1 # The pattern of genetic variability in a core collection of 2,021 cowpea

2 # accessions

3

4 Christopher J. Fiscus[1,#a], Ira A. Herniter[1,#b], Marimagne Tchamba[2], Rajneesh Paliwal[2], María

5 Muñoz-Amatriaín[3], Philip A. Roberts[4], Michael Abberton[2], Oluwafemi Alaba[1], Timothy J. Close[1,5],

6 Olaniyi Oyatomi[2*] and Daniel Koenig[1,5*]

7

8

9 [1]Department of Botany and Plant Sciences, University of California, Riverside, CA, USA

10 [2]International Institute of Tropical Agriculture, Ibadan, Nigeria

11 [3]Departamento de Biología Molecular, Universidad de León, 24071 León, Spain

12 [4]Department of Nematology, University of California, Riverside, CA, USA

13 [5]Institute for Integrative Genome Biology, University of California, Riverside, CA, USA

14

15 Present addresses:

16 [#a]Department of Ecology and Evolutionary Biology, University of California, Irvine, Irvine, CA

17 USA

18 [#b]Department of Plant Biology, Rutgers University, New Brunswick, NJ, USA 08901

19

20 *Corresponding authors: Daniel Koenig and Olaniyi Oyatomi

21 Emails: dkoenig@ucr.edu & o.oyatomi@cgiar.org

Abstract

Cowpea is a highly drought-adapted leguminous crop with great promise for improving agricultural sustainability and food security. Here, we report analyses derived from array-based genotyping of 2,021 accessions constituting a core subset of the world's largest cowpea collection, held at the International Institute of Tropical Agriculture (IITA) in Ibadan, Nigeria. We used this dataset to examine genetic variation and population structure in worldwide cowpea. We confirm that the primary pattern of population structure is two geographically defined subpopulations origining in West and East Africa, respectively, and that population structure is associated with shifts in phenotypic distribution. Furthermore, we establish the cowpea core collection as a resource for genome-wide association studies by mapping the genetic basis of several phenotypes, with a focus on seed coat pigmentation patterning and color. We anticipate that the genotyped IITA cowpea core collection will serve as a powerful tool for mapping complex traits, facilitating the acceleration of breeding programs to enhance the resilience of this crop in the face of rapid global climate change.

Introduction

Enhancing the sustainability of agriculture is essential to feeding the world population, especially in the context of climate change. Among crops, legumes are unique in their ability to improve soil fertility through symbiotic association with rhizobia, which reside within root nodules to facilitate nitrogen fixation (Nap and Bisseling 1990). This quality has made legumes important components of crop rotation and intercropping, both of which help to increase productivity (Reckling *et al.* 2016). In addition, grain legumes are notable for the high protein content of their grains (termed "pulses"), and their amino acid compositions complement those of cereals, making them exceptionally nutritious for both human and animal diets. Productivity gains in grain legumes over the past 50 years have been attributed primarily to expanded planting area rather than to advances in crop improvement as seen in other types of crops, such as cereals

2

48 (Foyer *et al.* 2016). As such, the development of genetic resources linking genotype to

49 phenotype is crucial to improve grain legumes and to advance crop research and breeding.

50   Cowpea (*Vigna unguiculata* (L.) Walp.) is a leguminous crop of global importance as a

51 source of both food and fodder. The crop is produced around the world with the majority of

52 contemporary production in West Africa (7.6 of 8.9 million tonnes in 2021) (Food and Agriculture

53 Organization of the United Nations 2022), where it is grown mainly for its protein-rich dry grain.

54 However, other parts of the plant are edible including the leaves, stems, and immature pods

55 (Weng *et al.* 2019; Boukar *et al.* 2019), which, in addition to culinary uses, are used in traditional

56 medicine (Abebe and Alemayehu 2022). Cowpea haulms, stems, and leaves also constitute a

57 nutritious fodder for livestock (Tarawali, S., Singh, B., Peters, M. & Blade, S. 1997). In addition

58 to its ability to grow in nutrient-poor soil, cowpea is also remarkable for having exceptional

59 resilience to drought and high temperatures (Hall 2004), rendering it an ideal candidate for

60 bolstering food system sustainability and for mitigating the repercussions of global climate

61 change.

62   Linguistic and archaeological evidence suggest that cowpea was domesticated at least

63 5,000 years ago in Sub-Saharan Africa (Herniter *et al.* 2020), presumably from *V. unguiculata*

64 ssp. *dekindtiana* var. *spontanea* (Pasquet and Padulosi 2012). The domestication region,

65 however, remains a subject of debate with proposed domestication centers in West Africa

66 (Vaillancourt and Weeden 1992; Padulosi and Ng 1997; Coulibaly *et al.* 2002; Ba *et al.* 2004),

67 East Africa (Xiong *et al.* 2018), or in both regions (Huynh *et al.* 2013; Xiong *et al.* 2016; Herniter

68 *et al.* 2020). Previous efforts to study genetic variation in contemporary cowpea germplasm

69 collections have indicated a strong pattern of genetic differentiation between germplasm from

70 West and East Africa (Huynh *et al.* 2013; Xiong *et al.* 2016; Muñoz-Amatriaín *et al.* 2017, 2021;

71 Fatokun *et al.* 2018; Herniter *et al.* 2020), with West Africa as the region of maximum diversity of

72 domesticated cowpea (Fatokun *et al.* 2018; Muñoz-Amatriaín *et al.* 2021).

73   Although genetic resources for cowpea are underdeveloped compared to cereal crops

74 such as maize, rice, and wheat, cowpea benefits from several available sequenced genomes

75 (Lonardi *et al.* 2019; Liang *et al.* 2023) as well as a high quality genotyping array

76 (Muñoz-Amatriaín *et al.* 2017). Several quantitative trait locus (QTL) mapping populations have

77 been developed to link genotype to phenotype in this species and have been useful in

78 identifying QTL for diverse traits such as flowering time (Andargie *et al.* 2013; Huynh *et al.* 2018;

79 Lo *et al.* 2018; Olatoye *et al.* 2019), disease resistance (Huynh *et al.* 2016; Santos *et al.* 2018),

80 seed coat color and patterning (Herniter *et al.* 2018, 2019), perenniality (Lo *et al.* 2020), and

81 several domestication-related traits such as seed shattering and yield (Lo *et al.* 2018).

82 Additionally, several genome-wide association studies (GWAS) have been conducted to

83 discover SNPs associated with flowering time (Paudel *et al.* 2021; Muñoz-Amatriaín *et al.* 2021),

84 disease resistance (Kpoviessi *et al.* 2022; Dong *et al.* 2022), seed size (Lo *et al.* 2019), seed

85 protein content (Chen *et al.* 2023), drought response (Ravelombola *et al.* 2021; Wu *et al.* 2021),

86 pod shattering (Lo *et al.* 2021), and salt tolerance (Ravelombola *et al.* 2022). Nevertheless, the

87 majority of QTL mapping studies in cowpea have been limited in their ability to fine-map

88 associated loci due to constrained sample sizes or low molecular marker densities.

