

## **Incomplete domestication of South American grain amaranth (*Amaranthus caudatus*) from its wild relatives**

Markus G. Stetter, Thomas Müller and Karl J. Schmid

Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Fruwirthstr. 21, 70599 Stuttgart / Germany

### **Keywords**

Amaranthus, genotyping-by-sequencing, genetic diversity, domestication, flow cytometry, orphan crop

### **Corresponding Author**

Name: Karl Schmid, Prof. Dr.

Address: Institute of Plant Breeding, Seed Science and Population Genetics (350)

University of Hohenheim

Fruwirthstraße 21

D-70599 Stuttgart / Germany

Tel: +49 711 459 23487

Fax: +49 711 459 24458

E-Mail address: [karl.schmid@uni-hohenheim.de](mailto:karl.schmid@uni-hohenheim.de)

Running title: Incomplete domestication of *A. caudatus*

## Incomplete Amaranth domestication

---

### 1 **Abstract**

2 Grain amaranth is a pseudo-cereal and an ancient crop of Central and South America. Of the  
3 three species of grain amaranth, *Amaranthus caudatus* is mainly grown in the Andean region.  
4 Several models of domestication were proposed including a domestication from the wild rela-  
5 tives *A. hybridus* or *A. quitensis*. To investigate the domestication history of *A. caudatus* and  
6 its relationship to the two wild relatives, we used genotyping-by-sequencing (GBS) to geno-  
7 type 119 amaranth accessions from the Andean region. We determined the genome sizes of  
8 the three species and compared phenotypic variation in two domestication-related traits, seed  
9 size and seed color. We show that the population genetic analysis based on 9,485 SNPs  
10 revealed very little genetic differentiation between the two wild species, suggesting they are  
11 the same species, but a strong differentiation between wild and domesticated amaranths. *A.*  
12 *caudatus* has a higher genetic diversity than its wild relatives and about 10% of accessions  
13 showed a strong admixture between the wild and cultivated species suggesting recent gene  
14 flow. Genome sizes and seed sizes were not significantly different between wild and domesti-  
15 cated amaranths, although a genetically distinct cluster of Bolivian accessions had significantly  
16 larger seeds. Taken together our analysis suggests that grain amaranth is an incompletely do-  
17 mesticated species, either because it was not strongly selected or because high levels of gene  
18 flow from its sympatric wild relatives counteract the fixation of key domestication traits in the  
19 domesticated *A. caudatus*.

## Incomplete Amaranth domestication

---

### 20 Introduction

21 The genus *Amarantus* L. comprises between 50 and 75 species and is distributed worldwide  
22 (Sauer, 1967; Costea & DeMason, 2001). Four species are cultivated as grain amaranths or  
23 leaf vegetables (Sauer, 1967; Brenner, 2000). The grain amaranths *Amaranthus caudatus*,  
24 *Amaranthus cruentus* and *Amaranthus hypochondriacus* originated from South and Central  
25 America. Amaranth is an ancient crop, archaeological evidence in Northern Argentina sug-  
26 gested that wild amaranth seeds were collected and used for human consumption during the  
27 initial mid-Holocene (8,000 - 7,000 BP; Arreguez *et al*, 2013). In the Aztec empire, amaranth  
28 was a highly valued crop and tributes were collected from the farmers that were nearly as high  
29 as for maize (Sauer, 1967). Currently, amaranth is promoted as a healthy food because of its  
30 favorable composition of essential amino acids and high micronutrient content.

31 The three grain amaranth species differ in their geographical distribution. *A. cruentus* and  
32 *A. hypochondriacus* are most common in Central America, whereas *A. caudatus* is cultivated  
33 mainly in South America. In the Andean region, *A. caudatus* grows in close proximity to the  
34 two wild *Amaranthus* species *A. hybridus* and *A. quitensis*, which are considered as potential  
35 ancestors (Sauer, 1967). Of these, *A. quitensis* was tolerated or cultivated in Andean home  
36 gardens and used for coloring in historical times.

37 Past research on the domestication of major crop plants revealed that crops from different plant  
38 families have similar domestication syndromes that include larger seeds, loss of seed shatter-  
39 ing, reduced branching, loss of seed dormancy and increased photoperiod insensitivity (Abbo  
40 *et al*, 2014; Hake & Ross-Ibarra, 2015). In addition to phenotypic changes, domestication  
41 strongly affected the structure of genetic diversity of domesticated plants and created a genetic  
42 signature of selection and drift because domestication is frequently associated with a strong ge-  
43 netic bottleneck (Doebley *et al*, 2006; Olsen & Wendel, 2013; Sang & Li, 2013; Nabholz *et al*,  
44 2014). The history of amaranth domestication is still under discussion. Sauer (1967) proposed  
45 two scenarios based on the morphology and geographic distribution of the different species.  
46 The first model postulates three independent domestication events, in which *A. hypochondria-*  
47 *cus* originated from *A. powellii*, *A. cruentus* from *A. hybridus*, and *A. caudatus* from *A. quitensis*.  
48 The second model proposes an initial domestication of *A. cruentus* from *A. hybridus* followed

## Incomplete Amaranth domestication

---

49 by a migration and intercrossing of *A. cruentus* with *A. powellii* in Central America and an intercrossing of *A. cruentus* with *A. quitensis* resulting in *A. caudatus* in South America. Another model based on SNP markers suggested that all three domesticated amaranths evolved from *Amaranthus hybridus*, but at multiple locations (Maughan *et al*, 2011). Most recently, Kietlinsky *et al*. (Kietlinski *et al*, 2014) proposed a single domestication *A. hybridus* in the Andes or in Mesoamerica and a subsequent spatial separation of two lineages leading to *A. caudatus* and *A. hypochondriacus* or two independent domestication events of *A. hypochondriacus* and *A. caudatus* from a single *A. hybridus* lineage in Central and South America. Taken together, the diversity of hypotheses indicates either a complex domestication history or insufficient data to strongly support a single model of domestication.

59 Despite its long history of cultivation, the domestication syndrome of cultivated amaranth is remarkably indistinct because it still shows strong photoperiod sensitivity and has very small shattering seeds (Sauer, 1967; Brenner, 2000). Other crops like maize that were cultivated at a similar time period in the same region exhibit the classical domestication syndrome (Sang & Li, 2013; Lenser & Theißen, 2013). This raises the question whether amaranth has a different domestication syndrome or whether genetic constraints, a lack of genetic variation or (agri)cultural reasons led to a distinct domestication pattern compared to other crops. The phenotypic analysis of amaranth domestication is complicated by the taxonomic uncertainty of wild amaranth species. Although *A. quitensis* was suggested to be the ancestor of *A. caudatus*, the state of *A. quitensis* as a separate species is under debate. Sauer (1967) classified it as species, but later it was argued that it is the same species as *A. hybridus* (Coons, 1978; Brenner, 2000). However, until today *A. quitensis* is treated as separate species and since genetic evidence for the status of *A. quitensis* as a separate species is based on few studies with limited numbers of markers, this topic is still unresolved (Mallory *et al*, 2008; Kietlinski *et al*, 2014).

73 The rapid development of sequencing technologies facilitates the large-scale investigation of the genetic history of crops and their wild relatives. Among available methods, reduced representation sequencing approaches such as genotyping-by-sequencing (GBS) allow a genome-wide and cost-efficient marker detection compared to whole genome sequencing (Elshire *et al*, 2011; Poland *et al*, 2012). Despite some biases associated with reduced representation sequencing, GBS and related methods are suitable and powerful approaches for studying inter-

## Incomplete Amaranth domestication

---

79 specific phylogenetic relationships (Cruaud *et al*, 2014) and intraspecific patterns of genetic  
80 variation in crop plants (Morris *et al*, 2013).

81 We used GBS and genome size measurements to characterize the genetic diversity and rela-  
82 tionship of cultivated *A. caudatus* and its putative wild ancestors *A. quitensis* and *A. hybridus*,  
83 and compared patterns of genetic structure with two domestication-related phenotypic traits  
84 (seed color and hundred seed weight). We tested whether domestication led to a reduction of  
85 genetic diversity and larger seed size in domesticated amaranth, and clarified the taxonomic  
86 relationship and gene flow with the close relatives. Our results indicated that *A. caudatus* has  
87 a history of domestication that may be considered as incomplete and is consistent with models  
88 of multiple domestication.

## 89 **Material and Methods**

### 90 **Plant material**

91 A total of 119 South American amaranth accessions of three *Amaranthus* species were ob-  
92 tained from the USDA gene bank (<http://www.ars-grin.gov/npgs/searchgrin.html>). Of  
93 these accessions, 89 were classified as *A. caudatus*, 17 as *A. hybridus*, seven as *A. quitensis*  
94 and six as interspecific hybrids according to the passport information (Figure S5). We selected  
95 the *A. caudatus* accessions based on the altitude of the collection site and focused on high-  
96 altitude populations (2,200 to 3,700 m). We further subdivided the species into populations  
97 according to their country of origin and included *A. caudatus* from Peru, Bolivia, *A. hybridus*  
98 from Peru, Bolivia, Ecuador, *A. quitensis* from Peru and Ecuador as well as hybrids from Peru  
99 and Bolivia. Accessions were planted in a field in Nürtingen (Germany), where a single young  
100 leaf of one representative plant per accession was sampled. From 12 accessions, three plants  
101 were sampled and sequenced individually for quality control.

