\text{M}$ ; New England Biolabs), 0.5  $\mu\text{l}$  of *Mse*I adapter (10  $\mu\text{M}$ ; New England Biolabs), 1.5  $\mu\text{l}$  of ATP (10 mM; New England Biolabs), 0.1  $\mu\text{l}$  of T4 DNA Ligase (2,000 U  $\mu\text{l}^{-1}$ ; New England Biolabs), and 5.4  $\mu\text{l}$  of biology-grade water. The ligated DNA samples were pooled and cleaned, using a 2:1 ratio of AMPure XP beads (Beckman Coulter, Brea, CA) to DNA solution and a magnetic particle concentrator (Invitrogen), with two washes in 95% ethanol and resuspension in 10 mM Tris-HCl (MilliporeSigma, St Louis, MO). Cleaned DNA pools were amplified using a 2X PhusionHF Master Mix (New England Biolabs) with thermocycler conditions as follows: 98 C for 30 s, 15 cycles (98 C for 10 s, 68 C for 30 s, 72 C for 30 s), and 72 C for 5 min. Gel electrophoresis on a 1% agarose gel was used to verify the presence of a smear before DNA samples were once again cleaned with AMPure XP beads. An Agilent Bioanalyzer 2100 was used to determine the average fragment length using the Agilent DNA1000 Kit (Agilent Technologies, Santa Clara, CA), and the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) was used to quantify concentrations of DNA. Samples with concentrations of less than 50 ng  $\mu\text{l}^{-1}$  were discarded before proceeding. These prepared sequencing libraries were pooled (within species) in equimolar concentrations and diluted to a final concentration of 10 nM in a library buffer consisting of 10 mM Tris-HCl with 0.05% v/v Tween-20 (MilliporeSigma). The pooled samples from *A. tuberculatus* and *A. palmeri* were submitted to the W. M. Keck Center at the Roy J. Carver Biotechnology Center at the University of Illinois, Urbana-Champaign, for single-end sequencing on one lane per species of an Illumina HiSeq2000. An additional qPCR assay was performed by the Keck Center on each pool to optimize nucleic acid concentrations before sequencing.

### Bioinformatics Analysis

Parallel workflows were followed in the analysis of the *A. tuberculatus* and *A. palmeri* RAD-Seq data sets. The Trait Analysis by aSSociation, Evolution and Linkage (TASSEL) v. 4.0 software

**Table 1.** Primer sets for sex discrimination in *Amaranthus tuberculatus* and *Amaranthus palmeri*.

Primer set	Forward primer	Reverse primer	Annealing temperature (C)
<i>A. tuberculatus</i> male-specific primer sets			
MU-976	ATTGGGCAGCAGTGT	CAAGACTCGGTCGAA	56
MU-657.2	GATGACCTACTTCAAATG	GGTTGTGCGTTCCAAA	52
MU-533	CATTGTTTGGTGTCTACTG	GTGACAGCAAAAAGTGTG	56
MU-505	CTGAATTCGTGCCACA	TGTATGTGCAAGTAGCC	52
<i>A. tuberculatus</i> female-specific primer sets			
FS-202.2	GGTCCCAACCATGAA	GAAGTTCACCTAGCC	57
FS-216.1	AATTTTCTTACCATCACGCG	AATTGACTGAGAGCCT	61
<i>A. palmeri</i> male-specific primer sets			
PAMS-1106	TCCACTAGTGAAAATATT	AACAAGAGCTATTTCCCTCTC	48
PAMS-940	GGACAAGGCACAACAC	AGCCTTCTCGCACTT	50