89   Here, we report the genetic characterization of a core collection of cowpea held at the

90 International Institute of Tropical Agriculture (IITA), a CGIAR research center in Ibadan, Nigeria.

91 The IITA cowpea collection represents the most comprehensive collection of worldwide cowpea

92 germplasm, currently consisting of 16,460 accessions collected from 91 countries (March 2022)

93 (IITA 2023). The bulk of the collection (~12,000 lines) consists of locally adapted landraces

94 (hereafter traditional cultivars). The core subset of the collection was selected to represent the

95 diversity of geographic origin and agrobotanical traits present among accessions in the full

96 collection (Mahalakshmi *et al.* 2007) and has been subsequently refined to include 2,082

97 accessions. To study genetic variation in the Core Collection, we made use of previously

98 generated SNP data (Close *et al.* 2023) consisting of 2,021 accessions genotyped at nearly

99 50,000 polymorphic sites using the Illumina Cowpea iSelect Consortium Array

100 (Muñoz-Amatriaín *et al.* 2017). We then leveraged this dataset to characterize the pattern of

101 population structure in global cowpea and demonstrated the utility of this resource for

102 genome-wide association studies by dissecting the genetic basis of variation in seed

103 pigmentation.

104

105 **Results**

106 *Genetic diversity of the IITA Cowpea Core Collection*

107 Genomic DNA was extracted from 2,021 accessions from the IITA Cowpea Core Collection and

108 scored at nearly 50,000 polymorphic sites using the Illumina Cowpea iSelect Consortium Array

109 (Close *et al.* 2023). Following removal of sites and samples with missing data and excess

110 heterozygosity, the dataset comprised genotypes from 1,991 accessions collected from 86

111 countries across 6 continents (Figure 1A). All major cowpea producing regions were

112 represented in the dataset with 1,547 accessions from Africa, 207 from Asia, 110 from North

113 America, 68 from Latin America, 43 from Europe and 14 from Oceania. The accessions mostly

114 consisted of 1,590 traditional cultivars, with an additional 206 breeding lines mostly from Nigeria

115 and the United States of America. Additionally, there were 57 uncultivated accessions,

116 presumed to be wild or weedy.

117   Pairwise genetic distances calculated between accessions across filtered sites revealed

118 many nearly identical accessions in the Core Collection (Figure S1). In total, we identified 130

119 clusters (comprising 400 accessions) of nearly identical accessions by applying hierarchical

120 clustering to the genetic distance matrix (Table S1, static tree cut threshold of 500 corresponded

121 to differences at ~0.5% or fewer sites). The vast majority of clusters consisted of either two or

122 three accessions (79 and 25 clusters, respectively), while 26 clusters consisted of 4 or more

123 accessions (max cluster size 15). Most clusters (72 / 130) contained accessions collected from

124 different countries with 40 clusters consisting of accessions from countries on different

125 continents, 19 clusters consisting of accessions collected from adjacent countries within Africa,

126 12 clusters consisting of accessions from non-adjacent countries within Africa, and 1 cluster

127 with accessions from adjacent countries in South America. The geographic distribution of

128 accessions within clusters suggests that accessions were dispersed between continents by

129 human activity and then resampled during the construction of the Core Collection. Of the 130

130 clusters, 86 consisted of accessions of the same germplasm status (e.g. traditional cultivars

131 only), 20 clusters contained both traditional cultivars and breeding lines, 12 clusters had

132 traditional cultivars and accessions of unknown status, 3 clusters contained breeding material

133 and accessions of unknown status, and 2 clusters contained wild or weedy accessions and

134 traditional cultivars. We randomly selected one accession from each of the 130 clusters for

135 subsequent analysis.

136   The final dataset consisted of 1,722 accessions genotyped for 48,004 biallelic SNP

137 markers. The included accessions reflect the geographic diversity of the IITA Cowpea

138 Collection, with lines originating from 85 countries and include ~83% (1,722 / 2,081) of

139 accessions from the IITA Cowpea Core Collection (Mahalakshmi *et al.* 2007) and 86% (324 /

140 376) of accessions included in the IITA Cowpea Mini-Core Collection (Fatokun *et al.* 2018). The

141 analyzed samples encompassed 1,394 traditional cultivars, 171 breeding and research lines, 32

142 weedy accessions, 17 wild accessions, and 110 lines with unknown germplasm status. While

143 the traditional cultivars constitute a global collection, the breeding and research material was

144 predominately collected from Africa and North America (98 and 51 accessions, respectively).

145 The presumed wild accessions were exclusively collected from Africa, while the presumed

146 weedy accessions originated solely from West Africa.

147   Crop dispersal and improvement can lead to genome-wide depletion of genetic diversity.

148 We sought to determine how these processes impacted cowpea by estimating average

149 expected heterozygosity (H_e) within the major geographic regions represented in our dataset. H_e

150 ranged from 0.14 in Northern Europe to 0.29 in South Asia among regions that contributed 3 or

151 more accessions (Figure 1B, Table S2). Within Africa, H_e was similar for all regions (H_e =

152 0.25-0.27) except for North Africa, which had a substantial reduction in diversity (H_e = 0.15).

153 Outside of Africa, H_e varied considerably, with higher values exceeding those observed in Africa

154 in South Asia, West Asia, North America, and Latin America. In contrast, all regions of Europe

155 and Oceania had reduced H_e compared to Sub-Saharan Africa.

156   Although H_e provides a summary of genetic variation within a population, the measure is

157 influenced by the number of segregating alleles and their frequencies. The number of

158 segregating sites per geographic region ranged from 13,660 in Northern Europe to 47,361 in

159 West Africa (Figure 1C, Table S2). The number of segregating sites per region was greatest in

160 Sub-Saharan Africa, with a loss of over 10% of segregating sites observed in all other regions

161 except for South Asia. The proportion of segregating sites was especially low in Oceania and

162 East Asia, though these regions were not well sampled in the IITA core.

163   Taken together, our analyses are consistent with the hypothesis that West Africa is the

164 center of domesticated cowpea genetic diversity with the largest number of segregating sites

165 observed in this region. We observed a reduction of genetic diversity in all regions out of Africa

166 except South Asia, which is consistent with founder effects upon dispersal. The elevation of H_e

167 in South Asia is largely due to the majority of alleles segregating at intermediate frequencies

168 (median minor allele frequency among segregating alleles = 0.25). The reduction of both H_e and

169 the proportion of segregating sites in North Africa suggests strong founder effect with low

170 genetic diversity among the founding lines. We also found evidence that cowpea improvement

171 has reduced genetic diversity, with breeding lines having an enrichment for rare alleles when

172 compared to traditional cultivars (Figure 1D). Of the 47,660 SNPs segregating amongst

173 traditional cultivars, only 46,298 segregated in the breeding material.

174

7

Patterns of population structure

After removal of nearly identical samples, we observed a multimodal distribution of pairwise genetic distances between accessions suggesting that strong population structure might exist in the core collection. Hierarchical clustering of accessions by genetic distance linked population structure to geographic origin and germplasm status (Figure 1E). The primary genetic pattern in the dataset consisted of two large clusters: one cluster composed of accessions from Eastern and Southern Africa and a second cluster encompassing accessions from Northern, Central, and Western Africa. Germplasm collected from outside Africa is interspersed within the two main clusters, consistent with the aforementioned pattern of migration out of Africa likely facilitated by human activity. We also observed that many wild and weedy accessions clustered together within the larger cluster. It remains uncertain whether this observation is a result of ascertainment bias of our genotyping approach or if accessions designated as wild or weedy are actually feral accessions that have escaped cultivation.