## Incomplete Amaranth domestication

---

### 102 **DNA extraction and library preparation**

103 Genomic DNA was extracted using a modified CTAB protocol (Saghai-Marooft *et al*, 1984). The  
104 DNA was dried and dissolved in 50-100  $\mu$ l TE and diluted to 100 ng/ $\mu$ l for further usage. Two-  
105 enzyme GBS libraries were constructed with a modified protocol from the previously described  
106 two-enzyme GBS protocol (Poland *et al*, 2012). DNA was digested with a mix of 2  $\mu$ l DNA,  
107 2  $\mu$ l NEB Buffer 2 (NEB, Frankfurt/Germany), 1  $\mu$ l ApeKI (4U/ $\mu$ l, NEB), 1  $\mu$ l HindIII (20U/ $\mu$ l,  
108 NEB) and 14  $\mu$ l ddH<sub>2</sub>O for 2 hours at 37°C before incubating for 2 hours at 75°C. Adapters  
109 were ligated with 20  $\mu$ l of digested DNA 5  $\mu$ l ligase buffer (NEB), T<sub>4</sub>- DNA ligase (NEB), 4  $\mu$ l  
110 ddH<sub>2</sub>O and 20  $\mu$ l of adapter mix containing 10 $\mu$ l barcode adapter (0.3 ng/ $\mu$ l) and 10  $\mu$ l common  
111 adapter (0.3ng/ $\mu$ l). Samples were incubated at 22°C for 60 minutes before deactivating ligase  
112 at 65°C for 30 minutes. Subsequently, samples were cooled down to 4°C. For each sequencing  
113 lane 5 $\mu$ l of 48 samples with different barcodes were pooled after adapter ligation. Samples  
114 of the different species were randomized over the 3 pools and different barcode lengths. The  
115 12 replicated samples were in each pool. The pooled samples were purified with QIAquick  
116 PCR purification kit (Qiagen, Hilden/Germany) and eluted in 50  $\mu$ l elution buffer before PCR  
117 amplification of the pools. The PCR was performed with 10  $\mu$ l of pooled DNA, 25  $\mu$ l 2x Taq  
118 Master Mix (NEB), 2  $\mu$ l PCR primer mix (25pmol/ $\mu$ l of each primer) and 13  $\mu$ l ddH<sub>2</sub>O for 5 min  
119 at 72°C and 30 sec at 98°C before 18 cycles of 10 sec at 98°C, 30 sec at 65°C and 30 sec at  
120 72°C after the 18 cycles 5 min of 72°C were applied and samples were cooled down to 4°C.  
121 Samples were purified again with QIAquick PCR purification kit (Qiagen) and eluted in 30 $\mu$ l  
122 elution buffer. Three lanes with 48 samples per lane were sequenced on an Illumina HighScan  
123 SQ with single end and 105 cycles on the same flow cell (see supporting data).

### 124 **Data preparation**

125 Raw sequence data were filtered with the following steps. First, reads were divided into sepa-  
126 rate files according to the different barcodes using Python scripts. Read quality was assessed  
127 with fastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Due to lower  
128 read quality towards the end of the reads, they were trimmed to 90 bp. Low quality reads were  
129 excluded if they contained at least one N (undefined base) or if the quality score after trimming

## Incomplete Amaranth domestication

---

130 was below 20 in more than 10% of the bases. Data from technical replicates were combined  
131 and individuals with less than 10,000 reads were excluded from further analysis (Table S5).  
132 The 12 replicated samples were used to detect a lane effect with an analysis of variance.

### 133 **SNP calling and filtering**

134 Since no high quality reference genome for *Amaranthus* sp. was available for read mapping,  
135 we used Stacks 1.19, for the *de novo* identification of SNPs in GBS data (Catchen *et al*,  
136 2011, 2013). The pipeline provided for Stacks `denovo_map.pl` was used to call SNPs from  
137 the processed data. Highly repetitive GBS reads were removed in the `ustacks` program with  
138 option `-t`. Additionally, the minimum number of identical raw reads required to create a stack  
139 was set to three and the number of mismatches allowed between loci when processing a single  
140 individual was two. Four mismatches were allowed between loci when building the catalog.  
141 The catalog is a set of non redundant loci representing all loci in the accessions and used as  
142 reference for SNP calling. SNPs were called with the Stacks tool `populations 1.19` without  
143 filtering for missing data using option `-r 0`. One individual, PI 511754, was excluded from  
144 further analysis because it appeared to be misclassified. According to its passport information  
145 it belonged to *A. hybridus*, but with all clustering methods it was placed into a separate cluster  
146 consisting only of this individual, which suggested it belongs to a different species. Therefore,  
147 we repeated the SNP calling without this individual. The SNPs were further filtered with `vcftools`  
148 (Danecek *et al*, 2011), by allowing a maximum of 60% missing values per SNP position.

### 149 **Inference of genetic diversity and population structure**

150 Nucleotide diversity ( $\pi$ ) weighted by coverage was calculated with a Python script implement-  
151 ing the formula of Begun *et al* (2007) which corrects for different sampling depths of SNPs  
152 in sequencing data. The confidence interval of  $\pi$  was calculated by bootstrapping the calcu-  
153 lation 10,000 times. Mean expected ( $H_{exp}$ ) and observed ( $H_{obs}$ ) heterozygosities based on  
154 SNPs were calculated with the R package `adegenet 1.4-2` (Jombart & Ahmed, 2011). The  
155 inbreeding coefficient (F) was calculated as:

$$156 \quad \frac{H_{exp} - H_{obs}}{H_{exp}}$$

## Incomplete Amaranth domestication

---

157 Weir and Cockerham weighted  $F_{ST}$  estimates were calculated with vcftools (Weir & Cocker-  
158 ham, 1984; Danecek *et al*, 2011). To infer the population structure, we used ADMIXTURE for a  
159 model-based clustering (Alexander *et al*, 2009) and conducted the analysis with different num-  
160 bers of predefined populations ranging from  $K = 1$  to  $K = 9$  to find the value of  $K$  that was most  
161 consistent with the data using a cross-validation procedure described in the ADMIXTURE man-  
162 ual. To avoid convergence effects we ran ADMIXTURE 10 times with different random seeds  
163 for each value of  $K$ . As a multivariate clustering method, we applied discriminant analysis of  
164 principal components (DAPC) implemented in the R-package adegenet (Jombart *et al*, 2010;  
165 Jombart & Ahmed, 2011) and determined the number of principal components (PCs) used  
166 in DAPC with the `optim.a.score` method. To investigate the phylogenetic relationship of the  
167 species, we calculated an uncorrected neighbor joining network using the algorithm Neighbor-  
168 Net (Bryant & Moulton, 2004) as implemented in the SplitsTree4 program (Huson & Bryant,  
169 2006). The Euclidean distance was calculated from the genetic data to construct a neighbor  
170 joining tree, which was bootstrapped 1,000 times with the pegas R-package (Paradis *et al*,  
171 2004). The migration between genetic groups was modeled with TreeMix (Pickrell & Pritchard,  
172 2012). For the TreeMix analysis we used the groups that were identified by ADMIXTURE ( $K = 5$ )  
173 without an outgroup, and allowed 4 migration events, as preliminary runs indicates 4 migration  
174 events to be the highest number. The tree was bootstrapped 1,000 times.

## 175 Genome size

176 To compare genome sizes between *Amaranthus* species, we measured the genome size of  
177 22 *A. caudatus*, 8 *A. hybridus* and 4 *A. quitensis* accessions. Plants were grown for four  
178 weeks in the greenhouse before one young leaf was collected for cell extraction. A tomato  
179 cultivar (*Solanum lycopersicum* cv Stupicke) was used as internal standard, because it has a  
180 comparable genome size that has been measured with high accuracy (DNA content = 1.96 pg;  
181 Dolezel *et al*, 1992). Fresh leaves were cut up with a razor blade and cells were extracted  
182 with CyStain PI Absolute P (Partec, Muenster/Germany). Approximately 0.5 cm<sup>2</sup> of the sample  
183 leaf was extracted together with similar area of tomato leaf in 0.5 ml of extraction buffer. The  
184 DNA content was determined with CyFlow Space (Partec, Muenster/Germany) flow cytometer  
185 and analyzed with FlowMax software (Partec, Muenster/Germany). For each sample, 10,000



## Incomplete Amaranth domestication

---

186 particles were measured each time. Two different plants were measured for each accession.

187 The DNA content was calculated as:

$$188 \quad \text{DNA content 2C [pg]} = \text{genome size tomato} \times \frac{\text{fluorescence amaranth}}{\text{fluorescence tomato}}$$

189 and the genome size (in Mbp) was calculated as followed:

$$190 \quad \text{genome size 1C [Mbp]} = (0.978 * 10^3) \times \frac{\text{DNA content 2C [pg]}}{2}$$

191 The conversion from pg to bp was calculated with 1pg DNA =  $0.978 \times 10^9$  bp (Dolezel *et al*,  
192 2003). Means were calculated using R software (Team) and an ANOVA was performed to infer  
193 differences in genome size for the species.

### 194 **Seed color and hundred seed weight**

195 For each accession we calculated the hundred seed weight (HSW) by weighting three samples  
196 of 200 seeds. Seed color was determined from digital images taken with a binocular (at 6.5x  
197 magnification) and by visual comparison to the GRIN descriptors for amaranth (<http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?159>). There were three colors present in  
198 the set of accessions, white, pink, which also indicates a white seed coat and dark brown.  
199 To infer how the species, assigned genetic groups or seed color influenced seed size, we  
200 conducted an ANOVA. Differences were tested with a LSD test implemented in the R package  
201 *agricolae* (<http://tarwi.lamolina.edu.pe/~fmendiburu/>)  
202

## 203 **Results**

### 204 **SNP identification by GBS**

205 To investigate genome-wide patterns of genetic diversity in cultivated amaranth and two putative  
206 ancestors, we genotyped a diverse panel of 119 amaranth accessions from the Andean region  
207 that we obtained from the USDA genebank. The sequencing data generated with a two-enzyme  
208 GBS protocol, consisted of 210 Mio. raw reads with an average of 1.5 Mio. reads per accession  
209 (Supporting information S2). We tested for a lane effect of the Illumina flow cell, by sequencing

## Incomplete Amaranth domestication

---

210 12 individuals on each of the three lanes used to sequence all accessions. A subsequent  
211 analysis of variance (ANOVA) of the read number did not show a lane effect (Table S1). Since  
212 a high-quality reference genome of an amaranth species was not available, we aligned reads  
213 *de novo* within the dataset to unique tags using Stacks (Catchen *et al*, 2011). The total length  
214 of the aligned reads was 16.6 Mb, which corresponds to approximately 3.3 % of the *A. caudatus*  
215 genome. For SNP calling, reads of each individual were mapped to the aligned tags. SNPs  
216 were called with parameters described in Materials and Methods, which resulted in 63,956  
217 SNPs. Since GBS data are characterized by a high proportion of missing values, we removed  
218 SNPs with more than 60% of missing values. After this filtering step, we obtained 9,485 biallelic  
219 SNPs with an average of 35.3 % missing data for subsequent analyses (Figure S1).