package was used to analyze the association of tag presence and sex. Reads originating from underrepresented individuals (those with fewer than 100 total reads) were discarded and attributed to error in library preparation or sequencing. The GBSeqToTagDBPlugin was then used to identify all reads occurring at least 10 times in the total data set. These reads were collapsed into tags (each tag representing a unique sequence read) and retained for downstream analysis. This plug-in also used the bar-code identifiers to demultiplex the data set and input the data into a local SQLite database. A custom Python script was written to count the occurrence of each tag in both males and females and the total number of individuals in which the tag occurred. The tags were subjected to 1,000 iterations of permutation analysis with replacement using the coin package in R, with sex treated as the response variable (Hothorn et al. 2006). In each iteration, the sex of each plant was randomly assigned as male or female. Once this was accomplished, a ratio of male:female occurrence was calculated and logged. After 1,000 iterations, the actual male:female ratio of each tag was compared with the distribution of all 1,000 occurrence ratios derived from the permutations (Hothorn et al. 2008). The resulting P-value reflected the probability of observing the actual male:female ratio derived from the RAD-Seq reads within the distribution of the 1,000 ratios created during the permutation process. Male-biased tags were defined as tags that were assigned P-values less than or equal to 0.001. Additionally, male-specific tags were defined as tags appearing only in male samples and never in female samples. Female-specific and female-biased tags were identified using corresponding criteria.

### Marker Design and Testing

PCR primers were designed to amplify a subset of male-specific tags from each species that had the highest frequency in their respective data sets, while being completely absent from all female individuals. Primers ranged in length from 15 to 18 bp and were designed using a combination of Web-based software packages, including BatchPrimer 3 (You et al. 2008), FastPCR (Kalendar et al. 2017), and OligoPerfect Designer by Life Technologies (ThermoFisher Scientific n.d.). Default settings were used for each program, and the results were compared. Optimal primer sets for each selected tag were designed with consideration of each program output to reduce self-dimerization, heterodimerization, melting temperature difference, and hairpin secondary structure. To ensure specificity, designed primers were aligned, using the Basic Local Alignment Search Tool (BLAST) software package, to all female reads of the same species, and only primers that showed no exact matches in females were examined further (Camacho et al. 2009).

*Amaranthus tuberculatus* male-specific primer sets were tested on separate pools of nondigested male and female DNA samples used in the RAD-Seq protocol. To prepare each sample pool, five diluted DNA samples of the same sex were mixed in equal concentration. Testing these primer sets on pooled samples saved time and reagents while selecting primer sets and protocol conditions that differentiate males from females. *Amaranthus tuberculatus* primer sets that were able to distinguish male from female pools were subsequently tested on individual plants from nine geographically diverse *A. tuberculatus* populations (from Ohio, Missouri, Nebraska, and six different Illinois counties). They were also tested on grain amaranth (*Amaranthus hypochondriacus* L.), both sexes of *A. palmeri*, and hybrids created by crossing *A. tuberculatus* with smooth pigweed (*Amaranthus hybridus* L.). Primer testing on additional species helped to ensure that primer sets were not only sex-specific, but also species-specific. For all *A. tuberculatus* marker tests, the components of each PCR mix included: 12.1 µl of biology-grade water, 5 µl of 5X Green GoTaq Flexi Buffer (Promega, Madison, WI), 2 µl of 25 mM MgCl<sub>2</sub> (Promega), 2 µl of 10 mM equal mixture of dNTPs (New England Biolabs), 1.25 µl each of forward and reverse primer diluted to a concentration of 10 µM (Integrated DNA Technologies), 1 µl of DNA template diluted to 1 to 100 ng µl<sup>-1</sup>, and 0.2 µl each of Taq polymerase (Promega) and dimethyl sulfoxide (DMSO; MilliporeSigma). Thermocycler settings were as follows: 95 C for 5 min; 35 cycles of 95 C 30 s, optimal annealing temperature 30 s, and 72 C 20 s; and 72 C for 5 min, wherein each primer set had its own optimal annealing temperature (Table 1).

One of the primer sets that was most specific in silico (lowest homology to any female sequence based on BLAST), MU\_657.2, was tested further to gain a better understanding of the reliability of this marker across hundreds of samples. *Amaranthus tuberculatus* plants from the original population used in the RAD-Seq assay were grown and DNA was extracted before flowering. Primer set MU\_657.2 was used to predict sex of these plants at the 3- to 4-leaf stage, and phenotypic data were compared with these predicted results.

*Amaranthus palmeri* male-specific primer sets were tested in a similar manner. The reaction mixture for both *A. palmeri* markers included: 13.6 µl of biology-grade water, 5 µl of 5X Green GoTaq Flexi Buffer (Promega), 1.5 µl of 25 mM MgCl<sub>2</sub> (Promega), 1.5 µl of 10 mM equal mixture of dNTPs (New England Biolabs), 1 µl each of forward and reverse primer diluted to a concentration of 10 µM (Integrated DNA Technologies), 1 µl of DNA template diluted to 1 to 100 ng µl<sup>-1</sup>, and 0.2 µl each of Taq polymerase and DMSO. Thermocycler settings for *A. palmeri* male-specific assays were identical to those for *A. tuberculatus* male-specific assays, with

annealing temperatures listed in Table 1. Primer sets that were able to distinguish male pools from female pools were tested across individuals sourced from the six geographically diverse populations of *A. palmeri* used in the RAD-Seq protocol. They were also tested on DNA extracted from male and female *A. tuberculatus*, grain amaranth, and male and female hybrids created by crossing *A. tuberculatus* with *A. hybridus*. To further validate these primer sets and better understand their reliability, *A. palmeri* plants from three field populations in Kansas were grown, and sex predictions determined through PCR assays were compared with phenotypic data as in *A. tuberculatus*.

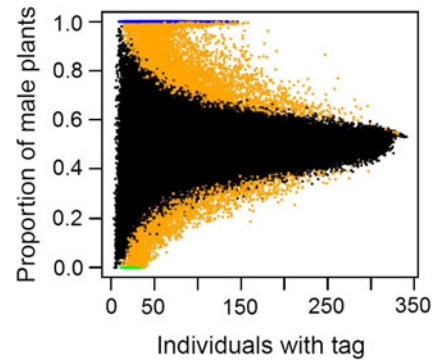
### In Silico Marker Testing

Short-read, shotgun sequence data at a depth of ~10X were generated from 23 male and 19 female *A. tuberculatus* plants from Illinois as part of an ecological study to better understand *A. tuberculatus* population structure (Kreiner et al. 2018). The sequence reads produced by this protocol were 150-bp paired-end reads with an approximate insert length of 300 bp. We compared by BLAST all *A. tuberculatus* male- and female-specific tags that were used to produce primer sets that reliably differentiated males from females from the RAD-Seq DNA samples to all reads of each male and female data set separately under default settings (Camacho et al. 2009). Furthermore, we used SOAPdenovo2 to assemble as much of the genome as possible for each individual after quality-trimming the reads with Trimmomatic (v. 0.38) under default paired-end settings (Bolger et al. 2014; Luo et al. 2012). Optimal *k*-mer size and coverage cutoff used in each assembly were identified with KmerGenie (Chikhi and Medvedev 2014). The four male-specific tags were compared by BLAST against these assemblies to find associated genic regions (Camacho et al. 2009). Genome segments showing perfect alignment were then compared using BLASTx against the National Center for Biotechnology Information's (NCBI) nonredundant protein sequence database to search for homology to known genic regions or transcripts.

### Investigation of Female-specific Sequences

To investigate tags appearing only in females, PCR primer sets were designed for two female-specific tags with the highest frequency in the data set and with no close alignments to male individuals in silico using the same software tools as indicated earlier to minimize self-dimerization, heterodimerization, melting temperature difference, and hairpin secondary structure. Reagent concentrations and thermocycler settings for these female-specific assays were similar to those for *A. palmeri* male-specific assays, with primers and annealing temperatures listed in Table 1. Primers were validated on DNA samples used in the RAD-Seq protocol.

One hypothesis for the female-specific markers was the presence of a cryptic Y region (Y') in a subset of the females, likely derived from the male-specific Y (MSY) region, but nonfunctional. We further hypothesized that, in combination with a true MSY region, this genotype (YY') would be missing key X genes, leading to a lethal phenotype, thus skewing the ratio of the population toward more females. To test for this, plants were grown from the same seed lot used to produce the plants involved in RAD-Seq and screened with the female-specific markers. All female plants were allowed to intermate with males from the same population, and seed was harvested from each female individually. Seed from 10 females testing positive for the female-specific markers (FS+) and from six females testing negative (FS-) were



**Figure 1.** Plot of individual tags from *Amaranthus tuberculatus* based on total number of individuals in which each was present (x axis) and male:female proportion (y axis). Male-specific tags are colored blue, female-specific tags green, sex-biased tags orange, and sex-neutral tags black.

germinated, and 108 seedlings from each female parent were transplanted into soil and grown until sex could be identified under greenhouse conditions. Sex of all flowering plants was documented, and sex ratios were compared among populations derived from FS+ and FS- females. A chi-square goodness-of-fit test was used to determine whether any of the tested populations' male:female ratio differed from the expected 1:1 at a confidence level of 95%.

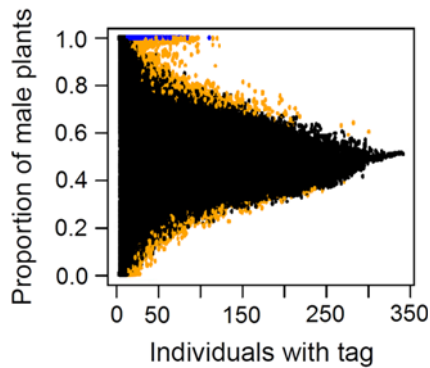
## Results and Discussion

### Bioinformatics Analysis

The RAD-Seq protocol generated 204,658,182 total *A. tuberculatus* reads, each consisting of 64 bp after adapter and bar-code trimming. After these raw reads were processed with the GBSSeqToTagDBPlugin of TASSEL 4.0, they were consolidated into 13,031,934 total tags (Bradbury et al. 2007; Glaubitz et al. 2014). A total of 522,712 tags that occurred at least 10 times throughout the entire data set were retained, representing 178 males and 175 females. These tags were plotted to visualize the number of individuals in which each appeared and in what gender proportion (Figure 1). Sex-specific tags were identified as tags that were solely found in one gender and were labeled as male specific or female specific. There were 2,754 male-specific tags (found in 25 or more males and zero females) and 723 female-specific tags (found in 25 or more females and zero males). To select the most likely male-specific tags for later marker design, a final filtering step retained any tag that occurred 500 times or more in males. Twenty-two male-specific tags passed these criteria.

A similar analysis pipeline was followed with the *A. palmeri* RAD-Seq data. A total of 396,075,230 reads were passed into GBSSeqToTagDBPlugin TASSEL 4.0 and reduced to 6,209,074 unique tags. As in *A. tuberculatus*, any tag appearing less than 10 times in the data set was discarded, resulting in 1,037,755 unique tags across 140 males and 152 females to be analyzed further. These remaining tags were subjected to the same permutation analysis as the *A. tuberculatus* tags (Figure 2). Of the male-specific tags, 345 appeared in 25 or more male individuals and zero females. There were no female-specific tags in the *A. palmeri* data set.

The higher proportion of male-biased to female-biased tags in the two data sets supports the hypothesis that males are the heterogametic sex in both species. In a traditional XY sex chromosome system, there should be no sequence that is specific to females (the



**Figure 2.** Plot of individual tags from *Amaranthus palmeri* based on total number of individuals in which each was present (x axis) and male:female proportion (y axis). Male-specific tags are colored blue, sex-biased tags orange, and sex-neutral tags black.

female-specific *A. tuberculatus* tags discovered in our RAD-Seq protocol are discussed under “Investigation of Female-specific Sequences”). In contrast, sequences nearby or inside a non-recombining region of the Y chromosome are liable to accumulate mutations and become specific to the Y chromosome, giving rise to male-biased and male-specific tags. The larger number of well-supported male-specific tags (those appearing solely in males and being represented by 25 or more individuals) in *A. tuberculatus* relative to *A. palmeri* suggests *A. tuberculatus* diverged from monoecy more recently than did *A. palmeri* (Bachtrog 2013; Puterova et al. 2018). It is hypothesized that a recently evolved dioecious species quickly accumulates transposable genomic segments in the non-recombining region of the Y chromosome. As evolutionary time goes on, the non-recombining region is hypothesized to degenerate, as deleterious mutations eliminate unnecessary genomic segments until the region is reduced to only the essential components for sex determination, as in the mammalian Y chromosomes. The proposition that *A. tuberculatus* evolved a dioecious reproduction system more recently than *A. palmeri* is also supported by previously conducted phylogenetic studies (Stetter and Schmid 2017; Ward et al. 2013).

### Marker Design and Testing

Male-specific tags from both species that showed the highest number of occurrences throughout their respective data sets were used to design PCR primer sets. These tags were assumed to be sequences that were part of the MSY and with little homology to sequences found in the X chromosome or any autosome. Thus, amplification should only occur in male samples. With just 64 bp per tag, options for primer design were limited. Several tags were excluded due to this primer design constraint, and even the successful primers (Table 1) required stringent annealing temperatures for sex discrimination.

Four *A. tuberculatus* primer sets and two *A. palmeri* primer sets were able to accurately distinguish male pools from female pools (Table 1). Figure 3 illustrates representative results of each *A. tuberculatus* and *A. palmeri* male-specific tag when tested on one male and one female of their own species. All four *A. tuberculatus* primer sets and both *A. palmeri* primer sets successfully differentiated males from females in groups of 30 plants of geographically diverse populations with no false positives or false negatives (unpublished data). The *A. palmeri* primer sets were tested on populations from Illinois, Georgia, Kansas, Missouri, Arkansas, and New Mexico (the same populations used for RAD-Seq), while the *A. tuberculatus* primer sets were tested on populations from

Missouri (also used in the original RAD-Seq), Ohio, Nebraska, and six different Illinois counties. Additionally, these primer sets did not show any amplicons in any other species, with the exception of *A. tuberculatus* primer sets showing amplicons in male *A. tuberculatus* × *A. hybridus* hybrids (unpublished data). Amplification in only male *A. tuberculatus* and male hybrid samples indicates both sex specificity and species specificity, given that male hybrid plants received the genetic component of maleness from the male *A. tuberculatus* parent (Trucco et al. 2009).

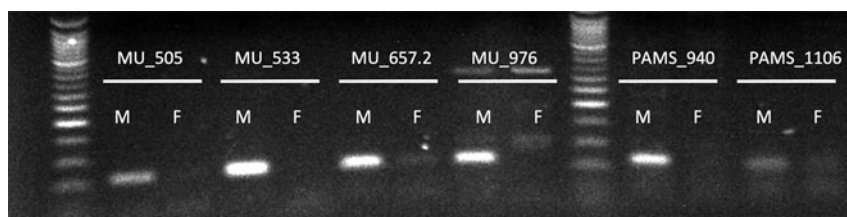
Primer set MU\_657.2 was used to screen 327 more *A. tuberculatus* plants originating from Illinois and Missouri, and it accurately predicted sex for 98% of the plants (Table 2). Similarly, *A. palmeri* primer sets PAMS\_940 and PAMS\_1106 were each able to predict the sex of 48 plants from three additional field populations in Kansas with 96% accuracy (Table 2). Sequence divergence in the marker regions could explain why the markers failed to detect some males. Additionally, the marker might not be entirely within the non-recombining MSY region. If so, recombination between the marker and the sex-determining locus would result in both false positives and false negatives. In both species, the validation of male-specific markers lends further strength to the hypothesis that males are the heterogametic sex.

No significant homology was found between male-specific tags in *A. tuberculatus* and *A. palmeri* (unpublished data). This lack of homology suggests that these two species evolved dioecious reproductive systems independently. Alternatively, we might have simply failed to recover the conserved sequences between the two species that are critical for sex determination.

### In Silico Marker Testing

Among the 23 male individuals, male-specific tags MU\_976, 533, and 505 all aligned perfectly (no mismatches or gaps in the alignment) to 18 males in silico. Male-specific tag MU\_657.2 aligned to 21 males perfectly. The two individuals in which MU\_657.2 was not present also did not show the presence of any of the other male-specific tags tested. None of the four male-specific tags were found in any of the 19 female individuals. This leads us to believe that variation in the Y chromosome makes these markers imperfect, although even without perfect sequence conservation across the tag, the primers may still yield a PCR product in males, giving a correct result. It is also possible that dioecious *Amaranthus* species exhibit some plasticity in sex expression (i.e., sex is not under absolute genetic control), as has been observed in other plants, including spinach (*Spinacia oleracea* L.) (Komai and Masuda 2004).

The sequenced *A. tuberculatus* genomes were used to search for contigs containing the male-specific markers. After trimming, approximately 90% of read sequence was retained for assembly. Because of the short read length and low depth of coverage, the average N50 of each genome assembly was only approximately 500 bp, with a maximum contig length of nearly 20,000 bp. Male-specific tags MU\_976, 533, and 505 showed no suitable alignments in any of the male assemblies, likely because reads containing these tags were either not used in the assembly or assembled to a length less than 300 bp. However, MU\_657.2 aligned perfectly to a single contig in each of the 21 male genomes to which it originally aligned. Contigs that housed the tag were 3,000 to 7,000 bp in length and were highly conserved. When these contigs were compared by BLAST to NCBI’s nonredundant protein sequence database, each contig shared the greatest homology with a Ty3 retrotransposon of the gypsy subclass found in beet (*Beta vulgaris* L.). Furthermore, the tag used to generate primer set MU\_657.2 was



**Figure 3.** Representative PCR products obtained from male (M) and female (F) *Amaranthus tuberculatus* (MU primer sets) or *Amaranthus palmeri* (PAMS primer sets) plants. Primers for each primer set are provided in Table 1. The major PCR product obtained from males migrated between the 50-bp and 100-bp bands of the DNA ladder.

**Table 2.** Results of male-specific PCR assays on male and female plants.<sup>a</sup>

Primer set	Male presence	Male absence	Female presence	Female absence
MU-657.2	188	7	1	161
PAMS-940	23	2	0	23
PAMS-1106	23	2	0	23

<sup>a</sup> Assays were performed on DNA extracted from plants in the 3- to 4-leaf stage, and sex was phenotypically determined for each plant upon flowering.

contained within the proposed retrotransposon. Perhaps a transposable element once jumped into the non-recombining region of the Y-chromosome, where it mutated and lost function, and further mutations became unique to the Y-chromosome and were captured by MU\_657.2.

### Investigation of Female-specific Sequences

We unexpectedly found 723 tags that only appeared in a subset of female individuals. The discovery of female-specific tags in our *A. tuberculatus* data set brought our hypothesis of a traditional XY sex chromosome system into question. Given that males are hypothesized to be the heterogametic sex, any genetic sequence found in females should also be found in males. To investigate these tags further, primer sets were designed based on the 10 most frequent female-specific tags found in *A. tuberculatus*. These sets were tested on the original RAD-Seq DNA samples. After testing primer sets and optimizing PCR conditions, two of the primer sets proved to be female-specific across the RAD-Seq samples and were used for downstream testing (Table 1).

Siblings of the plants used to generate the RAD-Seq data were screened for the presence of these female-specific tags. An initial screen of 72 plants with these markers showed hits to 19 females and 2 males (unpublished data). Females testing positive for the female-specific markers (FS+) showed no discernibly different phenotype when compared with females lacking the female-specific markers (FS-), and the two female-specific markers segregated in male:female ratios of 0.96:1 and 0.90:1 in FS+ progeny, indicating the lack of true female specificity. In addition, chi-square analysis of each progeny's sex ratio concluded that none significantly deviated from an expected 1:1 (male:female) distribution (unpublished data). This lack of lethality of a subset of the male offspring from FS+ females infers that the FS+ plants have two functional X regions, and the female-specific sequences do not belong to a cryptic Y region. Possibly the female-specific sequences we identified in the generation of plants in which we performed RAD-Seq were unique to that generation and arose from a mutational event (e.g., an insertion) in the X region of a male plant in the preceding generation.

In summary, our results are consistent with previous work that hypothesized males to be the heterogametic sex in both *A. tuberculatus* and *A. palmeri*. When high-quality reference genomes from males of the two species become available, the male-specific markers we identified should enable homing in on potential causative male-determining genes. In the long term, manipulation of such genes could form the basis of a novel genetic control strategy, perhaps through gene-drive technology, in which all plants of a population are converted to males. In the short term, the markers we identified are immediately valuable, for example, for setting up controlled crosses or for investigating environmental plasticity of sex in these two species.

**Author ORCID.** Patrick J. Tranel <https://orcid.org/0000-0003-0666-4564>.

**Acknowledgments.** The USDA National Institute of Food and Agriculture (AFRI project 2018-67013-27818) provided partial funding of this research. No conflicts of interest have been declared.

### References

- Adhikary D, Pratt DB (2015) Morphologic and taxonomic analysis of the weedy and cultivated *Amaranthus hybridus* species complex. *Syst Bot* 40:604–610
- Bachtrog D (2013) Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* 14:113–124
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10:421
- Charlesworth D, Mank JE (2010) The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. *Genetics* 186:9–31
- Chen K, Wang Y, Zhang R, Zhang H, Gao C (2019) CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu Rev Plant Biol* 70:667–697
- Chikhi R, Medvedev P (2014) Informed and automated *k*-mer size selection for genome assembly. *Bioinformatics* 30:31–37
- Cummins I, Wortley DJ, Sabbadin F, He Z, Coxon CR, Straker HE, Sellers JD, Knight K, Edwards L, Hughes D, Kaundun SS, Hutchins SJ, Steel PG, Edwards R (2013) Key role for a glutathione transferase in multiple-herbicide resistance in grass weeds. *Proc Natl Acad Sci USA* 110:5812–5817
- Davey JW, Blaxter ML (2010) RAD-Seq: next-generation population genetics. *Brief Funct Genomics* 9:416–423
- Davis AS, Schutte BJ, Hager AG, Young BG (2015) Palmer amaranth (*Amaranthus palmeri*) damage niche in Illinois soybean is seed limited. *Weed Sci* 63:658–668

- Gaines TA, Ward SM, Bukun B, Preston C, Leach JE, Westra P (2012) Interspecific hybridization transfers a previously unknown glyphosate resistance mechanism in *Amaranthus* species: interspecific transfer of glyphosate resistance. *Evol Appl* 5:29–38
- Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9:e90346
- Grant WF (1959) Cytogenetic studies in *Amaranthus*. I. Cytological aspects of sex determination in dioecious species. *Can J Bot* 37:413–417
- Harkess A, Zhou J, Xu C, Bowers JE, Van der Hulst R, Ayyampalayam S, Mercati F, Riccardi P, McKain MR, Kakrana A, Tang H, Ray J, Groenendijk J, Arikrit S, Mathioni S, Nakano M, Shan H, Telgmann-Rauber A, Kanno A, Yue Z, Chen H, Li W, Chen Y, Xu X, Zhang Y, Luo S, Chen H, Gao J, Mao A, Pires JC, Luo M, Kundra D, Wing RA, Meyers BC, Yi K, Kong H, Lavrijnsen P, Sunseri F, Falavigna A, Ye Y, Leebens-Mack JH, Chen G (2017) The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat Commun* 8:1279
- Henry IM, Akagi T, Tao R, Comia L (2018) One hundred ways to invent the sexes: theoretical and observed paths to dioecy in plants. *Annu Rev Plant Biol* 69:553–575
- Hothorn T, Hornik K, Van de Wiel MA, Zeileis A (2006) A Lego system for conditional inference. *Am Stat* 60:257–263
- Hothorn T, Hornik K, Van de Wiel MA, Zeileis A (2008) Implementing a class of permutation tests: the coin package. *J Stat Softw* 28:1–23
- Kafkas S, Khodaeiaminjan M, Güney M, Kafkas E (2015) Identification of sex-linked SNP markers using RAD sequencing suggests ZW/ZZ sex determination in *Pistacia vera* L. *BMC Genomics* 16:98
- Kalendar R, Khassenov B, Ramankulov Y, Samuilova O, Ivanov KI (2017) FastPCR: an in silico tool for fast primer and probe design and advanced sequence analysis. *Genomics* 109:312–319
- Komai F, Masuda K (2004) Plasticity in sex expression of spinach (*Spinacia oleracea*) regenerated from root tissues. *Plant Cell Tiss Org* 78:285–287
- Kreiner J, Giacomini D, Waithaka B, Bemm F, Lanz C, Hildebrandt J, Regalado J, Sikkema P, Tranel P, Weigel D, Stinchcombe J, Wright S (2018) Multiple modes of convergent adaptation in the spread of glyphosate-resistant *Amaranthus tuberculatus*. <https://www.biorxiv.org/node/147743.full>. Accessed: January 15, 2019.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18
- Murray MJ (1940) The genetics of sex determination in the family *Amaranthaceae*. *Genetics* 25:409–431
- National Academies of Sciences, Engineering, and Medicine (2016) *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*. Washington, DC: National Academies Press. Pp 56–58
- Neve P (2018) Gene drive systems: do they have a place in agricultural weed management? *Pest Manag Sci* 74:2671–2679
- Palaiokostas C, Bekaert M, Davie A, Cowan ME, Oral M, Taggart JB, Gharbi K, McAndrew BJ, Penman DJ, Migaud H (2013) Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *BMC Genomics* 14:566
- Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253
- Puterova J, Kubat Z, Kejnovsky E, Jesionek W, Cizkova J, Vyskot B, Hobza R (2018) The slowdown of Y chromosome expansion in dioecious *Silene latifolia* due to DNA loss and male-specific silencing of retrotransposons. *BMC Genomics* 19:153
- Sauer JD (1957) Recent migration and evolution of the dioecious amaranths. *Evolution* 11:11–31
- Stetter MG, Schmid KJ (2017) Analysis of phylogenetic relationships and genome size evolution of the *Amaranthus* genus using GBS indicates the ancestors of an ancient crop. *Mol Phylogenet Evol* 109:80–92
- ThermoFisher Scientific (n.d.) Primer Designer Tool for PCR & Sanger Sequencing. <https://www.thermofisher.com/us/en/home/life-science/sequencing/sanger-sequencing/pre-designed-primers-pcr-sanger-sequencing.html>. Accessed: March 1, 2018.
- Tranel PJ, Riggins CW, Bell MS, Hager AG (2011) Herbicide resistances in *Amaranthus tuberculatus*: a call for new options. *J Agric Food Chem* 59:5808–5812
- Tranel PJ, Trucco F (2009) 21st-century weed science: a call for *Amaranthus genomics*. Pages 53–81 in Stewart CN Jr, ed. *Weedy and Invasive Plant Genomics*. Ames, IA: Blackwell
- Trucco F, Tatum T, Rayburn AL, Tranel PJ (2009) Out of the swamp: unidirectional hybridization with weedy species may explain the prevalence of *Amaranthus tuberculatus* as a weed. *New Phytol* 184:819–827
- Ward SM, Webster TM, Steckel LE (2013) Palmer amaranth (*Amaranthus palmeri*): a review. *Weed Technol* 27:12–27
- Waselkov KE, Olsen KM (2014) Population genetics and origin of the native North American agricultural weed waterhemp (*Amaranthus tuberculatus*; *Amaranthaceae*). *Am J Bot* 101:1726–1736
- Wu C, Davis AS, Tranel PJ (2018) Limited fitness costs of herbicide-resistance traits in *Amaranthus tuberculatus* facilitate resistance evolution: fitness costs of herbicide resistance. *Pest Manag Sci* 74:293–301
- Xin Z, Chen J (2012) A high throughput DNA extraction method with high yield and quality. *Plant Methods* 8:26
- You FM, Huo N, Gu Y, Luo M, Ma Y, Hane D, Lazo GR, Dvorak J, Anderson OD (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* 9:253
- 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 Physiol* 166:1106–1118