To examine population structure in more detail, we next estimated the ancestries of each accession using ADMIXTURE (Alexander *et al.* 2009; Alexander and Lange 2011) (Figure 2A). Evanno's ΔK method (Evanno *et al.* 2005) suggested an optimum of two subpopulation clusters (Figure S2). We assigned accessions to the respective cluster that contributed 70% or more ancestry, resulting in 614 accessions in cluster 1, 680 accessions in cluster 2, and 428 accessions unassigned to a cluster (Table S1). Accessions were non-randomly distributed between clusters for 13 of the 17 geographic regions (Chi-squared tests, $p < 0.05$, Table S3). Cluster 1 contains the majority of accessions from East Africa (90%), Latin America and the Caribbean (75%), North America (71%), Southern Africa (72%), and Australia and New Zealand (67%). In contrast, cluster 2 consists of the majority of accessions from North Africa (92%), West Africa (59%), and Central Africa (52%). The majority of lines from East Asia (86%), Melanesia (83%), and Southeast Asia (82%) were unassigned to a cluster. Germplasm from Eastern Europe, Micronesia, Southern Europe, and West Asia did not have significant

enrichment for belonging to a particular cluster. Given the strong partitioning of accessions by geographic subregion within Africa, we hereafter refer to cluster 1 as the ESA (Eastern and Southern Africa) cluster and cluster 2 as the WNCA (West, North, and Central Africa) cluster. The two clusters have significant genetic differentiation (F_{st} = 0.292) and the ESA cluster has a slightly higher level of genetic diversity compared to the WNCA cluster (H_e = 0.25 for ESA, 0.22 for WNCA, 0.28 for unassigned).

To explore the possibility of population substructure in West Africa, we considered additional subpopulations. Considering K = 3 subpopulations reassigned 124 accessions from the WNCA cluster and 82 previously unassigned accessions to the new cluster 3. Additionally, 78 and 146 accessions from the ESA and WNCA clusters, respectively, were considered to be unassigned at K = 3. Cluster 3 consisted of nearly all accessions from North Africa in proximity to the Nile River Delta (57 / 62) and the majority from Southern Europe (20 / 25), East Asia (5 / 7), and Eastern Europe (6 / 10) (Figure 2A). We also observed that the ancestry previously attributed to the WNCA cluster in accessions found outside of Africa was reassigned to cluster 3. Cluster 3 accessions were nearly exclusively collected in arid climates (Köppen climate classification: BWh/BWk), while other clusters are found to be mostly confined to wet tropical climates (Figure S3), hinting at the possibility that cluster 3 ancestry includes alleles locally adapted to warmer, drier environments.

Modeling population structure with an additional subpopulation (K = 4) grouped 148 accessions previously unassigned to a cluster when considering fewer subpopulations to cluster 4 (Figure 2A,D). Cluster 4 consisted of 133 accessions from West Africa, 7 accessions from either Central or East Africa, and 8 accessions collected in other world regions. The West African cluster 4 accessions were not associated with a particular landform such as a river or mountain range and, like the majority of accessions in the WNCA cluster, originated from tropical savanna climates (Figure S3). This group of accessions likely represents early-flowering

226 traditional varieties that were cultivated alongside later maturing lines as a precaution against

227 drought (Mortimore *et al.* 1997).

228   A principal component analysis (PCA) of allele frequencies was largely consistent with

229 the population structure modeling. The first 10 PCs explain 27% of the variance in the dataset,

230 with nearly half of the variance (12.8%) explained by the first 2 PCs alone (Figure S4). Overall,

231 the distribution of individuals along the first two PCs was continuous, suggesting frequent

232 admixture between population groups (Figure 2B-D). PC 1 separates the ESA cluster from the

233 WNCA cluster (Figure 2B-D), while PC 2 separates cluster 3 from the other clusters (Figure 2C).

234 Cluster 4 is not resolved along the first two PCs and is instead resolved on PC 3 (Figure 2D,

235 Figure S5).

236   ESA ancestry was also overrepresented in the modern breeding material present in the

237 IITA core though a large proportion of accessions remained unassigned (Table S4), perhaps the

238 result of controlled crosses between accessions from the two clusters. Consistently, increasing

239 K did not reduce the number of unassigned accessions, pointing to true admixture rather than

240 cryptic population structure in the sample. In addition, we found limited evidence for contribution

241 from clusters 3 and 4 in the breeding germplasm.

242   Overall, our population structure analyses confirm that global cowpea is characterized by

243 genetic differentiation between germplasm from Western versus Eastern and Southern Africa

244 (Huynh *et al.* 2013; Xiong *et al.* 2016; Muñoz-Amatriaín *et al.* 2017, 2021; Fatokun *et al.* 2018;

245 Herniter *et al.* 2020). Further, our results suggest that germplasm derived from Northern,

246 Eastern, and Southern but not Western Africa represent the majority of ancestry in lines

247 distributed to other parts of the world. Breeding material appears to be enriched for ancestry

248 from the ESA cluster, potentially neglecting climate-adapted alleles from other clusters.

249

Phenotypic distributions of subpopulations

Our analyses of population structure in global cowpea indicate that cowpea germplasm is broadly genetically differentiated into two germplasm groups defined by geography (i.e. the ESA and WNCA clusters). To understand the degree to which population structure is linked to phenotype, we leveraged the IITA Cowpea Characterization and Cowpea Evaluation datasets (IITA 2023) to compare the distributions of 34 quantitative phenotypes and the frequencies of 37 qualitative phenotypes between the two subpopulation clusters (Materials and Methods, Table S5, Table S6). As a caveat, the phenotyping data we analyze here was collected through IITA's cowpea breeding program through numerous field trials conducted across both time (1967-present) and location. Therefore, the data represent reference phenotype values used for breeding efforts and do not correspond to values measured as part of a single experiment.

We identified 17 quantitative phenotypes with significantly different distributions between the WNCA and ESA subpopulation clusters (Wilcoxon test Bonferroni-corrected $p < 0.05$, Figure 3A, Table S7). Overall, the majority of differences were found in life history and morphological traits and not in disease susceptibility traits. Accessions in the ESA cluster generally matured earlier (Figure 3B), were more branched (Figure 3C), had larger lateral organs (Figure 3D-E), and produced more seeds that were generally smaller in size compared to the WNCA cluster (Figure 3F). Additionally, accessions in the ESA cluster were more likely to have smooth compared to rough testa texture, raceme in all layers of the canopy rather than in the upper leaf layers, intermediate plant growth habit with lower branches touching the ground, more twining, and higher susceptibility to anthracnose compared to the WNCA cluster (Figure S6).

We also observed differences in the incidence of qualitative traits between the ESA and WNCA subpopulation clusters. In total, there were 17 traits with differential frequency between the two clusters (Fisher's Exact Test Bonferroni-corrected $p < 0.05$, Figure 2G, Table S8). Many of the phenotypes with significant differences were seed coat pigmentation color and patterning traits. For instance, black and white eye colors and the hilum ring, small eye, Watson, and

11

276 speckled pigmentation patterns were more common in the WNCA cluster compared to the ESA

277 cluster (see below for a detailed description of these patterns). In contrast, the brown, red, and

278 tan eye colors were more common in the ESA cluster. Additional phenotypes with differential

279 frequencies between the two clusters included several leaflet shapes, with globose and

280 sub-globose leaflet shapes more common in ESA and the sub-hastate shape more common in

281 the WNCA cluster. We also observed that flower pigmentation is more common in the ESA

282 cluster and incidence of target leaf spot was more common in the WNCA cluster.

283   Taken together, the data suggests that ESA and WNCA cowpea are differentiated across

284 a wide suite of traits spanning plant architecture, pigmentation, and life history.

285

286 *Decay of linkage disequilibrium in the cowpea genome*

287 To establish the IITA core collection as a valuable resource for identifying the genetic basis of

288 agronomically relevant traits using GWAS in cowpea, we examined the patterns of

289 genome-wide and chromosome-specific linkage disequilibrium (LD) between SNP markers.

290 Overall, we found that LD decays to R^2 = 0.20 by 121.5 Kbp on average, with full decay to the

291 genome-wide background by 2.76 Mbp (Figure 4A). We also observed differences in the rate of

292 LD decay between the ESA and WNCA subpopulation clusters, with more rapid decay in the

293 WNCA cluster (R^2 = 0.20 by 136 Kbp and 89 Kbp, genome-wide background by 3.57 Mbp and

294 2.97 Mbp, respectively).

295   The rate of LD decay also varied between chromosomes. LD decay to R^2 = 0.20 ranged

296 from 68 Kbp on Vu08 to 178 Kbp on Vu01 and decay to chromosome-specific background levels

297 of LD ranged from 721 Kbp on Vu04 to 2.48 Mbp on Vu01 (Figure 4B, Figure S7). As expected

298 given the genome-wide data, the rate of LD decay was more rapid across nearly all

299 chromosomes in the WNCA cluster compared to the ESA cluster.

300   The resolution of genetic mapping in GWAS is determined by the recombination rate,

301 which determines LD decay, and the density of genetic markers near an associated variant. In

12

302 regions with dense markers and high recombination (i.e. rapid LD decay), there is a higher

303 likelihood of pinpointing an associated locus within a narrow genomic window. Using our

304 measures of LD decay, we estimate that an associated locus will contain ~25 genes and ~60

305 SNPs on the array on average. This estimate is based on the average distance of 5,611 bp

306 between adjacent genes, an average gene length of 3,843 bp, and average distance of 3,997

307 bp between adjacent SNPs as measured in the reference genome (Muñoz-Amatriaín *et al.*

308 2017; Lonardi *et al.* 2019). However, we acknowledge that these estimates should be

309 considered to be coarse due to the non-uniform distribution of genes and SNPs and variation in

310 the local recombination rate within the genome.

311

312 *The genetic architecture of seed pigmentation patterning*

313 To demonstrate the utility of the genotyped IITA core collection to map the genetic basis of

314 agronomically important traits, we focused on studying the genetic basis of seed pigmentation.

315 Variation in the color and pattern of cowpea seed pigmentation is a particularly conspicuous

316 example of the rich phenotypic diversity of the IITA collection. Cowpea seeds range from having

317 no pigmentation (called "no color") to complete pigmentation throughout the seed coat

318 (so-called "full coat"). Between these extremes, various pigmentation patterns exist, traditionally

319 classified qualitatively. The "small eye" or "Eye 1" pattern displays pigmentation in a narrow

320 margin around the seed hilum with an indistinct margin, while the "large eye" or "Eye 2" pattern

321 displays pigmentation encompassing the hilum with a wide, distinct margin and covering a

322 significant portion of the ventral surface (Herniter *et al.* 2019). The "Watson" pattern, named

323 after a corresponding variety, features pigmentation surrounding the hilum and extending to the

324 micropylar end of the seed, with an unclear boundary (Spillman 1911). Similarly, the "Holstein"

325 pattern, also named after a variety possessing the pattern, exhibits solid pigmentation over most

326 of the seed coat, with patches of color on the remaining portion of the seed. In addition to

327 pigmentation patterning, seed coat color exhibits extraordinary variation across a spectrum of

13

328 colors including white/cream, tan/buff, brown, red, maroon, and purple/black (Harland 1919,

329 1920; Spillman and Sando 1930; Saunders 1960; Drabo *et al.* 1988).

330   Seed coat pigmentation patterning in cowpea has been described as being controlled by

331 a three locus system consisting of the color factor (*C* locus), Watson factor (*W* locus), and

332 Holstein factor (*H* locus) (Spillman and Sando 1930; Saunders 1960; Drabo *et al.* 1988). The *C*

333 locus is epistatic to the other two loci and determines if pigmentation is present throughout the

334 seed coat or just restricted to the eye. The *W* and *H* loci meanwhile act as modifiers and

335 determine whether the pattern of pigmentation in the seed coat develops into the Watson,

336 Holstein, or full coat patterns. Previous work based on QTL mapping in biparental populations

337 has proposed *Vigun07g110700* as a candidate gene for the *C* locus, *Vigun09g13900* for the *W*

338 locus, and *Vigun10g163900* for the *H* locus (Herniter *et al.* 2019).

339   We first sought to confirm the previously proposed candidate loci for the *W*, *H*, and *C*

340 factors using genome-wide association study in the IITA core collection. Our mapping results for

341 pigmentation patterning yielded three distinct associated loci: one locus on Vu09 and two loci on

342 Vu10 (Figure 5, Table S9). The locus on Vu09 was associated with all five pigmentation

343 patterns, spanned 1.38 Mbp (Vu09:30,035,583-30,035,583), and contained 116 genes (Figure

344 5F). We identified 2_01960 as the tag SNP for this locus and found that the small eye, large

345 eye, and Holstein pigmentation patterns were associated with the A allele (MAF = 0.146) at this

346 locus while the Watson and full coat patterns were associated with the major frequency G allele.

347 Examination of the genes in this significant interval found that *Vigun09g13990*, a WD-repeat

348 gene proposed as a candidate gene for the *W* locus (Herniter *et al.* 2019), is present in the

349 region between the tag SNP and a linked upstream significant SNP. Furthermore, the candidate

350 gene is surrounded by other significant variants that are not linked with the tag SNP, suggesting

351 the presence of additional associated haplotypes segregating in this region.

352   The first locus on Vu10 was associated with the small eye, Watson, and full coat

353 pigmentation patterns. The tag SNP for this locus is 2_31919, with the C allele (MAF = 0.221)

14

354 associated with the small eye and Watson patterns and the major frequency T allele associated

355 only with the full coat pattern. The locus spans 12 KB (Vu10:38,454,820-38,467,161) and

356 overlaps with *Vigun10g165400*, an R2-R3 MYB transcription factor (Figure 5G). R2-R3 MYB

357 genes are reported to have a role in transcriptional regulation of flavonoid biosynthesis as part

358 of MYB-bHLH-WDR complexes (Xu *et al.* 2015).

359   The genome-wide association study for the Holstein pattern yielded two peaks on Vu10

360 that are several megabases apart. The first peak has tag SNP 2_15395, which is moderately

361 linked (R^2 = 0.346) to 2_23640, a SNP that is ~52 KB upstream of *Vigun10g163900*, an E3

362 ubiquitin ligase previously proposed as a candidate gene for the *H* locus (Herniter *et al.* 2019)

363 (Figure 5H). There are also several SNPs surrounding the gene that did not reach significance

364 per the conservative Bonferroni threshold but are close to the threshold and would have likely

365 been significant with increased power. The second peak spans 189 KB

366 (Vu10:41100865-41290623) and contains 23 genes, but does not yield a clear candidate gene.

367   Our results confirm that previously reported candidate genes for the *W* locus and *H*

368 locus could be mapped with the Core Collection GWAS panel. Although we were unable to map

369 the *C* locus on Vu07 in the five main seed pigmentation patterns, the locus was mapped in the

370 GWA mapping for hilum ring and eye absent patterns (Figure S8).

371

372 *The genetic architecture of seed color*

373 In cowpea, seed coat color varies on a continuous spectrum with several colors typically

374 described qualitatively such as white/cream, tan/buff, brown, red, blue, purple, and black.

375 Previous genetic studies of the segregation of seed coat color phenotypes revealed that coat

376 color is inherited independently of pigmentation pattern and is likely controlled by at least five

377 unlinked major genes (Saunders 1960; Drabo *et al.* 1988). To date, only the locus controlling

378 black seed color, the *Bl* locus, has been identified (Herniter *et al.* 2018). To attempt to identify

379 the loci associated with additional seed coat colors, we mapped the genetic basis of the red,

tan, black, blue, and white seed coat colors using GWAS in the Core Collection (Figure 6). In total, we found 4 major peaks associated with seed coat color: one each on Vu03 (red), Vu05 (black), Vu06 (blue), and Vu10 (No color/white).

We first sought to confirm if the Vu05 locus previously proposed to be associated with black seed coat color (Herniter *et al.* 2018) could be mapped via GWAS in the core collection. The GWAS for black seed coat yielded a major 121 KB peak on Vu05 (Vu05:3,104,538-3,225,877), spanning 13 genes (Figure 6A, F). The locus was also associated with the tan seed coat color (Figure 6C). The minor C allele at tag SNP 2_19309 (MAF = 0.463) was associated with the black seed coat phenotype while the major T allele was associated with the tan seed coat phenotype. Among the genes in the region, we found *Vigun05g039500*, a R2-R3 MYB transcription factor previously identified as a candidate gene for the *Bl* locus, which is associated with black seed coat and purple pod tip (Herniter *et al.* 2018).

The peak on Vu03 is associated with both the red and tan seed colors (Figure 6B-C, G), spans 274 KB (Vu03:10933603-11208404) and contains 25 genes. The T allele (MAF = 0.102) at the tag SNP 2_20787 is associated with red seed color, while the major frequency G allele is associated with the tan color. The locus maps to a pair of anthocyanin reductase (ANR) genes: *Vigun03g118700/Vigun03g118800*, with the tag SNP a predicted missense mutation in the former gene. ANR acts in the biosynthesis of flavan-3-ols and proanthocyanidins (condensed tannins) in plants (Xie *et al.* 2003). Mutations in ANR in *Arabidopsis thaliana* have been shown to result in the accumulation of anthocyanins in the seed coat (Devic *et al.* 1999), making this locus a compelling candidate for the red seed color phenotype.

The GWAS for blue seed color resulted in a ~155 KB peak on Vu06 (Vu06:13491073-13645944) that spans 5 genes (Figure 6D). SNP 2_22603, which is strongly linked to the tag SNP and lies directly upstream, is a predicted synonymous or intron mutation within the gene *Vigun06g030900* (Figure 6H). This gene is a homolog of *AT3G59030*

16

405 (*TRANSPARENT TESTA12*) which has been shown to be required for sequestration of

406 flavonoids in the seed coat epithelium tissue of *A. thaliana (Debeaujon et al. 2001)*.

407   Our attempt to map the genetic basis of white/cream seed color yielded one of the peaks

408 on Vu10 (2_31919) observed in seed coat patterning GWAS (Figure 5). Additionally, we did not

409 observe any significant SNPs in the GWA mapping of brown seed coat color, likely due the fact

410 that this phenotype can be heat-induced (Pottorff *et al.* 2014). In total, our GWAS was able to

411 map a previously identified locus for the black seed coat color and propose novel loci

412 associated with the red, tan, and blue seed coat colors.

413

414 *Mapping of additional phenotypes*

415 Including the seed coat pigmentation pattern and color phenotypes discussed previously, we

416 attempted to map the genetic basis of 71 distinct phenotypes in total. Altogether, 42 of the

417 genome-wide association studies had one or more significant SNPs (range 1 to 106 SNPs,

418 median 8). Of the 748 SNPs that were associated with variation in a phenotype, 698 were

419 associated with variation in only one phenotype and 50 SNPs were associated with more than

420 one phenotype (Table S9).

421   We found numerous interesting association peaks with clear underlying candidate

422 genes. For instance, GWAS for presence of flower pigmentation had a significant SNP adjacent

423 to *Vigun07g148800*, an E3 ubiquitin ligase. Although E3 ubiquitin-ligases are implicated in

424 numerous biological processes, they have been reported to be negative regulators of the

425 flavonoid biosynthesis pathway transcriptional activation complex (Shin *et al.* 2015; Herniter *et*

426 *al.* 2019). We also found candidate genes associated with disease resistance. GWAS for

427 incidence of rust yielded two peaks with clear candidates: a significant SNP on Vu02 in a

428 12-oxophytodienoate reductase (*Vigun02g192500*) involved in the Jasmonic acid biosynthesis

429 pathway and a very strong peak on Vu09 spanning 90 KB (Vu09:2156441-2246970) containing

430 three TIR-NBS-LRR family disease resistance genes with substantial sequence similarity:

17

431 *Vigun09g027450*, *Vigun09g027600* (DSC1), and *Vigun09g027700* (suppressor of npr1-1). In

432 sum, our analyses demonstrate the utility of the genotyped IITA Cowpea Core Collection for

433 discovering variants associated with phenotypic variation in cowpea.

434

435 **Discussion**

436 The purpose of this study was to utilize the recently reported SNP genotyping data to examine

437 population structure and the potential for genome-wide association mapping in the IITA Cowpea

438 Core Collection. The accessions in this collection are representative of the genotypic and

439 phenotypic diversity of extant cultivated cowpea germplasm. To our knowledge, the dataset

440 constitutes the largest cohort of cowpea that have been genotyped genome-wide with a single

441 technology and we anticipate that the resources linked to this work will be of considerable utility

442 to the cowpea research and breeding community.

443

444 *Genetic relationships in global cowpea*

445 Our analysis of population structure revealed that global cowpea is broadly genetically

446 differentiated into two major subpopulations defined by geography, with separation between

447 germplasm collected in Eastern and Southern Africa (ESA cluster) from germplasm collected in

448 West, Central, and North Africa (WNCA cluster). The genetic differentiation between germplasm

449 from West Africa and East Africa has been observed previously in smaller subsets of globally

450 collected cowpea germplasm (Huynh *et al.* 2013; Xiong *et al.* 2016; Muñoz-Amatriaín *et al.*

451 2017, 2021; Fatokun *et al.* 2018; Herniter *et al.* 2020), and our observations confirm that this is

452 the major organizing principle of cowpea diversity. However, we did observe that between

453 9-12% (depending on the number of subpopulations considered) of accessions collected in

454 West Africa were assigned to the ESA cluster (cluster 1). While this pattern is consistent with

455 the proposed domestication of cowpea in West Africa, our data cannot rule out the possibility of

456 a separate domestication center in East Africa. Additional study of wild cowpea in both West

18

457 Africa and East Africa could be useful for reconstructing the mode and tempo of domestication

458 of this crop.

459   Although the IITA Cowpea Core Collection primarily consists of traditional cultivars from

460 West Africa, the population structure analyses did provide some insight into the dispersal of

461 cowpea germplasm around the globe. The majority of germplasm from Latin America and the

462 Caribbean, North America, and Oceania clustered within the ESA cluster, suggesting dispersal

463 to these locales from Eastern and Southern Africa. In contrast, germplasm collected from East

464 Asia, Melanesia, and Southeast Asia was mostly assigned to cluster 3 and thus likely

465 represents migrants dispersed out of North Africa, perhaps through the Sabaean Lane, an

466 ancient trade route connecting East Africa with modern-day India (Murdock 1959; Herniter *et al.*

467 2020). Further, our analyses indicated a limited role of cluster 2 (West Africa specific) ancestry

468 in the dispersal of cowpea out of Africa, with only a few accessions, all present in Latin America

469 and the Caribbean, assigned to this cluster when considering more than 2 subpopulations. Our

470 analysis of phenotypic differences between the ESA and WNCA clusters suggested that ESA

471 germplasm is associated with earlier flowering, suggesting that this phenotype was important for

472 the worldwide dispersal of cowpea.

473   Compared to other surveys of global population structure in cowpea, our analysis

474 yielded fewer "optimal" clusters. We attribute this result to our use of the Evanno method

475 (Evanno *et al.* 2005) to decide the number of subpopulations, which has been shown previously

476 to be biased to select K = 2 (Janes *et al.* 2017). Therefore, we considered up to two additional

477 subpopulations, identifying a cluster of accessions collected from the Nile Delta in Northern

478 Egypt in close proximity to the Mediterranean Sea and a cluster of early flowering accessions in

479 West Africa, both of which were also discovered in the UCR Minicore, a mini-core subset of a

480 cowpea diversity collection held at the University of California, Riverside (Muñoz-Amatriaín *et al.*

481 2021). In contrast to previous studies, we did not observe subpopulation clusters consisting of

482 only breeding lines or clusters consisting solely of germplasm collected outside of Africa. It is

19

483 possible that these signals are also present in the data and would have been resolved if we had

484 improved sampling from continents beside Africa, if we had considered additional

485 subpopulations, or if we had conducted the analysis omitting the traditional cultivars from West

486 Africa that constitute the bulk of the collection. Our analysis also did not identify the cluster of

487 yardlong bean (subspecies sesquipedalis) previously identified in other collections of global

488 germplasm (Muñoz-Amatriaín *et al.* 2021) as the IITA Cowpea Core Collection only has three

489 accessions known to be from this group.

490   In addition to the patterns of population structure, we found that a substantial number of

491 accessions are nearly identical to one or more other accessions in the Core Collection. Our

492 analysis indicated that the majority of clusters of identical accessions consisted of lines

493 collected across geopolitical boundaries, with many clusters containing collections transcending

494 continents. A primary objective when selecting accessions for the core collection was to ensure

495 that the chosen lines were representative of the geographic and phenotypic variation present in

496 the full IITA cowpea collection (Mahalakshmi *et al.* 2007). Consequently, the occurrence of

497 dispersed identical lines within the core collection may be a result of sampling across

498 geographic gradients (i.e. sampling an accession that had been dispersed across a geopolitical

499 border in both the original and dispersed location in an effort to have representatives from all

500 regions where cowpea is found) or due to sampling of lines subject to phenotypic plasticity upon

501 selection of the core collection subset.

502

503 *A genome-wide association mapping panel for cowpea*

504 A primary objective of this study was to determine whether or not the IITA Cowpea Core

505 Collection constitutes a suitable panel for genome-wide association mapping in domesticated

506 cowpea. The rate of decay of linkage disequilibrium dictates the precision with which significant

507 associations can be linked to specific candidate genes. The decay of linkage disequilibrium (LD)

508 in the cowpea core collection is slow when compared to other plant systems (1.5 KB in maize

20

509 (Remington *et al.* 2001) and 10 KB in *A. thaliana (Kim et al. 2007)*), which is consistent with

510 observations in smaller previously genotyped collections (Muñoz-Amatriaín *et al.* 2021; Sodedji

511 *et al.* 2021). The slow decay of LD may be a feature common to legumes as similar rates have

512 been reported in cultivated soybean (Zhou *et al.* 2015). However, the slow decay of LD in

513 cowpea is partly due to strong population structure, as LD decays somewhat more quickly within

514 the WNCA population cluster.

515 Despite the slow decay of LD in cowpea, we were able to use genome-wide association

516 to identify hundreds of loci associated with variation in dozens of traits. We note that the

517 phenotype values used for our analyses were collected across many field studies spanning

518 years and locations, leaving us unable to control for the contribution of environmental variance

519 to phenotypic variance. We anticipate that GWAS of phenotypes collected in a single

520 experiment will have even greater success in identifying the genetic basis of trait variation using

521 the IITA Cowpea Core Collection.

522

523 *The genetic basis of seed coat pigmentation in cowpea*

524 Our efforts to map the genetic basis of seed pigmentation patterning using GWAS yielded

525 several candidate loci that were previously discovered using QTL mapping in F2 and advanced

526 RIL populations (Herniter *et al.* 2019). The consistency between the loci mapped as part of our

527 work and previous reports suggests that the loci proposed previously likely regulate seed

528 pigmentation pattern more generally in global cowpea and support the three locus model for the

529 genetic control of seed pigmentation patterning (Spillman and Sando 1930; Saunders 1960;

530 Drabo *et al.* 1988).

531 Through GWAS, we were able to map a second locus on Vu10 that likely contributes to

532 seed pigmentation patterning (Figure 5G). For this locus, we propose *Vigun10g165400*, a

533 R2-R3 MYB transcription factor, as a candidate gene. Analysis of patterns of linkage

534 disequilibrium at this locus suggests that this locus is distinct from the candidate gene at the *H*

21

535 locus despite being in close proximity (~110 KB downstream). This locus has been observed in

536 a previous attempt to map the genetic basis of seed pigmentation patterning (Herniter *et al.*

537 2019), but was disqualified due to not being observed in the associated haplotype blocks of all

538 the tested F2 populations. Our results support the hypothesis that multiple genes in close

539 proximity of Vu10 likely regulate seed pigmentation patterning. Further work is needed to

540 determine the molecular basis of this trait and to demonstrate the efficacy of the proposed

541 candidates.

542   Seed coat color has been studied for over 100 years in cowpea (Harland 1919, 1920)

543 and the existence of at least five genes affecting this trait have been proposed (Saunders 1960;

544 Drabo *et al.* 1988). However, the full genetic control of seed coat color has yet to be elucidated.

545 Here, we propose that the *R* locus controlling red seed color is an anthocyanin reductase

546 (ANR), either *Vigun03g118700* or *Vigun03g118800*. Although both genes lie under the

547 associated peak, the tag SNP is a predicted missense variant in *Vigun03g118700*, which is also

548 expressed at a high level during early seed development in the reference cultivar IT97K-499-35,

549 which has a black-eyed phenotype. In contrast, *Vigun03g118800* is expressed at a low level

550 across many tissues and developmental timepoints (Yao *et al.* 2016). Thus, we consider

551 *Vigun03g118700* to be the likely causal locus for the red seed coat color phenotype.

552   ANR is active in the flavonoid biosynthesis pathway, where it competes with

553 3-O-glucosyltransferase to convert cyanidin into condensed tannins or anthocyanins,

554 respectively (Kovinich *et al.* 2012). Mutations in ANR have been shown to result in the rapid

555 accumulation of anthocyanins during early seed development in *Arabidopsis thaliana* (Devic *et*

556 *al.* 1999) and knockdown of ANR in soybean yields a red seed coat phenotype (Kovinich *et al.*

557 2012). Our results suggest that a non-functional ANR gene is responsible for the red seed coat

558 phenotype in cowpea.

559   We also propose *Vigun06g030900*, a homolog of *TRANSPARENT TESTA12 (TT12)*, as

560 a candidate gene for the blue seed color phenotype. *TT12* is a MATE transporter which

22

561 mediates anthocyanin deposition in *A. thaliana* seeds (Debeaujon *et al.* 2001; Marinova *et al.*

562 2007). It remains to be seen whether blue seed color might be driven by changes in activity or

563 regulation of *Vigun06g030900*.

564   A full genetic model of seed pigmentation patterning and color in cowpea based on our

565 GWAS is presented in Table 1.

566

567 *Future directions*

568 The genotyped IITA Cowpea Core Collection represents a valuable tool for dissecting the

569 genetic basis of agronomic traits in cowpea. We anticipate that the resources produced by this

570 work will be useful for marker-assisted breeding and will facilitate the production of new cultivars

571 in response to consumer preference and changing environmental conditions. To facilitate the

572 latter aim, this collection could be used to identify climate-associated alleles, as many of the

573 traditional cultivars have known collection locations. This knowledge could be especially useful

574 for ensuring the resilience of cowpea into the future.

575

576 **Materials and Methods**

577 *Accession collection sites*

578 Collection sites for each accession were provided by the IITA (IITA 2023). World regions are

579 described according to the United Nations M49 standard (United Nations Statistics Division).

580

581 *Genotyping and variant filtering*

582 Accessions were genotyped using the Illumina Cowpea iSelect Consortium Array which assays

583 nearly 50,000 SNPs (Muñoz-Amatriaín *et al.* 2017) as described in (Close *et al.* 2023). All

584 alleles reported in this work belong to the Watson strand.

585   SNPs with greater than 95% of missing calls across samples were filtered first, followed

586 by filtering of samples with greater than 50% missing calls or more than 20% heterozygous calls

23

587 across sites. Near identical lines were identified by calculating pairwise Hamming distance

588 between samples with plink 1.9 (Chang *et al.* 2015), applying hierarchical clustering to the

589 distance matrix, and defining clusters of accessions using a static tree cut at height of 500. We

590 randomly selected one sample from each cluster to retain for subsequent analysis.

591   From the filtered SNP set, we also produced a set of SNPs in approximate linkage

592 equilibrium by first subsetting to SNPs with minor allele frequency greater than or equal to 0.01

593 with no more than 5% missing calls. The filtered SNPs were then pruned pairwise in 50 SNP

594 windows with a step size of 10 and an R^2 threshold of 0.20. All SNP filtering and pruning was

595 done using plink 1.9 (Chang *et al.* 2015). The pruned set, composed of 4,245 SNPs, was used

596 for all population structure analyses.

597

598 *Variant effect prediction*

599 The predicted effect of variants were determined using Ensembl Variant Effect Predictor v. 95.0

600 (McLaren *et al.* 2016) using the *V. unguiculata* v. 1.0 reference genome assembly and v. 1.2 of

601 the genome annotation (Lonardi *et al.* 2019).

602

603 *Genetic diversity statistics*

604 Allele frequencies and pairwise genetic distances were calculated using plink v.1.9 (Chang *et al.*

605 2015). Expected heterozygosity was calculated according to the following formula:

606 $\frac{1}{N} \sum_{n=1}^{N} (1 - p^2 + q^2)$, where N is the total number of sites and p and q are the frequencies of

607 the major and minor frequency alleles at each site, respectively.

608

609 *Population structure*

610 Population structure was analyzed using ADMIXTURE 1.3.0 (Alexander *et al.* 2009; Alexander

611 and Lange 2011) for K = 1 to 10 with 10 replicates per K. The most likely K value was

24

612 determined using the Evanno method (Evanno *et al.* 2005). ADMIXTURE runs were

613 post-processed and visualized using the pophelper 2.3.1 R package (Francis 2017). Samples

614 were assigned to the group contributing 70% or more of the sample's estimated ancestry

615 fraction. All other samples were considered to be unassigned.

616   Principal component analysis on allele frequencies was done using the adegenet v. 2.1.5

617 (Jombart 2008) R package. The mean allele frequency at a locus was substituted for missing

618 genotype calls.

619

620 *Phenotypic analyses*

621 The IITA Cowpea Characterization and Cowpea Evaluation phenotypic datasets were

622 downloaded from the International Institute of Tropical Agriculture (IITA) website (IITA 2023) and

623 consisted of 15 quantitative and 23 qualitative traits. The data were collected as part of the IITA

624 cowpea breeding program across multiple field trials spanning many years and thus represent

625 reference values used for breeding. For the analysis, seed color pigmentation traits were

626 defined according to the majority color of the seed and similar traits were combined as

627 described in Table S6. "Small eye" is synonymous with the "Eye 1" pattern and "Large eye" is

628 synonymous with the Eye 2 pattern described by (Herniter *et al.* 2019). "Self-colored" is

629 described herein as the synonymous "full coat" pattern (Singh and International Institute of

630 Tropical Agriculture 2014). Qualitative traits were coded into binary "case/control" variables

631 using the fastDummies 1.6.3 R package (Kaplan 2020) or considered to be discrete numerical

632 variables (Table S4). Differentiation in phenotype distribution between population structure

633 clusters was tested using Mann Whitney Wilcoxon tests for quantitative phenotypes and

634 Fisher's Exact test for qualitative phenotypes in R (R Core Team 2013). Two-sided tests were

635 used for all analyses and a Bonferroni adjusted threshold was used to determine significance at

636 α = 0.05.

637

25

Linkage disequilibrium

Pairwise linkage disequilibrium between markers was calculated with plink 1.9 (Chang *et al.* 2015). Since calculation of pairwise LD for all markers in a dataset can be computationally prohibitive, we only calculated LD between markers that were within 10 Mbp of each other. Genome-wide and chromosome-specific background levels of LD were determined by first calculating the mean R^2 value in 500 bp distance bins and then taking the median value across all bins.

Genome-wide association mapping

Univariate mixed linear models were fitted with GEMMA 0.98.4 (Zhou and Stephens 2012). The models included a centered relatedness matrix to correct for spurious associations due to population structure. Only sites with minor allele frequency greater than 0.01 and missing in fewer than 5% of samples were tested in this analysis. Continuous quantitative phenotypes were log-transformed prior to analysis. Qualitative phenotypes were encoded in binary as 1 or 0 corresponding to presence or absence of the trait. Bonferroni correction was used to correct p-values for multiple testing and significance was assessed at α = 0.05. Linked significant SNPs ($R^2 > 0.20$) within 1 Mbp were considered to define a single peak and the SNP with the lowest p-value per peak was assigned to be the "tag SNP." Candidate genes were identified by examining the overlaps between significant peaks and loci described in the *Vigna unguiculata* v.1.2 genome annotation (Lonardi *et al.* 2019).

Gene expression data

The *Vigna unguiculata* Gene Expression Atlas (VuGEA) data (Yao *et al.* 2016) were retrieved from the Legume Information System (Berendzen *et al.* 2021) at https://data.legumeinfo.org.

Data Availability

Acknowledgements

References

Abebe B. K., and M. T. Alemayehu, 2022 A review of the nutritional use of cowpea (Vigna unguiculata L. Walp) for human and animal diets. Journal of Agriculture and Food Research 10: 100383. https://doi.org/10.1016/j.jafr.2022.100383

Alexander D. H., J. Novembre, and K. Lange, 2009 Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19: 1655–1664. https://doi.org/10.1101/gr.094052.109

Alexander D. H., and K. Lange, 2011 Enhancements to the ADMIXTURE algorithm for individual

687   ancestry estimation. BMC Bioinformatics 12: 246. https://doi.org/10.1186/1471-2105-12-246

688 Andargie M., R. S. Pasquet, G. M. Muluvi, and M. P. Timko, 2013 Quantitative trait loci analysis

689   of flowering time related traits identified in recombinant inbred lines of cowpea (Vigna

690   unguiculata). Genome 56: 289–294. https://doi.org/10.1139/gen-2013-0028

691 Ba F. S., R. S. Pasquet, and P. Gepts, 2004 Genetic diversity in cowpea [Vigna unguiculata (L.)

692   Walp.] as revealed by RAPD markers. Genet. Resour. Crop Evol. 51: 539–550.

693   https://doi.org/10.1023/B:GRES.0000024158.83190.4e

694 Berendzen J., A. V. Brown, C. T. Cameron, J. D. Campbell, A. M. Cleary, *et al.*, 2021 The

695   legume information system and associated online genomic resources. Legum. Sci. 3.

696   https://doi.org/10.1002/leg3.74

697 Boukar O., N. Belko, S. Chamarthi, A. Togola, J. Batieno, *et al.*, 2019 Cowpea (Vigna

698   unguiculata): Genetics, genomics and breeding. Plant Breed. 138: 415–424.

699 Chang C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, *et al.*, 2015 Second-generation

700   PLINK: rising to the challenge of larger and richer datasets. Gigascience 4: 7.

701   https://doi.org/10.1186/s13742-015-0047-8

702 Chen Y., H. Xiong, W. Ravelombola, G. Bhattarai, C. Barickman, *et al.*, 2023 A Genome-Wide

703   Association Study Reveals Region Associated with Seed Protein Content in Cowpea.

704   Plants 12. https://doi.org/10.3390/plants12142705

705 Close T., O. Oyatomi, Y.-N. Guo, R. Paliwal, M. Muñoz-Amatriaín, *et al.*, 2023 SNP genotypes of

706   the international institute of tropical agriculture Cowpea Core. Dryad

707   https://datadryad.org/stash/dataset/doi:10.6086/D19Q37

708 Coulibaly S., R. S. Pasquet, R. Papa, and P. Gepts, 2002 AFLP analysis of the phenetic

28

709   organization and genetic diversity of Vigna unguiculata L. Walp. reveals extensive gene

710   flow between wild and domesticated types. Theor. Appl. Genet. 104: 358–366.

711   https://doi.org/10.1007/s001220100740

712 Debeaujon I., A. J. Peeters, K. M. Léon-Kloosterziel, and M. Koornneef, 2001 The

713   TRANSPARENT TESTA12 gene of Arabidopsis encodes a multidrug secondary

714   transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat

715   endothelium. Plant Cell 13: 853–871. https://doi.org/10.1105/tpc.13.4.853

716 Devic M., J. Guilleminot, I. Debeaujon, N. Bechtold, E. Bensaude, *et al.*, 1999 The BANYULS

717   gene encodes a DFR-like protein and is a marker of early seed coat development. Plant J.

718   19: 387–398. https://doi.org/10.1046/j.1365-313x.1999.00529.x

719 Dong J., Y. Song, B. Wang, X. Wu, Y. Wang, *et al.*, 2022 Identification of Genomic Regions

720   Associated with Fusarium Wilt Resistance in Cowpea. NATO Adv. Sci. Inst. Ser. E Appl.

721   Sci. 12: 6889. https://doi.org/10.3390/app12146889

722 Drabo I., T. A. O. Ladeinde, J. B. Smithson, and R. Redden, 1988 Inheritance of eye pattern and

723   seed coat colour in Cowpea (Vigna unguiculata [L.] Walp.). Plant Breed. 100: 119–123.

724 Evanno G., S. Regnaut, and J. Goudet, 2005 Detecting the number of clusters of individuals

725   using the software STRUCTURE: a simulation study. Mol. Ecol. 14: 2611–2620.

726 Fatokun C., G. Girma, M. Abberton, M. Gedil, N. Unachukwu, *et al.*, 2018 Genetic diversity and

727   population structure of a mini-core subset from the world cowpea (Vigna unguiculata (L.)

728   Walp.) germplasm collection. Sci. Rep. 8: 16035.

729 Food and Agriculture Organization of the United Nations, 2022 FAOSTAT Statistical Database.

730   Food and Agricultural Organization of the United Nations- FAOSTAT.

Foyer C. H., H.-M. Lam, H. T. Nguyen, K. H. M. Siddique, R. K. Varshney, *et al.*, 2016 Neglecting legumes has compromised human health and sustainable food production. Nat Plants 2: 16112. https://doi.org/10.1038/nplants.2016.112

Francis R. M., 2017 pophelper: An R package and web app to analyse and visualise population structure. Molecular Ecology Resources 17: 27–32.

Hall A. E., 2004 Breeding for adaptation to drought and heat in cowpea. Eur. J. Agron. 21: 447–454. https://doi.org/10.1016/j.eja.2004.07.005

Harland S. C., 1919 Inheritance of certain characters in the cowpea (vigna sinensis). J. Genet. 8: 101–132. https://doi.org/10.1007/BF02983490

Harland S. C., 1920 Inheritance of certain characters in the cowpea (Vigna sinensis). II. J. Genet. 10: 193–205. https://doi.org/10.1007/BF03007981

Herniter I. A., M. Muñoz-Amatriaín, S. Lo, Y.-N. Guo, and T. J. Close, 2018 Identification of Candidate Genes Controlling Black Seed Coat and Pod Tip Color in Cowpea (Vigna unguiculata [L.] Walp). G3 Genes|Genomes|Genetics 8: 3347–3355.

Herniter I. A., R. Lo, M. Muñoz-Amatriaín, S. Lo, Y.-N. Guo, *et al.*, 2019 Seed Coat Pattern QTL and Development in Cowpea (Vigna unguiculata [L.] Walp.). Front. Plant Sci. 10: 1346. https://doi.org/10.3389/fpls.2019.01346

Herniter I. A., M. