Development, Characterization, and Linkage Mapping of Single Nucleotide Polymorphisms in the Grain Amaranths (Amaranthus sp.)

P. J. Maughan,* S. M. Smith, D. J. Fairbanks, and E. N. Jellen

Abstract
The grain amaranths (Amaranthus sp.) are important pseudocereals native to the New World. During the last decade they have garnered increased international attention for their nutritional quality, tolerance to abiotic stress, and importance as a symbol of indigenous cultures. We describe the development of the first single nucleotide polymorphism (SNP) assays for amaranth. In addition, we report the characterization of the first complete genetic linkage map in the genus. The SNP assays are based on KASP genotyping chemistry and were detected using the Fluidigm dynamic array platform. A diversity screen of 41 accessions of the cultivated amaranth species and their putative ancestor species (Amaranth hybridus L.) showed that the minor allele frequency (MAF) of these markers ranged from 0.05 to 0.5 with an average MAF of 0.27 per SNP locus. One hundred and forty-one of the SNP loci were considered highly polymorphic (MAF ≥ 0.3). Linkage mapping placed all 411 markers into 16 linkage groups, presumably corresponding to each of the 16 amaranth haploid chromosomes. The map spans 1288 cM with an average marker density of 3.1 cM per marker. The work reported here represents the initial first steps toward the genetic dissection of agronomically important characteristics in amaranth.

The genus Amaranthus (Caryophyllales: Amaranthaceae) contains three domesticated grain species, collectively referred to as the grain amaranths (A. hypochondriacus L., A. cruentus L., and A. caudatus L.; Sauer, 1976). These species, along with their putative progenitor species (A. hybridus L., A. quitensis Kunth, and A. powellii S. Watson) are classified in what is termed the A. hybridus complex and are considered paleo-allotetraploids (2n = 4x = 32; Greizerstein and Poggio, 1994; Greizerstein and Poggio, 1995; Pal and Khoshoo, 1982). Amaranth was a major domesticated food crop of the pre-Columbian New World civilizations, likely having been domesticated multiple times over thousands of years ago (Mallory et al., 2008; Sauer, 1950). The importance of the amaranths to these ancient civilizations cannot be overstated; indeed, the Spaniards recorded that the Aztec emperor Moctezuma II required a tribute of approximately 200,000 bushels (5,727 t) per year of amaranth, an amount nearly equal to the annual maize tribute (Sauer, 1950, 1967, 1993). At the time of the Spanish conquest, cultivation of the grain amaranths was suppressed due to their deeply rooted use in indigenous religious practices (Iturbide and Gispert, 1994; Sauer, 1976, 1993). In the last few decades the grain amaranths have begun to reclaim some of their importance, largely due to the potential nutritional impact of their seed consumption on human health (Bressani et al., 1992; Tucker, 1986).

Amaranth grain is higher in fiber (8%) and fat (7 to 8%) than most cereals (Breene, 1991; Pedersen et al., 1987). Crude grain protein content ranges from 12.5 to 22.5% on a dry matter basis (Becker et al., 1981; Bressani et al., 1987, 1992; Tucker, 1986).

Published March, 2011
1989; Pedersen et al., 1987), and the crude protein content is relatively rich in the essential amino acid lysine normally limited in other cereal crops (0.73 to 0.84% of amaranth’s total protein content; Bressani et al., 1987). Amaranth oil is also highly nutritious, containing a relatively high content of squalene (7–8%; Bressani et al., 1987), which has been shown to reduce cholesterol levels in humans (Berger et al., 2003; Martirosyan et al., 2007). The grain amaranths exhibit C₄ photosynthesis and grow rapidly under heat and drought stress, and they tolerate a variety of unfavorable abiotic conditions, including high salinity, acidity, or alkalinity, making them uniquely suited for subsistence agricultural. By implication, amaranth has the potential for significant impact on malnutrition (Emokaro et al., 2007).

Genetic markers are essential tools for modern plant breeding research programs (Eathington et al., 2007). They are particularly important for germplasm conservation, for core-collection characterization (Diwan et al., 1995; Tanksley and McCouch, 1997), and in breeding applications, such as marker-assisted selection (MAS). The first step toward the development of genetic markers for amaranth was the discovery and characterization of 179 microsatellite markers by Mallory et al. (2008). Unfortunately, only 37 of microsatellite markers segregated in their intraspecific A. cruentus F₂ mapping population resulting in a sparsely populated linkage map. A significant advance in the number of available markers occurred in 2009 when Maughan et al. (2009) reported the utilization of a novel genomic reduction strategy linked with next-generation sequencing to identify 27,658 putative single nucleotide polymorphisms (SNPs) among four diverse amaranth accessions. Single nucleotide polymorphisms, defined as single base changes, are the most abundant type of DNA polymorphism found in eukaryotic genomes (Batley et al., 2003; Garg et al., 1999). Compared to microsatellites-based markers, SNPs exhibit a lower mutation rate and thus are less problematic in population genetic analyses (Xu et al., 2005). Single nucleotide polymorphisms have already been utilized in a wide array of research areas, including association studies (Andrew et al., 2008), conservation genetics (Cramer et al., 2008), and genetic diversity analysis (Kawuki et al., 2009) and are fast becoming the marker system of choice in marker-assisted plant breeding programs (Batley and Edwards, 2007).

The goals of this project were to (i) develop the first large scale set of functional SNP assays for amaranth, (ii) evaluate the informativeness (minor allele frequency) of the SNPs on a diversity panel consisting of the three grain amaranth species and their putative wild progenitor (A. hybridus), and (iii) construct of the first complete genetic linkage map for the genus.

Materials and Methods

Plant Materials and DNA Extraction

The SNPs were used to genetically characterize a set of 41 diverse genotypes from the A. hybridus complex, including 11 A. caudatus accessions, 10 A. cruentus accessions, 10 A. hypochondriacus accessions, and 10 A. hybridus accessions. Additionally, two accessions of both A. powellii and A. retroflexus L. and a single accession of A. tuberculatus (Moq.) J. D. Sauer were included in the diversity panel. These species were included to determine the utility of the SNP assays in these more distantly related but important weedy relatives (Table 1). All seed samples were obtained from the USDA collection (USDA, Iowa State University, Ames, IA; Table 1). For linkage analysis, an interspecific F₂ population was developed from a cross of PI 481125 (maternal parent; A. hypochondriacus) and PI 642741 (A. caudatus). The F₂ population consisted of 134 plants produced by self-fertilizing a single F₁ plant. All plants were grown in greenhouses at Brigham Young University, Provo, UT; in 12-cm pots using Sunshine Mix II (Sun Grow, Bellevue, WA) supplemented with Osmocote fertilizer (Scotts, Marysville, OH). Plants were maintained at 18°C under broad-spectrum halogen lamps with a 12-h photoperiod.

Total genomic DNA was extracted from 30 mg of freeze-dried leaf tissue from a single plant for each sample (diversity panel and F₂ population) according to procedures previously described (Sambrook et al., 1989), with modifications described by Todd and Vodkin (1996). Extracted DNA was quantified using a Nanodrop sample (diversity panel and F₂ population) according to procedures previously described (Sambrook et al., 1989), with modifications described by Todd and Vodkin (1996). Extracted DNA was quantified using a Nanodrop (ND 1000 Spectrophotometer; NanoDrop Technologies Inc., Montchanin, DE) and diluted to 20 ng μl⁻¹ in 1/10 TE buffer (10 mM Tris and 1 mM ethylenediaminetetraacetic acid [EDTA] [pH 8.0]).

Single Nucleotide Polymorphism Primer Design

A total of 480 putative SNPs were selected for genotyping from the 11,038 SNPs in silico reported to be polymorphic between the parents of our mapping population (Maughan et al., 2009). These 480 putative SNPs were selected based on several parameters, specifically that they showed no significant homology to the RepeatMasker (v. 3.2.9 Arabidopsis) database (Smit et al., 2010) or to the Arabidopsis chloroplast (GenBank accession NC_000932 [Benson et al., 2009]) and mitochondrial (NC_001284) genomes as determined by BLASTn (E-value < 1 × 10⁻⁵) (Altschul et al., 1997). These parameters were used to remove putative SNPs that were potentially extranuclear in origin or matched repetitive DNA sequences. The remaining sequences were then processed by the primer design software PrimerPicker (KBiosciences, 2009) using default design parameters and the first 480 SNPs were selected for primers synthesis. Primers were synthesized by Integrated DNA Technologies Inc. (Iowa City, IA). Primer sequence information for each of the functional KASPar SNP assays is provided in Supplemental Table S1.

Single Nucleotide Polymorphism Genotyping

The SNPs were genotyped by competitive allele-specific polymerase chain reaction (PCR) KASPar chemistry (KBioscience Ltd., Hoddesdon, UK) using the Fluidigm (Fluidigm Corp., South San Francisco, CA) nanofluidic 96.96 dynamic array (Wang et al., 2009). For genotyping on the 96.96 dynamic array chip using the KASPar chemistry, a 5-μL sample mix, consisting of 2.25 μl genomic
DNA (20 ng μL⁻¹), 2.5 μL of 2x KBiosciences Allele Specific PCR (KASP) reagent Mix (KBiosciences Ltd.), and 0.25 μL of 20x GT sample loading reagent (Fluidigm Corp., South San Francisco, CA) was prepared for each DNA sample. Similarly, a 4 μL 10x KASP Assay, containing 0.56 μL of the KASP assay primer mix (allele specific primers at 12 μM and the common reverse primer at 30 μM), 2 μL of 2x Assay Loading Reagent (Fluidigm Corp., South San Francisco, CA), and 1.44 μL DNase-free water was prepared for each SNP assay. The assay mix and sample mix were then loaded onto a 96.96 dynamic array chip, mixed, and thermal cycled using an integrated fluidic circuit (IFC) Controller HX and FC1 thermal cycler (Fluidigm Corp., South San Francisco, CA) according to the manufacturer's protocols. Thermal cycling consisted of an initial thermal mix cycle (70°C for 30 min; 25°C for 10 min) a hot-start Taq polymerase activation step (94°C for 15 min) followed by a touchdown amplification protocol as follows: 10 cycles of 94°C for 20 sec, 65°C for 1 min (decreasing 0.8°C per cycle), 26 cycles of 94°C for 20 sec, 57°C for 1 min, and then hold at 20°C for 30 sec. End-point fluorescent images of the chip were acquired on an EP-1 imager (Fluidigm Corp., South San Francisco, CA) and the data analyzed with Fluidigm SNP genotyping Analysis Software (Fluidigm, 2011).

### Single Nucleotide Polymorphism

#### Diversity Data Analysis

Alleles for each SNP were scored as present, absent, or missing (failed to amplify) and converted into a binary matrix to determine minor allele frequencies (MAFs) for each SNP locus. The genetic distance among genotypes was calculated based on the matrices of allele frequencies using Nei distance (Nei and Li, 1979). The clustering criterion used was neighbor joining and the resulting dendrogram was unrooted. Robustness of the topology of the cladogram was evaluated by bootstrap analysis (1000 replicates) of the data set.

#### Genetic Linkage Map Construction

Marker segregation was analyzed for conformity to Mendelian ratios expected in an F₂ population using a chi-squared test. Markers were initially grouped based on independence logarithm of odds (LOD) scores ≥ 5.0 using the G2 statistic as calculated by JoinMap 4 for recombination frequency (Van Ooijen, 2006). Markers within groups were then ordered using the regression mapping algorithms as described by Stam (1993), with the modification that the squares of the LODs are used as weights, to assign more weight to informative loci. Successive rounds of marker placement were utilized to add loci to the map. After the addition of each locus, a ripple test was applied to test for goodness-of-fit and assure the optimal map order.

### Results and Discussion

#### Single Nucleotide Polymorphism Assay Development

Maughan et al. (2009) previously reported the identification of 11,038 putative SNPs between the accessions

<table>
<thead>
<tr>
<th>Name</th>
<th>Species</th>
<th>Geographical location</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI 481125</td>
<td>A. hypochondriacus L.</td>
<td>India†</td>
</tr>
<tr>
<td>PI 477914</td>
<td>A. hypochondriacus</td>
<td>Mexico</td>
</tr>
<tr>
<td>PI 540446</td>
<td>A. hypochondriacus</td>
<td>Mexico</td>
</tr>
<tr>
<td>PI 558499</td>
<td>A. hypochondriacus</td>
<td>Nebraska, United States</td>
</tr>
<tr>
<td>PI 619259</td>
<td>A. hypochondriacus</td>
<td>Nebraska, United States</td>
</tr>
<tr>
<td>PI 633589</td>
<td>A. hypochondriacus</td>
<td>Chihuahua, Mexico</td>
</tr>
<tr>
<td>PI 23369</td>
<td>A. hybíridus L.</td>
<td>Brazil</td>
</tr>
<tr>
<td>PI 23891</td>
<td>A. hybíridus</td>
<td>Czech Republic</td>
</tr>
<tr>
<td>PI 25132</td>
<td>A. hybíridus</td>
<td>Nigeria</td>
</tr>
<tr>
<td>PI 26852</td>
<td>A. hybíridus</td>
<td>Portugal</td>
</tr>
<tr>
<td>PI 500249</td>
<td>A. hybíridus</td>
<td>Zambia</td>
</tr>
<tr>
<td>PI 568179</td>
<td>A. hybíridus</td>
<td>Iowa, United States</td>
</tr>
<tr>
<td>PI 603886</td>
<td>A. hybíridus</td>
<td>Ohio, United States</td>
</tr>
<tr>
<td>PI 605351</td>
<td>A. hybíridus</td>
<td>Greece</td>
</tr>
<tr>
<td>PI 632247</td>
<td>A. hybíridus</td>
<td>North Carolina, United States</td>
</tr>
<tr>
<td>PI 636181</td>
<td>A. hybíridus</td>
<td>Delaware, United States</td>
</tr>
<tr>
<td>PI 572261</td>
<td>A. powelli subsp. bouchioni (Thell.) Costea &amp; Carretero</td>
<td>Germany</td>
</tr>
</tbody>
</table>

†All origin information is derived from the Germplasm Resources Information Network (2011). Several accessions were collected in the Old World although they originate in the Americas according to Sauer (1967).

‡Reclassified here as A. hypochondriacus based on data reported herein.

§Cv. Plainsman.

PI 481125 and PI 642741 using a genomic reduction strategy based on restriction-site conservation and

---

**Table 1. Amaranthus accessions used in the single nucleotide polymorphism (SNP) diversity assay screens.** Plant Introduction (PI) 481125 and PI 642741 are the parents of the mapping population.
The diversity panel consisted of 46 samples, representing seven *Amaranthus* species. Limiting the analysis to just the grain amaranths (*A. caudatus*, *A. cruentus*, and *A. hypochondriacus*; *n* = 31) and their putative wild ancestor (*A. hybridus*; *n* = 10), a total of 414 (828 alleles) of the SNP markers were polymorphic, producing clearly separated genotypic clusters that could be scored with high confidence. Since SNPs are predominantly biallelic, the maximum MAF value of a SNP is 0.5 (which occurs when both alleles are present at equal frequencies in the test population). Across the full panel of grain amaranth and *A. hybridus* accessions, MAF values ranged from 0.05 to 0.5 with an average MAF value of 0.27 per locus (Supplemental Table S1; Fig. 1). Considering a SNP with a MAF ≥ 0.3 to be highly polymorphic, 141 (34%) of the SNP loci were highly polymorphic (Table 2).

Within the *A. caudatus*, *A. cruentus*, *A. hypochondriacus*, and *A. hybridus* subgroups, a total of 136, 35, 186, and 263 SNP assays were polymorphic, respectively (Table 2). *Amaranthus hypochondriacus* showed the highest total number of polymorphic SNP markers, while *A. cruentus* showed the lowest genetic diversity of the grain species with only 35 polymorphic markers. The reduced level of genetic diversity observed in *A. cruentus* is consistent with other observations using different types of genetic markers, including microsatellites, restriction fragment length polymorphism (RFLP), isozyme, and amplified fragment length polymorphism (AFLP) (Chan and Sun, 1997; Mallory et al., 2008; Xu and Sun, 2001). Chan and Sun (1997) suggested that the decreased level of genetic diversity observed in *A. cruentus* might be a result of a specialized domestication process where only a small subset of the original *A. hybridus* population was subjected to intense artificial selection to select for specific agronomic characteristics. Mallory et al. (2008) speculated that the limited and uniform cultivation range of *A. cruentus* might have further reduced the level of genetic diversity within the species. Conversely, *A. hybridus*, the putative wild progenitor species of the grain amaranths, showed the most genetic diversity of all the species included in the diversity panel. Indeed, *A. hybridus* showed a sevenfold increase in polymorphic loci when compared to *A. cruentus* and approximately twice as many polymorphic SNPs when compared to the other grain amaranths species (*A. caudatus* and *A. hypochondriacus*). The higher genetic diversity observed within *A. hybridus* is consistent with an expectation that a wild progenitor species should be genetically more diverse than a derived domesticated species, specifically as a result of genetic drift and selection (Hilu, 1995). Moreover, the observation that nearly 94% (388) of the grain amaranth sequence-based SNP assays worked with *A. hybridus* samples is notable in that it further confirms the close ancestral relationship between the grain amaranths and *A. hybridus*.

### Table 2. Summary analysis of single nucleotide polymorphism (SNP) marker results, including sample size, total number of polymorphic SNP loci, total alleles observed, minor allele frequency (MAF) range and average, and total highly polymorphic microsatellites. A total of 480 putative SNP markers were initially screened.

<table>
<thead>
<tr>
<th></th>
<th><em>Amaranthus hypochondriacus</em></th>
<th><em>A. cruentus</em></th>
<th><em>A. caudatus</em></th>
<th><em>A. hybridus</em></th>
<th><em>A. hybridus</em> complex†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>11</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>41</td>
</tr>
<tr>
<td>Polymorphic SNPs</td>
<td>186</td>
<td>35</td>
<td>136</td>
<td>263</td>
<td>414</td>
</tr>
<tr>
<td>MAF range</td>
<td>0.05–0.5</td>
<td>0.05–0.5</td>
<td>0.05–0.5</td>
<td>0.05–0.5</td>
<td>0.05–0.5</td>
</tr>
<tr>
<td>Average MAF</td>
<td>0.20</td>
<td>0.18</td>
<td>0.22</td>
<td>0.24</td>
<td>0.27</td>
</tr>
<tr>
<td>Highly polymorphic‡</td>
<td>41</td>
<td>7</td>
<td>43</td>
<td>84</td>
<td>141</td>
</tr>
</tbody>
</table>

†Includes *A. hypochondriacus*, *A. cruentus*, *A. caudatus*, and *A. hybridus* accessions.

‡Highly polymorphic SNP: MAF ≥ 0.3.
The grain amaranths are members of the genus *Amaranthus* (subfamily Amaranthoideae), which contains several other important plant species including several of the most damaging weedy species in the United States collectively referred to as the “pigweeds” (Basu et al., 2004; Wassom and Tranel, 2005). Various studies have already shown the utility of molecular markers for clarifying taxonomic relationships among the weedy species of the genus *Amaranthus* (Wassom and Tranel, 2005; Wetzel et al., 1999), yet taxonomic questions still exist and the need for additional, easy to use, genetic markers remains high. We evaluated the transferability and utility of these SNP assays on three distinctly related weedy amaranth species, specifically *A. powellii* (Powell amaranth), *A. retroflexus* (redroot pigweed), and *A. tuberculatus* (Moq.) J. D. Sauer (common waterhemp). Of the 414 SNP assays that were polymorphic in the grain amaranths and their putative ancestor (*A. hybridus*), 256 (62%) of the assays produced high confidence genotypic calls in both accessions of *A. powellii*, and *A. retroflexus*, whereas only 158 (38%) produced a high confidence genotypic call with the single *A. tuberculatus* accession included in the diversity panel (Supplemental Table S1). Between the two *A. powellii* accessions included in the diversity panel, 26 (10%) of the SNPs were polymorphic, while only three (1%) were polymorphic between the two *A. retroflexus* accessions. We note that the origins of the two accessions included in the analysis for both *A. powellii* and *A. retroflexus* accessions are geographically distinct (Table 1), suggesting that (i) the *A. retroflexus* population is potentially much less diverse than the *A. powellii* population—an intriguing proposition considering that *A. retroflexus* is among the most widely distributed weeds in the world (Holm et al., 1997) or, perhaps more likely, that (ii) the SNPs identified from the cultivated amaranths produce an inherent bias, such that species that are taxonomically closer (*A. powellii*) have a higher probability of sharing the genetic polymorphism (via an orthologous relationship).

**Phylogenetic Analysis**

Several hypotheses have been proposed for the evolutionary origins of the grain amaranths. The first hypothesis is based on geography and suggests that all three grain amaranths evolved independently, specifically *A. caudatus* from *A. quitensis* in South America, *A. cruentus* from *A. hybridus* in Central America, and *A. hypochondriacus* from *A. powellii* in Mexico (Sauer, 1967, 1976). The second hypothesis, based on plant and seed morphology, suggests that *A. hybridus* gave rise to *A. cruentus*, which in turn hybridized with *A. powellii* and *A. quitensis* to give rise to *A. hypochondriacus* and *A. caudatus*, respectively (Sauer, 1967, 1976). A third hypothesis, proposed more recently by Mallory et al. (2008), suggested that *A. hybridus* is the progenitor species for all three domesticated species but that each was derived from independent domestication events from genetically differentiated populations of *A. hybridus*. Our results support the designation of *A. hybridus* as the progenitor species of all three grain amaranth species. Indeed, neighbor-joining analysis reveals that *A. caudatus*, *A. cruentus*, and *A. hypochondriacus* are monophyletic, while *A. hybridus* was polyphyletic with *A. hybridus* accessions in each of the three grain amaranth monophyletic clades (Fig. 2). We note that *A. powellii* (a previously suggested progenitor) formed a monophyletic group distinct from any of the grain amaranths. Obviously a larger investigation, including substantially more accessions of the grain amaranth progenitor species (*A. hybridus*, *A. powellii*, and *A. quitensis*), is still needed to finely dissect the origins of the grain amaranths. However, we expect that taxonomic identification of these *Amaranthus* species may be ambiguous due to reciprocal gene flow via outcrossing in regions where the species are sympatric—a situation we have observed in Peruvian and Mesoamerican centers of origin of the crop species. The transferability, ease of use, and the highly polymorphic nature of the SNP assays reported here should facilitate such an investigation.

**Linkage Map Construction**

The parents of the mapping population were initially chosen based on the published research by Maughan et al. (2009) that showed that PI 481125 and PI 642741 were genetically diverse. The paternal parent, PI 642741, has an easily identifiable dominant phenotypic marker (red stem color) that facilitated the identification of true hybrid F1 plants. Both accessions are classified as *A. caudatus* accessions within the USDA Germplasm Resources Information Network (GRIN) system (Germplasm Resources Information Network, 2011); however, our phylogenetic analysis clearly places PI 481125 within the *A. hypochondriacus* clade (Fig. 2). An independent genotypic analysis of a second plant sample of PI 481125 also grouped with the *A. hypochondriacus* accessions, suggesting that PI 481125 was originally misclassified and should be reclassified as *A. hypochondriacus*. This reclassification also agrees well with the levels of genetic diversity identified by Maughan et al. (2009), where interspecific comparisons should be genetically more diverse than intraspecific comparisons. Based on these observations, our mapping population should be categorized as interspecific.

A total of 419 SNP loci were genotyped using the KASP genotyping chemistry on a Fluidigm IFC. Included in the genotyping experiment were 134 F1 individuals, the parental genotypes, and a synthetic heterozygote (consisting of eqimolar quantities of DNA from the two parental samples). Of the 419 SNP assays, 411 (98%) produced genotypic clusters that could be easily scored. The remaining eight assays were lost due to issues associated with loading the IFC. Of these 411 assays, 373 (91%) produced three clearly separated clusters and were scored in a codominant fashion (1:2:1; Fig. 3), while the remaining 9% produced only two clearly separated clusters and were scored in a dominant fashion (3:1; Fig. 3). The dominant SNP assays are likely due to preferential amplification of one of the SNP allele-specific primers in the heterozygous samples (Walsh et al., 1992).

To validate the genotyping process, eight random SNPs were regenotyped and compared across all 134 F1
individuals. Thus a total of 1072 data points, from two separate fluidic chips, were compared for genotyping accuracy. Twenty-five (2%) of the comparisons contained a missing value and thus could not be included in the comparative analysis. Of the remaining 1047 comparisons, 1021 (98%) were identical matches while 26 (2%) were mismatches, of which all were conflicts between homozygous to heterozygous calls. Interestingly, 35% of the conflicts were accounted for by just three of the F$_1$ individuals, suggesting that the genotypic conflicts might be related to a DNA source and not specifically to the genotyping methodology per se. Indeed, the 17 DNA samples with genotypic conflicts averaged 15 times more missing data than samples without any observed genotype conflicts ($n = 117$), supporting the conclusion that the genotyping conflicts were likely the result of problematic DNA samples. Consequently, the genotypic data for all 17 individuals were removed before linkage map construction.

At a minimum LOD score of 5.0, pairwise linkage analysis grouped all 411 SNP markers into 16 linkage groups, presumably corresponding to each of the amaranth haploid chromosomes ($2n = 32$; Fig. 4). The distribution of the markers within the linkage groups varied from 9 to 47 SNPs per linkage group. Regression mapping of the pairwise linkage groups successfully ordered all SNP markers within their respective linkage groups. The centiMorgan (cM) distance, corrected with the Kosambi mapping function, spanned by the SNP markers in the linkage groups ranged from a low of 23 to 144 cM. The total map consisted of 411 SNP loci and spanned 1288 cM. The largest interval between two linked markers was 27 cM on linkage group 12, while the average distance between all loci was 3.1 cM. Most intervals (93%) were less than 10 cM apart.

Of the 411 SNPs utilized for mapping, 22 (9.7%) showed significant segregation distortion ($p < 0.0001$). Since the map is based on an interspecific mapping population ($A. caudatus \times A. hypochondriacus$), segregation distortion was not unexpected. Indeed, segregation distortion in interspecific crosses has been reported to reach levels as high as 68.5% of the markers (Paterson et

Figure 2. Unrooted neighbor joining tree showing the genetic relationship among accessions of the *Amaranthus hybridus* complex (*A. caudatus*, *A. cruentus*, *A. hypochondriacus*, and *A. hybridus*) genotypes based on single nucleotide polymorphism (SNP) marker data. Bootstrap support values are given at each node. Individuals in the tree are identified by their abbreviated species. *Amaranthus hybridus* accessions are identified with stars (*). The boxed accession, PI 481125, was originally classified as *A. caudatus*. 

MAUGHAN ET AL.: SNPS IN THE GRAIN AMARANTHS
Skewed SNP markers mapped to a total of nine different clusters on seven linkage groups (Fig. 4). The presence of markers clusters skewed to a single parental genotype has been attributed to chromosomal regions containing possible gametophytic or zygotic viability factors (Lu et al., 2002; Zamir and Tadmor, 1986) and/or underlying genetic factors (i.e., quantitative trait loci [QTL]) conferring a selective advantage for the particular growing conditions used to produce the mapping population. We note that significant morphological differences in seedling morphology, growth rate, and seed production were observed among the F2 plants.

Flanking sequences for each of the 411 mapped SNPs were compared to the GenBank refseq_protein database (Benson et al., 2009) using BLASTX (Altschul et al., 1997). Twenty-four (5.8%) of query sequences returned significant (E-value < 1 × 10^{-10}) homologies to the refseq_protein database and were mapped to 11 of the 16 linkage groups. The low homology to the refseq_protein database is consistent with the genome reduction methodology, which randomly samples the genome. Significant homologies to well annotated genes included SPK1-guanylnucleotide exchange factor activity (NP_193367), ATMRP10-ATPase (NP_191829), WRKY20-transcription factor (NP_567752), SULTR3.4-sulfate transporter (NP_188220), RAP2.2-DNA transcription factor (NP_566482), OVA1-methionine-tRNA ligase activity (NP_191100), and ATCUL2-ubiquitin protein ligase binding (NP_171797). Considering a refseq_protein database hit rate of 5.8%, we conclude that many of the 11,038 in silico SNP loci that were originally reported by Maughan et al. (2009) should be located in or very near (±150 bp) gene sequences, suggesting that a subsequent linkage map of amaranth could be based almost solely on genic-SNP sequences. Indeed, after removal of the SNPs with significant homologies to the Arabidopsis extranuclear genomes (eight SNP loci) and RepeatMasker database (272 SNP loci) (Smit et al., 2010), a BLASTX analysis of the 11,038 in silico SNP loci identified 711 SNPs with significant sequence homologies (E-value < 1 × 10^{-10}) to genic sequences in the refseq_protein database. The mapping population is being selfed (currently at F_{2.5}) and expanded to 200 individuals to form a recombinant inbred line population that should
provide an immortalized mapping population for the amaranth community as well as provide the first population readied for QTL analysis.

**Conclusions**

We report the first high-density, complete genetic linkage map of amaranth. The SNP markers and linkage map reported are essential steps toward the development of marker-assisted selection programs for recalcitrant traits of agronomic importance in amaranth. The SNP assays were developed on the KBioscience KASPar genotyping chemistry using a Fluidigm IFC Access array. The utilization of this chemistry combined with the nano-fluidic Fluidigm chip reduced the overall data point cost to US$0.05 per data point—an important feature considering that the implementation of marker-assisted breeding strategies often requires the generation of thousands of data points per population (Eathington et al., 2007). Compared to other markers systems (e.g., AFLPs or simple sequence repeats [SSRs]), the SNP assays reported here are relatively inexpensive and easy to genotype. Indeed, a single 96.96 Fluidigm IFC is capable of producing 9216 genotypic data points in a single run (~3 h) with little technical expertise, and since each genotyping reaction is done on a nanoliter scale, the consumable reagent costs (i.e., Taq polymerase and primers) is only $0.001 data point (the remainder of the cost is the IFC). If a Fluidigm EP1 system is unavailable (a significant capital investment), the same KASPar SNP assays can be read on a standard fluorescence resonance energy transfer (FRET) plate reader—an important consideration for laboratories in the developing world, where amaranth is cultivated and where capital equipment may be limited. Worthy of note is the upfront cost of the allele specific

Figure 4. A 16 group linkage map constructed from an interspecific *Amaranthus hypochondriacus* × *A. caudatus* *F₂* population (*2n* = 32). Distances are shown in centiMorgans (cM) corrected with the Kosambi mapping function. Single nucleotide polymorphism (SNP) loci showing segregation distortion (*p* < 0.001) to PI 642741 or PI 481125 are identified with blackened or shaded boxes, respectively.
and common reverse primers needed for each of the KASPar genotyping assays. At the current commercial minimum synthesis scale (25 nM), each KASPar genotyping primer set costs approximately $11 to manufacture and is sufficient to run approximately six million Fluidigm IFC-based genotyping reactions.

The SNP markers reported here will be of particular value in ongoing efforts to characterize extensive amaranth germplasm collections and the development of core collections needed for existing and emerging amaranth breeding programs in the Andes, Mexico, Asia, and sub-Saharan Africa (Diwan et al., 1995; Tanksley and McCouch, 1997).

Supplemental Information Available
Supplemental material is available free of charge at http://www.crops.org/publications/cs.

Supplemental Table S1 provides the GenBank dbSNP accession ID, polymorphism type, KASPar primer sequences (A1, A2, and Common Reverse), minimum allele frequency (MAF), and cross species amplification (CSA) for all 419 functional amaranth single nucleotide polymorphism (SNP) assays tested.

Acknowledgments
This research was supported by grants from the Erza Taft Benson Agriculture and Food Institute and the Holmes Family Foundation. We gratefully acknowledge D. Brenner (USDA-ARS, Iowa State University) for taxonomic suggestions and seed contribution. We are also grateful for the technical help of Zachary Danielson during DNA extraction.

References


Fluidigm. 2011. Fluidigm SNP Genotyping Analysis v.3.0.2. Fluidigm Corporation, South San Francisco, CA.


Maughan, P.J., S.M. Yourstone, E.N. Jellen, and J.A. Udall. 2009. SNP discovery via genomic reduction, barcoding and 454-pyro...


KBiosciences. 2009. PrimerPicker Lite for KASPar v.0.26. KBiosciences Ltd., Hoddesdon, UK.