### 220 Inference of population structure

221 To infer the genetic relationship and population structure of *A. caudatus* and its putative ances-  
222 tors, we used three different methods, ADMIXTURE, Discriminant Analysis of Principal Compo-  
223 nents (DAPC) and phylogenetic reconstruction with an uncorrected neighbor-joining network.  
224 The ADMIXTURE analysis with three predefined groups ( $K = 3$ ), which corresponds to the  
225 number of *Amaranthus* species included in the study, did not cluster accessions by their species  
226 origin, but grouped the *A. caudatus* accessions into two distinct clusters and combined the two  
227 wild accessions *A. hybridus* and *A. quitensis* into a single cluster. This analysis indicates a  
228 clear separation between domesticated and the wild Amaranths, but the two wild amaranths  
229 appeared to be a single genetic group because with higher values of  $K$  did not lead to subdivi-  
230 sion of the two wild species into separate clusters that corresponds to the species assignment  
231 (Figure 1). Cross-validation showed that  $K = 5$  was most consistent with the data (Supplemen-  
232 tary Figure S2), which produced three different groups of *A. caudatus* accessions that included  
233 a few wild amaranth accessions, and two wild amaranth clusters that both consist of *A. hy-*  
234 *bridus* and *A. quitensis* accessions. The two wild amaranth clusters differ by the geographic  
235 origin because one cluster contains both *A. hybridus* and *A. quitensis* accessions from Peru  
236 and the other cluster from Ecuador. This indicates a strong strong geographic differentiation  
237 among wild ancestors.

238 The *A. caudatus* accessions clustered into three groups that also showed geographic differenti-

## Incomplete Amaranth domestication

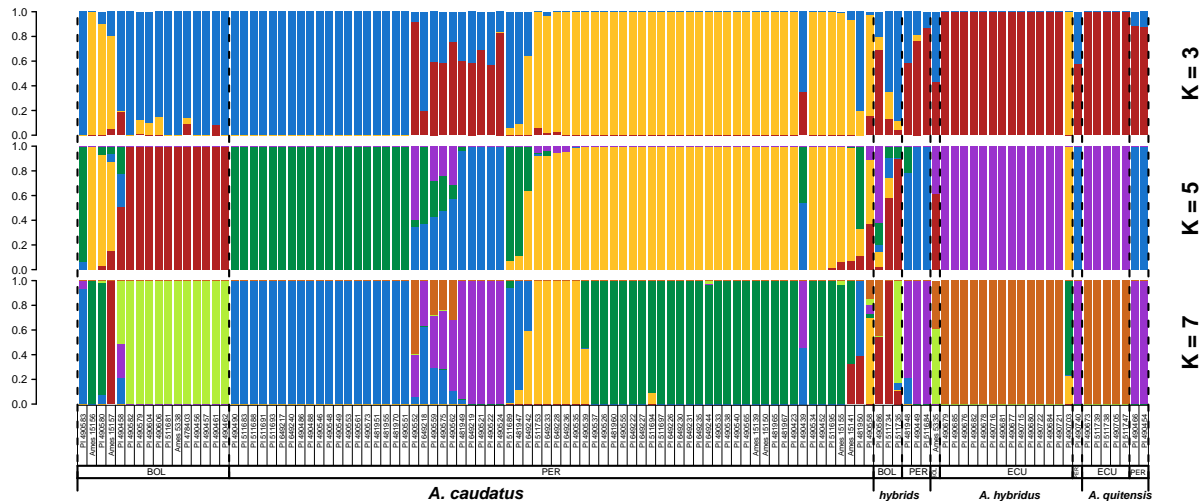


Figure 1: Model based clustering analysis with different numbers of clusters ( $K=3, 5, 7$ ) with ADMIXTURE. The clusters reflect the number of species in the study ( $K=3$ ), the number of single populations (species per country of origin,  $K=7$ ) and the optimal number as determined by cross validation ( $K=5$ ). Individuals are sorted by species and country of origin (BOL=Bolivia, PER = Peru and ECU = Ecuador) as given by their passport data.

239 ation. The first cluster consisted of individuals from Bolivia (Figures 2 and 1;  $K = 5$ , red color).  
240 *A. caudatus* accessions from Peru were split into two clusters of which one predominantly rep-  
241 resents a region from North Peru (Huari Province; Figures 2 and 1;  $K = 5$ , yellow color),  
242 whereas the second cluster contains individuals distributed over a wide geographic range that  
243 extending from North to South Peru ( $K = 5$ , green color). Ten *A. caudatus* accessions from  
244 the Cuzco region clustered with the three accessions of wild amaranths from Peru ( $K = 5$ ,  
245 blue color). These ten accessions showed admixture with the other cluster of wild amaranths  
246 and with a Peruvian cluster. Accessions that were labeled as 'hybrids' in their passport data,  
247 because they express a set of phenotypic traits of different species, clustered with different  
248 groups. 'Hybrids' from Bolivia were highly admixed, whereas 'hybrids' from Peru clustered with  
249 the Peruvian wild amaranths (Figure 1). Taken together, the population structure inference  
250 with ADMIXTURE identified a clear separation between the wild and domesticated amaranth  
251 species and genetic differentiation among domesticated amaranths but also gene flow between  
252 populations.

253 The inference of population structure with a discriminant analysis of principal components  
254 (DAPC) and Neighbor-Joining network produced very similar results as ADMIXTURE. The first

## Incomplete Amaranth domestication

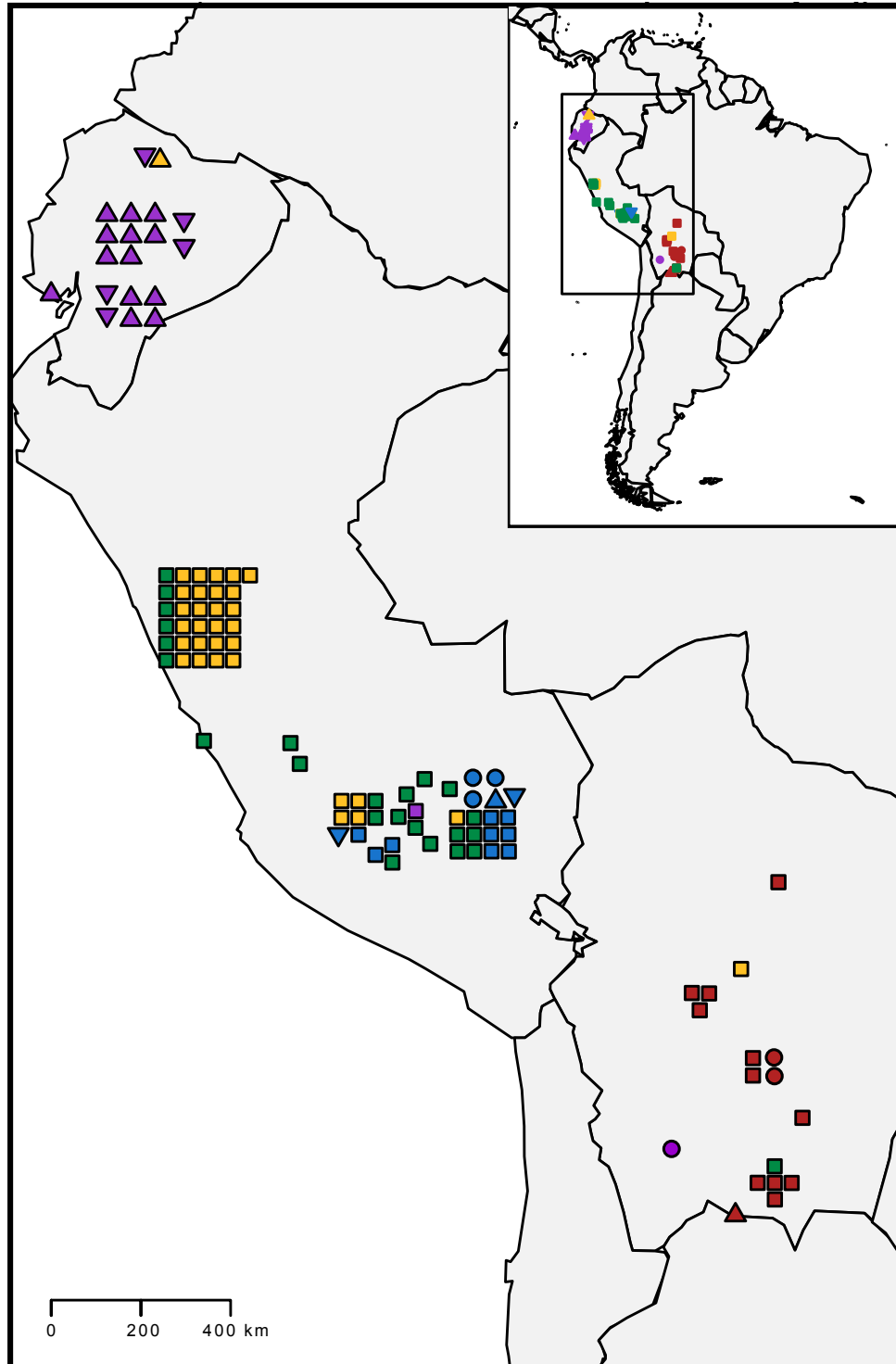


Figure 2: Geographic distribution of accessions for which data was available from passport information. Locations are not exact geographic locations because location data was given as country province. Colors are given by ADMIXTURE with  $K=5$  (Figure 1). Species are indicated by shapes. *A. caudatus* (□), *A. hybridus* (△), *A. quitensis* (▽) and hybrids between species (○)

255 principal component of the DAPC analysis which we used to cluster accessions based on their

256 species explained 96% of the variation and separated the two wild species from the domesti-

## Incomplete Amaranth domestication

257 cated *A. caudatus* (Figure S3A). In a second DAPC analysis that was based on information on  
258 species and country of origin (Figure S3B) the first principal component explained 55% of the  
259 variation and separated most of the wild from the domesticated amaranths. The second princi-  
260 pal component explained 35% of the variation and separated the Peruvian from the Bolivian *A.*  
261 *caudatus* accessions.

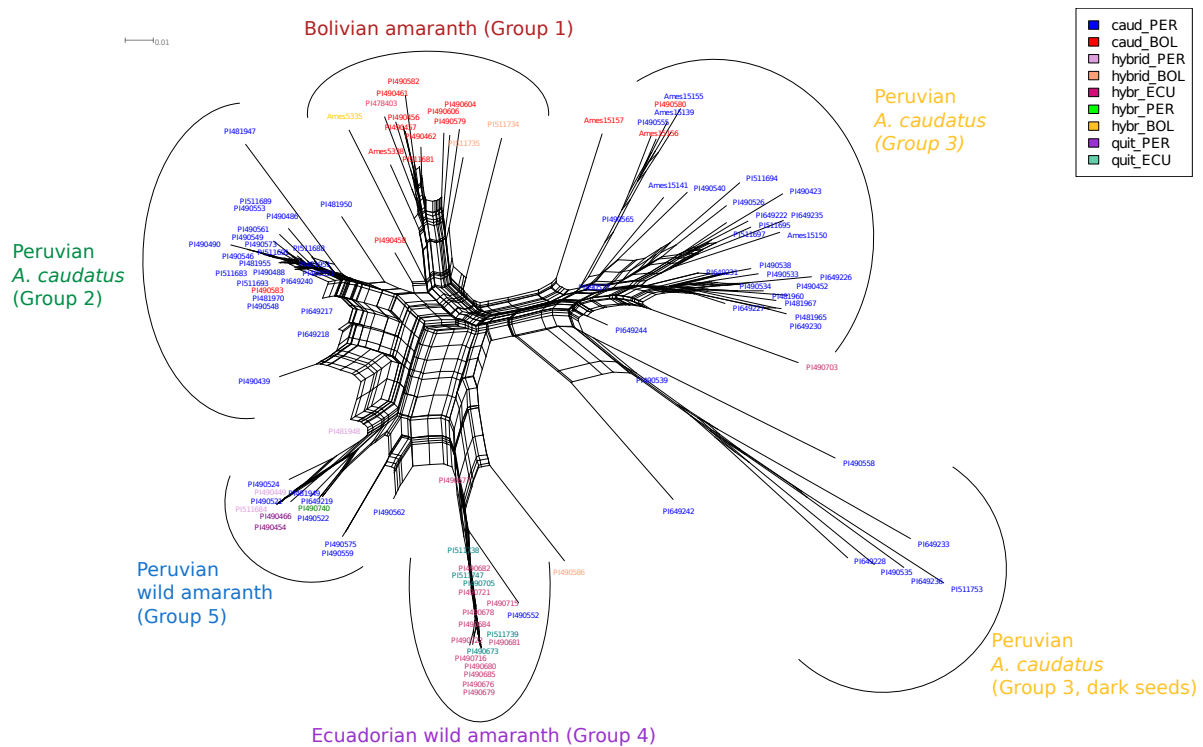


Figure 3: Neighbor-joining network of 113 amaranth accessions from six potential populations. Different colors indicate the species and origin according to gene bank information. *A. caudatus* from Peru (blue) and from Bolivia (red), *A. hybridus* from Ecuador (magenta), from Peru (green) and Bolivia (yellow), *A. quitensis* from Ecuador (turquoise) and Peru (purple) and hybrids between species from Peru (salmon) and Bolivia (light orange). Arches show genetic clusters as inferred with ADMIXTURE (K = 5).

262 The phylogenetic network outlines the relationships between the different clusters (Figure 3).  
263 It shows two distinct groups of mainly Peruvian *A. caudatus* accessions and a group of ac-  
264 cessions with a wide geographic distribution (Figure 2; green color). The latter is more closely  
265 related to the Bolivian *A. caudatus* and the wild relatives. The strong network structure between  
266 these three groups suggests a high proportion of shared polymorphisms or a high level of re-  
267 cent gene flow. In contrast, the clade with *A. caudatus* accessions from Northern Peru are more  
268 separated from the other clades which indicates a larger evolutionary distance, less ongoing

## Incomplete Amaranth domestication

---

269 gene flow with the wild ancestors or stronger selection (Figure 2; yellow color). They are split  
270 into two groups, of which the smaller includes only accessions with dark seeds. In a bifurcating  
271 phylogenetic tree, ten domesticated amaranth accessions clustered within the same clade as  
272 the wild species (Figure S4). The same clustering was also obtained with ADMIXTURE and  
273  $K = 7$  (Figure 1).

Table 1: Weir and Cockerham weighted  $F_{ST}$  estimates between populations based on the taxonomic assignment of their passport data. The group of wild amaranth are *A. hybridus* and *A. quitensis* taken together.

	$F_{ST}$
<i>A. caudatus</i> x <i>A. hybridus</i>	0.319
<i>A. caudatus</i> x <i>A. quitensis</i>	0.274
<i>A. caudatus</i> x wild amaranth	0.322
<i>A. hybridus</i> x <i>A. quitensis</i>	0.041
<i>A. caudatus</i> (PER) x <i>A. caudatus</i> (BOL)	0.132

274 To quantify the level of genetic differentiation between the species and groups within *A. cau-*  
275 *datu*s, we estimated weighted  $F_{ST}$  values using the method of Weir and Cockerham (Weir &  
276 Cockerham, 1984).  $F_{ST}$  values between *A. caudatus* and the wild *A. hybridus* and *A. quiten-*  
277 *sis* species were 0.31 and 0.32, respectively (Table 1), and 0.041 between *A. hybridus* and *A.*  
278 *quitensis* based on the taxonomic assignment. The latter reflects the high genetic similarity  
279 of the accessions from both species observed above. Within *A. caudatus* subpopulations, the  
280  $F_{ST}$  between *A. caudatus* populations from Peru and Bolivia was 0.132, three times higher than  
281 between *A. hybridus* and *A. quitensis*. The above analyses suggested that some individuals  
282 may be misclassified in the passport information, and we therefore calculated  $F_{ST}$  values of  
283 population sets defined by ADMIXTURE. Although such  $F_{ST}$  values are upward biased, they  
284 allow to evaluate the relative level of differentiation between groups defined by their genotypes.  
285 The comparison of  $F_{ST}$  values showed that the three *A. caudatus* groups (groups 1-3) are less  
286 distant to the Peruvian (group 5) than to the Ecuadorian wild amaranths (group 4; Table S2).  
287 A tree constructed with TreeMix, which is based on allele frequencies within groups (Figure 4),  
288 suggests gene flow from the Peruvian *A. caudatus* (group 2) to Peruvian wild amaranth (group  
289 5) and, with a lower confidence level, between wild amaranths from Ecuador (group 4) into  
290 Bolivian *A. caudatus* (group 1), as well as from Bolivian *A. caudatus* to Peruvian *A. caudatus*

## Incomplete Amaranth domestication

291 (Group 2).

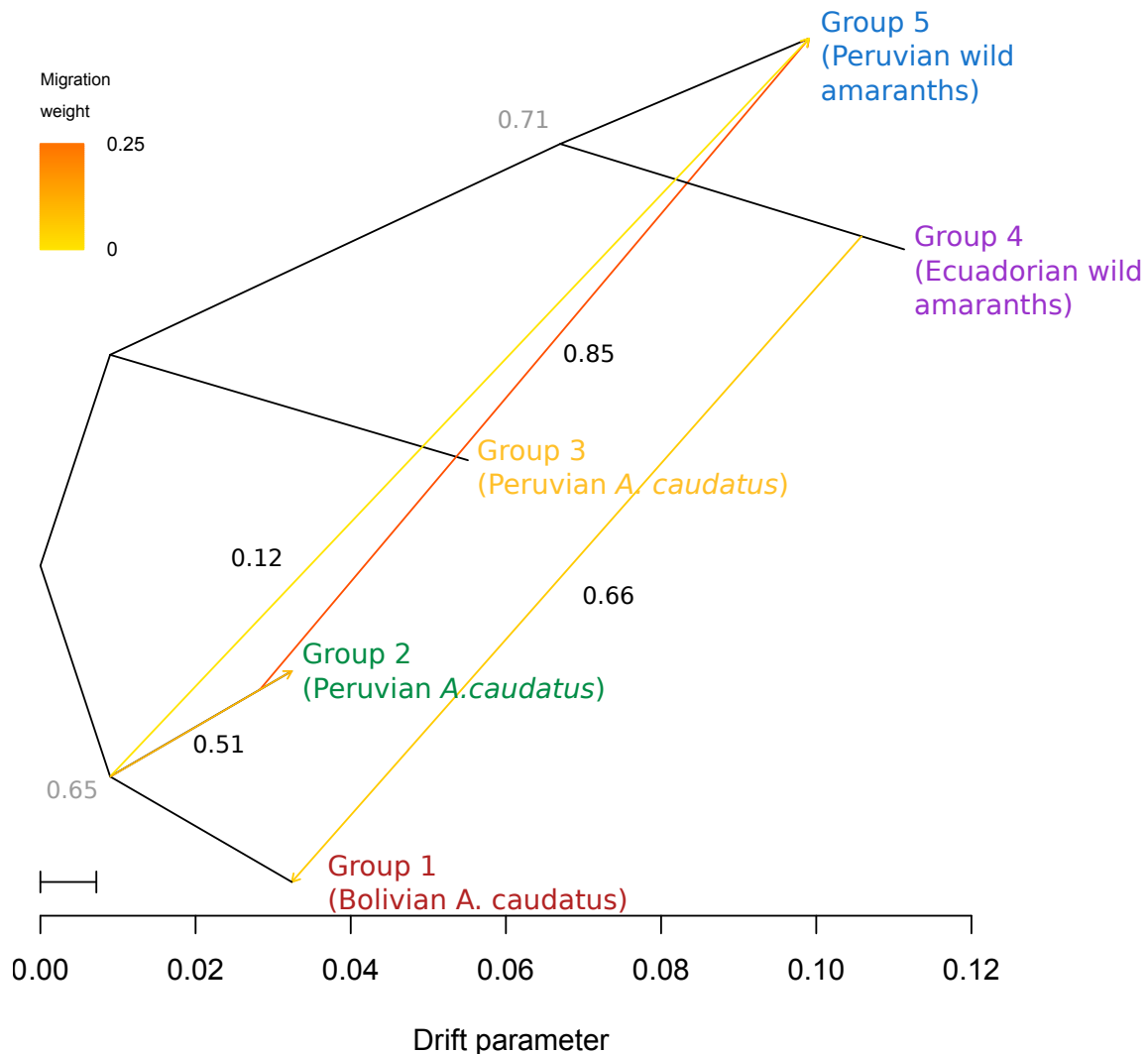


Figure 4: Tree of five genetic clusters of South American amaranths inferred with TreeMix. The genetic clusters which were used to calculate the tree were inferred with ADMIXTURE. Groups 1 to 3 represent *A. caudatus* clusters from Peru and Bolivia, group 4 represents wild amaranth from Ecuador and group 5 wild amaranth from Peru. The migration events are colored according to their weight. Numbers at branching points and on the migration arrow represent bootstrapping results based on 1,000 runs.

## 292 Analysis of genetic diversity

293 We further investigated whether domestication reduced genetic diversity in *A. caudatus* com-  
294 pared to wild amaranths (Table 2). All measures of diversity were higher for the cultivated  
295 than the wild amaranths. For example, nucleotide diversity ( $\pi$ ) was about two times higher in

## Incomplete Amaranth domestication

296 *A. caudatus* than in the two wild species combined. The diversity values of the accessions  
 297 classified as hybrids showed intermediate values between wild and domesticated populations  
 298 supporting their hybrid nature. The inbreeding coefficient,  $F$ , was highest in the domesticated  
 299 amaranth but did not differ from the wild amaranths combined. In contrast accessions classified  
 300 as 'hybrids' and *A. quitensis* showed lower inbreeding coefficients. Within the groups of acces-  
 301 sions defined by ADMIXTURE, genetic diversity differed substantially. The wild amaranths from  
 302 Ecuador had the lowest ( $\pi = 0.00031$ ) while the group from northern Peru showed the highest  
 303 level of nucleotide diversity ( $\pi = 0.00111$ ; Table S3). Figure 5 shows that even though the over-  
 304 all diversity of *A. caudatus* was higher a substantial proportion of sites were more diverse in the  
 305 wild amaranths ( $\pi_{caud} - \pi_{wild} < 0$ ; Figure 5).

Table 2: Genetic diversity parameters for the three putative *Amaranthus* species and the wild amaranth (*A. hybridus* and *A. quitensis*).  $\pi$  is the nucleotide diversity over all sites,  $CI_{\pi}$  is the 95% confidence interval of  $\pi$ ,  $H_{exp}$  the mean expected heterozygosity for the variant sites and  $SD_{He}$  its standard deviation,  $H_{obs}$  the mean observed heterozygosity and  $SD_{Ho}$  its standard deviation.  $F$  is the inbreeding coefficient and  $SD_F$  its standard deviation.

Population	$\pi$	$CI_{\pi}$	$H_{exp}$	$SD_{He}$	$H_{obs}$	$SD_{Ho}$	$F$	$SD_F$	$\theta_w$
<i>A. caudatus</i>	0.00117	$\pm 0.00002$	0.175	0.167	0.049	0.140	0.688	0.462	0.00123
<i>A. hybridus</i>	0.00061	$\pm 0.00001$	0.085	0.135	0.041	0.170	0.679	0.608	0.00073
<i>A. quitensis</i>	0.00059	$\pm 0.00001$	0.076	0.169	0.040	0.170	0.451	0.763	0.00048
Wild amaranth	0.00062	$\pm 0.00002$	0.090	0.140	0.041	0.166	0.681	0.591	0.00070
Hybrids	0.00091	$\pm 0.00001$	0.112	0.179	0.060	0.173	0.436	0.645	0.00107



## Incomplete Amaranth domestication

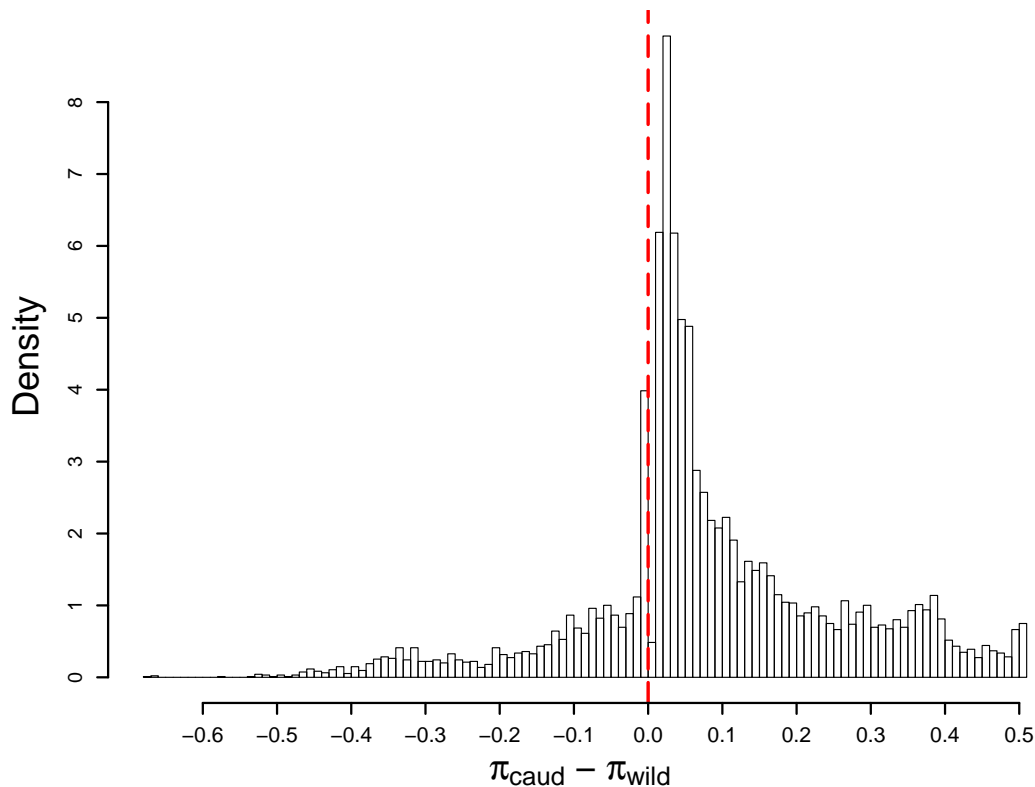


Figure 5: Per site difference in nucleotide diversity ( $\pi$ ) between domesticated amaranth (*A. caudatus*) and wild amaranth (*A. hybridus* and *A. quitensis*)

### 306 **Genome size in wild and cultivated amaranth**

307 Although the genomic history of amaranth species still is largely unknown, genome sizes and  
308 chromosome numbers are highly variable within the genus *Amaranthus* (<http://data.kew.org/cvalues/>). This raises the possibility that the domestication of *A. caudatus* was accom-  
309 panied by polyploidization events as observed in other crops. We therefore tested whether a  
310 change in genome size played a role in the context of domestication by measuring the genome  
311 size of multiple individuals from all three species with flow cytometry. The mean genome size  
312 of *A. caudatus* was 501.93 Mbp, and the two wild ancestors did not differ significantly from  
313 this value (Table 3) indicating that polyploidization did not play a role in the recent evolution of  
314 domesticated amaranth.  
315

## Incomplete Amaranth domestication

Table 3: Genome size of representative group of individuals for each species. There are no significant differences between genome sizes ( $p \leq 0.05$ ). The number of individuals per population is  $N$  and SD is the standard deviation for each parameter.

	$N$	DNA content (pg)	SD	genome size (Mbp)	SD
<i>A. caudatus</i>	22	1.026	0.026	501.93	12.74
<i>A. hybridus</i>	8	1.029	0.025	502.96	12.20
<i>A. quitensis</i>	4	1.021	0.016	499.07	7.91

### 316 **Seed color and seed size as potential domestication traits.**

317 In grain crops, grain size and seed color are important traits for selection and likely played  
318 a central role in domestication of numerous plants (Abbo *et al*, 2014; Hake & Ross-Ibarra,  
319 2015). To investigate whether these two traits are part of the domestication syndrome in grain  
320 amaranth, we compared the predominant seed color of the different groups of accessions and  
321 measured their seed size. The seeds could be classified into three colors, white, pink and  
322 brown. The white and pink types have both a white seed coat, but the latter has red cotyledons  
323 that are visible through the translucent seed coat. A substantial number of seed samples  
324 (19) from the genebank contained seeds of other color up to a proportion of 20%. One *A.*  
325 *caudatus* accession from Peru (PI 649244) consisted of 65% dark seeds and 35% white seeds  
326 in the sample. No accession from the two wild species or hybrid accessions had white seeds,  
327 whereas the majority (74%) of *A. caudatus* accessions had white (70%) or pink (4%) seeds,  
328 and the remaining (26%) brown seeds (Figure 6 A). We also compared the seed color of groups  
329 defined by ADMIXTURE ( $K = 5$ ; Figure 1), which reflect genetic relationship and may correct  
330 for mislabeling of accessions (Figure 6 B). None of these groups had only white seeds, but  
331 clusters that mainly consist of accessions from the wild relatives had no white seeds at all. In  
332 contrast to seed color, the hundred seed weight (HSW) of the different *Amaranthus* species did  
333 not significantly differ between wild and cultivated amaranths. The mean HSW of *A. caudatus*  
334 was 0.056 g and slightly higher than the HWS of *A. hybridus* (0.051 g) and *A. quitensis* (0.050  
335 g; Figure 6 C and Table S4). Among the groups identified by ADMIXTURE ( $K = 5$ ), one  
336 group showed a significantly higher HSW than the other groups, while the other four groups  
337 did not differ in their seed size. The group with the higher HSW consisted mainly of Bolivian  
338 *A. caudatus* accessions and had a 21 % and 35 % larger HSW than the two groups consisting

## Incomplete Amaranth domestication

339 mainly of Peruvian *A. caudatus* accessions, respectively (Figure 6 D). An ANOVA also revealed  
340 that seed color has an effect on seed size because white seeds are larger than dark seeds  
341 (Table 4).

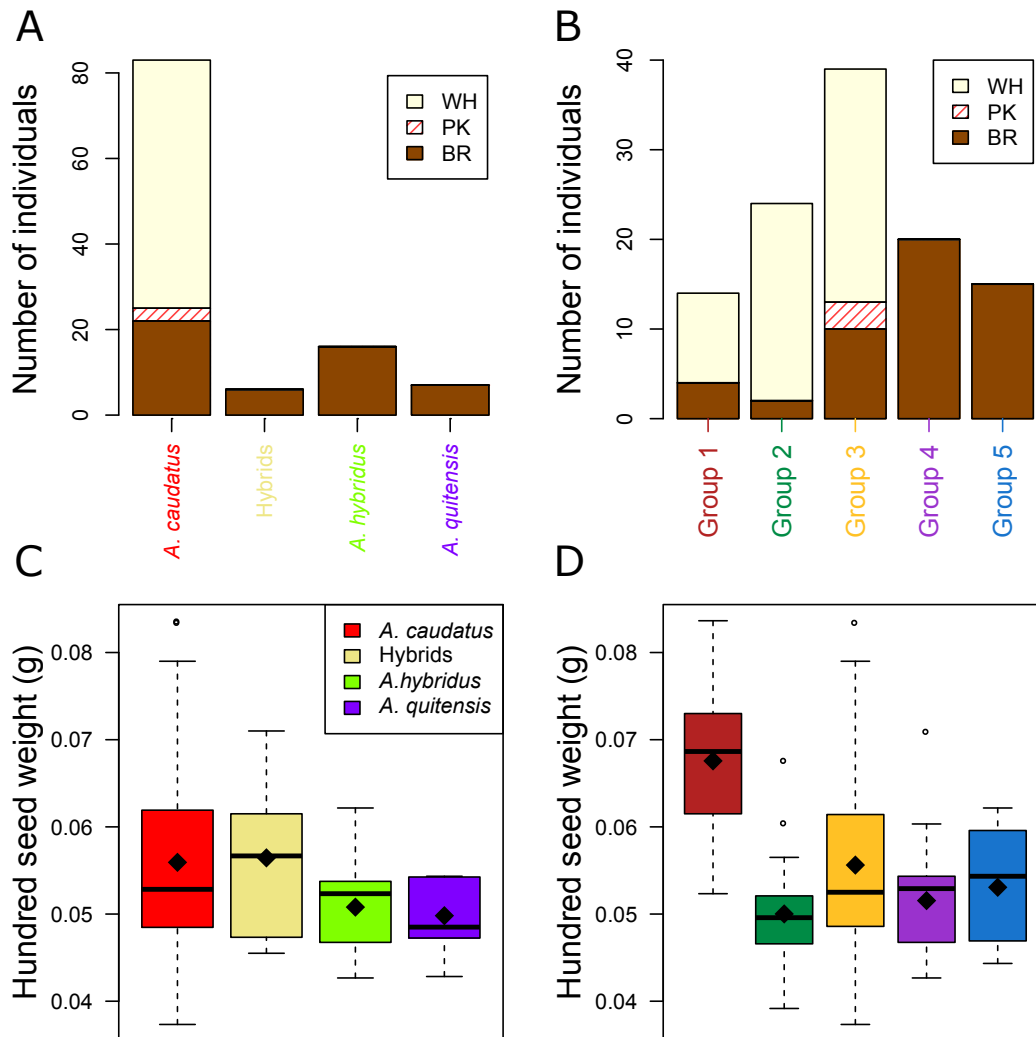


Figure 6: Hundred seed weight (**A,B**) and predominant seed color (**C,D**) by *Amaranthus* species (**A,C**) and groups identified with ADMIXTURE for K=5 (**B,D**) where group 1 (red) resembles *A. caudatus* from Bolivia, group 2 (green) and 3 (yellow) *A. caudatus* from Peru, group 4 (purple) represents wild amaranth from Ecuador and group 5 (blue) wild amaranth from Peru. Seed colors were white (WH), pink (PK) and dark brown (BR). While there were no significant differences in seed size between the species, Group 1 had significantly higher hundred seed weight ( $p \leq 0.05$ ) than the other groups.

## Incomplete Amaranth domestication

---

Table 4: Analysis of variance for the hundred seed weight in dependence of the Seed color and Population as determined by ADMIXTURE

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Seed color	2	0.000657	0.0003285	4.657	0.0116 *
Group	4	0.003151	0.0007877	11.165	1.46e-07 ***
Seed color:Group	2	0.000042	0.0000209	0.297	0.7440
Residuals	103	0.007266	0.0000705		

## Discussion

### Genotyping-by-sequencing of amaranth species

The genotyping of wild and cultivated amaranth accessions revealed a strong genetic differentiation between wild and cultivated amaranths and a high level of genetic differentiation within domesticated *A. caudatus*. We based our sequence assembly and SNP calling on a *de novo* assembly of GBS data with Stacks because currently no high quality reference sequence of an amaranth species is available. Stacks allows SNP calling without a reference genome by constructing a reference catalog from the data and includes all reads in the analysis (Catchen *et al*, 2011). Since *de novo* assembled fragments are not mapped to a reference, they are unsorted and do not allow to investigate differentiation along genomic regions but the data are suitable for the analysis of genetic diversity and population structure (Catchen *et al*, 2013). GBS produces a large number of SNPs (Poland *et al*, 2012; Huang *et al*, 2014), albeit with a substantial proportion of missing values. Missing data lead to biased estimators of population parameters such as  $\pi$  and  $\theta_w$  (Arnold *et al*, 2013) and need to be accounted for if different studies are compared. The comparison of accessions and groups within a study is possible, however, because all individuals were treated with the same experimental protocol. We filtered out sites with high levels of missing values to obtain a robust dataset for subsequent population genomic analysis. Compared to previous studies on amaranth genetic diversity (Maughan *et al*, 2009, 2011; Khaing *et al*, 2013; Jimenez *et al*, 2013; Kietlinski *et al*, 2014), our study combines a larger number of accessions and more genetic markers, which allowed us to assess the genetic diversity and population structure on a genome-wide basis.

## Incomplete Amaranth domestication

---

### 363 ***A. quitensis* and *A. hybridus* are not different species**

364 The two wild relatives *A. quitensis* and *A. hybridus* do not appear to be separate species in  
365 our analyses but form two distinct subgroups of Peruvian and Ecuadorian wild amaranths that  
366 both consist of accessions from the two species. It was suggested before that *A. quitensis* is  
367 the same species as *A. hybridus* (Coons, 1978), but the passport information regarding the  
368 species of genebank accessions was not changed and *A. quitensis* is still considered as a sep-  
369 arate species in these records. The taxonomic differentiation between the two species rests on  
370 a single morphological trait, namely the shape of the tepals, which are very small and prone to  
371 misidentification (Sauer, 1967). The high phenotypic similarity of *A. quitensis* and *A. hybridus*  
372 is supported by our analyses which showed that they are very closely related and mainly sep-  
373 arated by their geographic origin, from Peru and Ecuador. The  $F_{ST}$  value between the two wild  
374 species was lower than between the two *A. caudatus* groups from Peru and Bolivia (Tables 1  
375 and S2). A close relationship is also supported by the highly similar genome sizes of all three  
376 species, although the genus *Amaranthus* harbors species with very different genome sizes  
377 due to variation in chromosome numbers and ploidy levels (Baohua & Xuejie, 2002; Rayburn  
378 *et al*, 2005). In contrast to our results, a recent study found evidence for a genetic differen-  
379 tiation between *A. hybridus* and *A. quitensis* (Kietlinski *et al*, 2014). This discrepancy may  
380 result from the different composition of samples because our sample consists of accessions of  
381 both species from the Andean region whereas Kietlinski *et al.* (2014) used *A. hybridus* and *A.*  
382 *quitensis* accessions with little geographic overlap between the two species. Our  $F_{ST}$  values  
383 also indicate that Peruvian and Ecuadorian wild amaranths show a high level of differentiation  
384 ( $F_{ST} = 0.579$ ; Table S2), which is similar to the differentiation between one of two Peruvian  
385 *A. caudatus* groups and the wild amaranths from Peru ( $F_{ST} = 0.553$ ). In summary, under the  
386 assumption that the passport information of the wild amaranths is correct, we propose that *A.*  
387 *quitensis* and *A. hybridus* are a single species. The high level of intraspecific differentiation in  
388 both wild and cultivated amaranth is relevant for investigating domestication because the ge-  
389 netic distance between groups of cultivated amaranth is related to the geographic distance of  
390 the wild ancestors.

## Incomplete Amaranth domestication

---

### 391 **Diversity of South American amaranth**

392 In numerous crops, domestication was associated with a decrease in genome-wide levels of  
393 diversity due to bottleneck effects and strong artificial selection of domestication traits (Gepts,  
394 2014). In contrast, the overall genetic diversity in our sample of domesticated amaranths was  
395 higher than in the two wild relatives. The distribution of diversity between the GBS fragments  
396 includes genomic regions with reduced diversity in *A. caudatus*, which may reflect selection in  
397 some genomic regions (Figure 5). Without a reference genome it is not possible to position  
398 reads on a map to identify genomic regions that harbor putative targets of selection based on  
399 an inference of the demographic history. Despite the indirect phenotypic evidence for selection,  
400 the higher genetic diversity of domesticated grain amaranth may result from a strong gene flow  
401 between wild and domesticated amaranths. Gene flow between different amaranth species  
402 has been observed before (Trucco *et al*, 2005) and is also consistent with the observation of  
403 six highly admixed accessions classified as 'hybrids' in the passport data and which appear  
404 to be interspecific hybrids (Figure 1 and Table 2). Gene flow between *A. caudatus* and other  
405 *Amaranthus* species in different areas of the distribution range could explain a higher genetic  
406 diversity in the domesticated amaranth, which is also consistent with the strong network struc-  
407 ture (Figure 3) and the TreeMix analysis (Figure 4). Taken together, cultivated *A. caudatus* is  
408 unusual in its higher overall genetic diversity compared to its putative wild ancestors, which is  
409 uncommon in domesticated crops.

### 410 **Amaranth domestication syndrome**

411 Despite its long history of cultivation, diverse uses for food and feed and its high importance dur-  
412 ing the Aztec period, grain amaranth does not display the classical domestication syndrome as  
413 strongly as other crops (Sauer, 1967). On one hand, domesticated amaranth shows morpho-  
414 logical differentiation from wild amaranths like larger and more compact inflorescences (Sauer,  
415 1967) and a level of genetic differentiation (Table 1) which is comparable to the level of differ-  
416 entiation of other domesticated crops and their wild relatives (Sunflower: 0.22 (Mandel *et al*,  
417 2011); common bean: 0.1-0.4 (Papa *et al*, 2005), pigeonpea: 0.57-0.82 (Kassa *et al*, 2012)).  
418 On the other hand, the individual flowers of a plant do not mature synchronously and produce

## Incomplete Amaranth domestication

---

419 very small seeds that are shattered (Brenner, 2000). In contrast to wild amaranths, which  
420 all have dark brown seeds, the predominant seed color of cultivated grain amaranth is white,  
421 which suggests that selection for seed color played a role in the history of *A. caudatus*. How-  
422 ever, dark-seeded accessions are present in all three groups of *A. caudatus* defined by the  
423 genotypic data, which indicates that white seed color is not a fixed trait. Similarly, seed sizes  
424 between wild and domesticated amaranths are not significantly different with the exception of  
425 *A. caudatus* accessions with white seeds from Bolivia (Figure 6), which have larger seeds.  
426 The increased seed size in this group and in white seeds in general indicates past selection  
427 for domestication-related traits, but only in specific geographic regions or in certain types of  
428 amaranth, and not in the whole domesticated crop species.

429 Possible explanations for the incomplete fixation of domestication traits in South American  
430 grain amaranth include weak selection, genetic constraints or ongoing gene flow. First, weak  
431 selection of putative domestication traits may reflect that they were not essential for domes-  
432 tication. Although white seeds are predominant in cultivated amaranth and unambiguously  
433 a domestication-related trait under selection, other seed colors may have been preferred for  
434 different uses with the consequence that genes for white seed color were not fixed. Sec-  
435 ond, domestication traits may experience genetic constraints that limit phenotypic variation.  
436 Genes controlling domestication traits that are part of simple molecular pathways, have mini-  
437 mal pleiotropic effects, and show standing functional genetic variation have a higher chance of  
438 fixation than genes with high pleiotropic or epistatic interactions (Doebley *et al*, 2006; Lenser  
439 & Theißen, 2013). Numerous genes with these characteristics were cloned and characterized  
440 in major crops like rice, barley and maize and shown to contribute to the distinct domestication  
441 syndrome such as a loss of seed shattering, larger seed size and compact plant architecture.  
442 Since the molecular genetic basis of domestication traits in amaranth is unknown, the lack of  
443 a strong domestication syndrome and a lack of fixation of putative domestication traits despite  
444 a long period of cultivation may result from genetic constraints which limited the origin and se-  
445 lection of domestication phenotypes. A third explanation is ongoing gene flow between wild  
446 and domesticated amaranth that may prevent or delay the formation of a distinct domestication  
447 syndrome and contributes to the high genetic diversity (Table 2), similar seed size (Figure 6 C),  
448 and the presence of dark seeds (Figure 6) in cultivated amaranth. Both historical and ongoing

## Incomplete Amaranth domestication

---

449 gene flow are likely because amaranth has an outcrossing rate between 5% and 30% (Jain  
450 *et al*, 1982). In South America, wild and domesticated amaranths are sympatric over wide ar-  
451 eas and the wild *A. hybridus* and *A. quitensis* were tolerated in the fields and home gardens  
452 with *A. caudatus* (Sauer, 1967), where they may have intercrossed. Gene flow between wild  
453 and domesticated plants has also been observed in maize and teosinte in the Mexican high-  
454 lands, but did not have a major influence on the maize domestication syndrome (Hufford *et al*,  
455 2013). Further support for ongoing gene flow in amaranth is given by the presence of hybrids  
456 and admixed accessions in our sample with evidence for genetic admixture and dark seeds  
457 that demonstrate the phenotypic effects of introgression. Since the dark seed color is dominant  
458 over white color (Kulakow *et al*, 1985), dark seeds could have efficiently removed by selection  
459 despite gene flow. Therefore, gene flow likely is not the only explanation for the lack of a distinct  
460 domestication syndrome.

461 Our data are consistent with the model by Kietlinski *et al.* (2014) who proposed a single do-  
462 mestication of *A. caudatus* and *A. hypochondriacus* in Central America followed by migration  
463 of *A. caudatus* to South America. Gene flow between wild amaranths and *A. caudatus* in the  
464 Southern distribution range (Peru and Bolivia) may explain the higher genetic diversity of the  
465 latter despite a strong genetic differentiation. The second model of Kietlinski *et al.* of two in-  
466 dependent domestication events from a single *A. hybridus* lineage that ranged from Central  
467 America to the Andes is supported under the assumption that domestication occurred in South  
468 Peru because of the strong differentiation between Ecuadorian and Peruvian wild amaranths  
469 (Table S2). Since the Peruvian group of wild amaranths inferred with ADMIXTURE comprises  
470 *A. quitensis* and *A. hybridus*, but also *A. caudatus* accessions, the latter may represent acces-  
471 sions from the center of domestication.

## 472 Conclusions

473 The genotypic and phenotypic analysis of wild and domesticated South American grain ama-  
474 ranths suggest that *A. caudatus* is an incompletely domesticated crop species. Key domestica-  
475 tion traits such as the shape of inflorescences, seed shattering and seed size are rather similar  
476 between wild and cultivated amaranths and there is strong evidence of ongoing gene flow from



## Incomplete Amaranth domestication

---

477 its wild ancestor despite selection for domestication traits like white seeds. Although grain ama-  
478 ranth is an ancient crop of the Americas, genomic and phenotypic signatures of domestication  
479 differ from other, highly domesticated crops that originated from single domestication events  
480 like maize (Hake & Ross-Ibarra, 2015). In contrast, the history of cultivated amaranth may in-  
481 clude multiregional, multiple and incomplete domestication events with frequent and ongoing  
482 gene flow from sympatric wild relatives, which is more similar to the history of species like rice,  
483 apple or barley (Londo *et al*, 2006; Cornille *et al*, 2012; Poets *et al*, 2015). The classical model  
484 of a single domestication in a well-defined center of domestication may not sufficiently reflect  
485 the history of numerous ancient crops. Our study further highlights the importance of a compre-  
486 hensive sampling to study the domestication of amaranth. All three domesticated amaranths,  
487 *A. caudatus*, *A. cruentus* and *A. hypochondriacus*, as well as all wild relatives throughout the  
488 whole distribution range should be included in further studies to fully understand and model the  
489 domestication history of Central and South American amaranth.

## 490 **Acknowledgments**

491 We thank David Brenner (USDA-ARS) and Julie Jacquemin for discussions and Elisabeth  
492 Kokai-Kota for support with the GBS library preparation and sequencing. The work was funded  
493 by an endowment of the Stifterverband für die Deutsche Wissenschaft to K. J. S.

## References

- 494
- 495 Abbo S, van Oss RP, Gopher A, Saranga Y, Ofner I, Peleg Z (2014) Plant domestication ver-  
496 sus crop evolution: a conceptual framework for cereals and grain legumes. *Trends in Plant*  
497 *Science*, **19**, 351 – 360.
- 498 Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unre-  
499 lated individuals. *Genome Research*, **19**, 1655–64.
- 500 Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and  
501 introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*,  
502 **22**, 3179–90.
- 503 Arreguez GA, Martínez JG, Ponessa G (2013) *Amaranthus hybridus* L. ssp. *hybridus* in an ar-  
504 chaeological site from the initial mid-Holocene in the Southern Argentinian Puna. *Quaternary*  
505 *International*, **307**, 81–85.
- 506 Baohua S, Xuejie Z (2002) Chromosome numbers of 14 species in *Amaranthus* from China.  
507 *Acta Phytotaxonomica Sinica*, **40**, 428–432.
- 508 Begun DJ, Holloway AK, Stevens K, *et al* (2007) Population genomics: whole-genome analysis  
509 of polymorphism and divergence in *Drosophila simulans*. *PLoS Biology*, **5**, e310.
- 510 Brenner D (2000) Genetic resources and breeding of *Amaranthus*. *Plant Breeding Reviews*,  
511 **19**, 228–285.
- 512 Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of  
513 phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255–65.
- 514 Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool  
515 set for population genomics. *Molecular Ecology*, **22**, 3124–40.
- 516 Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and  
517 genotyping loci *de novo* from short-read sequences. *G3 (Bethesda, Md.)*, **1**, 171–82.
- 518 Coons MP (1978) The status of *Amaranthus hybridus* L. in South America. *Cienc. Nat.*, **18**.
- 519 Cornille A, Gladieux P, Smulders MJM, *et al* (2012) New insight into the history of domesti-  
520 cated apple: Secondary contribution of the european wild apple to the genome of cultivated  
521 varieties. *PLoS Genetics*, **8**, e1002703.
- 522 Costea M, DeMason D (2001) Stem morphology and anatomy in *Amaranthus* L. (*Amaran-*  
523 *thaceae*), taxonomic significance. *Journal of the Torrey Botanical Society*, **128**, 254–281.
- 524 Cruaud A, Gautier M, Galan M, *et al* (2014) Empirical assessment of RAD sequencing for  
525 interspecific phylogeny. *Molecular Biology and Evolution*, **31**, 1272–4.
- 526 Danecek P, Auton A, Abecasis G, *et al* (2011) The variant call format and VCFtools. *Bioinfor-*  
527 *matics (Oxford, England)*, **27**, 2156–8.
- 528 Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell*,  
529 **127**, 1309–21.
- 530 Dolezel J, Bartos J, Voglmayr H, Greilhuber J (2003) Nuclear DNA content and genome size of  
531 trout and human. *Cytometry. Part A : The Journal of the International Society for Analytical*  
532 *Cytology*, **51**, 127–8; author reply 129.

## Incomplete Amaranth domestication

---

- 533 Dolezel J, Sgorbati S, Lucretti S (1992) Comparison of three DNA fliiorocliromes for flow cyto-  
534 metric estimatioe of nuclear DNA conteet in plants. *Physiologia Plantarum*, **85**, 625–631.
- 535 Elshire RJ, Glaubitz JC, Sun Q, *et al* (2011) A robust, simple genotyping-by-sequencing (gbs)  
536 approach for high diversity species. *PLoS ONE*, **6**, e19379.
- 537 Gepts P (2014) The contribution of genetic and genomic approaches to plant domestication  
538 studies. *Current Opinion in Plant Biology*, **18**, 51–9.
- 539 Hake S, Ross-Ibarra J (2015) Genetic, evolutionary and plant breeding insights from the do-  
540 mestication of maize. *eLife*, **4**, 1–8.
- 541 Huang YF, Poland JA, Wight CP, Jackson EW, Tinker NA (2014) Using genotyping-by-  
542 sequencing (GBS) for genomic discovery in cultivated oat. *PLoS ONE*, **9**, e102448.
- 543 Hufford MB, Lubinsky P, Pyhäjärvi T, Devengenzo MT, Ellstrand NC, Ross-Ibarra J (2013) The  
544 genomic signature of crop-wild introgression in maize. *PLoS Genetics*, **9**, e1003477.
- 545 Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies.  
546 *Molecular Biology and Evolution*, **23**, 254–67.
- 547 Jain S, Hauptil H, Vaidya K (1982) Outcrossing rate in grain amaranths. *Journal of Heredity*,  
548 **73**, 71–72.
- 549 Jimenez FR, Maughan PJ, Alvarez A, *et al* (2013) Assessment of genetic diversity in Peru-  
550 vian amaranth (*Amaranthus caudatus* and *A. hybridus*) germplasm using single nucleotide  
551 polymorphism markers. *Crop Science*, **53**, 532.
- 552 Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP  
553 data. *Bioinformatics (Oxford, England)*, **27**, 3070–1.
- 554 Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new  
555 method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.
- 556 Kassa MT, Penmetsa RV, Carrasquilla-Garcia N, *et al* (2012) Genetic patterns of domestication  
557 in pigeonpea (*Cajanus cajan* L. millsp.) and wild cajanus relatives. *PLoS ONE*, **7**.
- 558 Khaing AA, Moe KT, Chung JW, Baek HJ, Park YJ (2013) Genetic diversity and population  
559 structure of the selected core set in *Amaranthus* using SSR markers. *Plant Breeding*, **132**,  
560 165–173.
- 561 Kietlinski KD, Jimenez F, Jellen EN, Maughan PJ, Smith SM, Pratt DB (2014) Relationships  
562 between the weedy (*Amaranthaceae*) and the grain amaranths. *Crop Science*, **54**, 220.
- 563 Kulakow P, Hauptli H, Jain S (1985) Genetics of grain amaranths I. mendelian analysis of six  
564 color characteristics. *Journal of Heredity*, **76**, 27–30.
- 565 Lenser T, Theißen G (2013) Molecular mechanisms involved in convergent crop domestication.  
566 *Trends in Plant Science*, pp. 1–11.
- 567 Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA (2006) Phylogeography of asian wild  
568 rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza*  
569 *sativa*. *Proceedings of the National Academy of Sciences*, **103**, 9578–9583.
- 570 Mallory MA, Hall RV, McNabb AR, Pratt DB, Jellen EN, Maughan PJ (2008) Development and  
571 characterization of microsatellite markers for the grain amaranths. *Crop Science*, **48**, 1098.

## Incomplete Amaranth domestication

---

- 572 Mandel JR, Dechaine JM, Marek LF, Burke JM (2011) Genetic diversity and population struc-  
573 ture in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L.  
574 *Theoretical and Applied Genetics*, **123**, 693–704.
- 575 Maughan P, Smith S, Fairbanks D, Jellen E (2011) Development, characterization, and linkage  
576 mapping of single nucleotide polymorphisms in the grain amaranths (*Amaranthus* sp.). *The*  
577 *Plant Genome Journal*, **4**, 92.
- 578 Maughan PJ, Yourstone SM, Jellen EN, Udall Ja (2009) SNP discovery via genomic reduction,  
579 barcoding, and 454-pyrosequencing in amaranth. *The Plant Genome Journal*, **2**, 260.
- 580 Morris GP, Ramu P, Deshpande SP, *et al* (2013) Population genomic and genome-wide asso-  
581 ciation studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of*  
582 *Sciences*, **110**, 453–458.
- 583 Nabholz B, Sarah G, Sabot F, *et al* (2014) Transcriptome population genomics reveals severe  
584 bottleneck and domestication cost in the african rice (*Oryza glaberrima*). *Molecular Ecology*,  
585 **23**, 2210–27.
- 586 Olsen KM, Wendel JF (2013) A bountiful harvest: genomic insights into crop domestication  
587 phenotypes. *Annual Review of Plant Biology*, **64**, 47–70.
- 588 Papa R, Acosta J, Delgado-Salinas a, Gepts P (2005) A genome-wide analysis of differentiation  
589 between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. *Theoretical and*  
590 *Applied Genetics*, **111**, 1147–1158.
- 591 Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R  
592 language. *Bioinformatics*, **20**, 289–290.
- 593 Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide  
594 allele frequency data. *PLoS Genetics*, **8**.
- 595 Poets AM, Fang Z, Clegg MT, Morrell PL (2015) Barley landraces are characterized by geo-  
596 graphically heterogeneous genomic origins. *Genome Biology*, **16**, 173.
- 597 Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic  
598 maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach.  
599 *PLoS ONE*, **7**, e32253.
- 600 Rayburn AL, McCloskey R, Tatum TC, Bollero GA, Jeschke MR, Tranel PJ (2005) Genome size  
601 analysis of weedy species. *Crop Science*, **45**, 2557.
- 602 Saghai-Marooif MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-  
603 length polymorphisms in barley: mendelian inheritance, chromosomal location, and popu-  
604 lation dynamics. *Proceedings of the National Academy of Sciences of the United States of*  
605 *America*, **81**, 8014–8018.
- 606 Sang T, Li J (2013) Molecular genetic basis of the domestication syndrome in cereals. In  
607 *Cereal Genomics II* (edited by PK Gupta, RK Varshney), pp. 319–340. Springer Netherlands,  
608 Dordrecht.
- 609 Sauer J (1967) The grain amaranths and their relatives: a revised taxonomic and geographic  
610 survey. *Annals of the Missouri Botanical Garden*, **54**, 103–137.

## Incomplete Amaranth domestication

---

- 611 Team RC (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for  
612 Statistical Computing, Vienna, Austria.
- 613 Trucco F, Jeschke MR, Rayburn AL, Tranel PJ (2005) *Amaranthus hybridus* can be pollinated  
614 frequently by *A. tuberculatus* under field conditions. *Heredity*, **94**, 64–70.
- 615 Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure.  
616 *Evolution*, **38**, 1358–1370.

## 617 **Data Accessibility**

- 618 The original genomic data will be available on the European Nucleic Archive (ENA). Scripts and  
619 phenotypic raw data are available under Dryad (<http://datadryad.org/>).

## 620 **Author Contributions**

- 621 M.G.S. and K.J.S. designed research; M.G.S. and K.J.S. performed research; T.M. contributed  
622 analytic tools; M.G.S. analyzed data; and M.G.S. and K.J.S. wrote the paper.

## 623 **Conflict of interest**

- 624 The authors declare no conflict of interest.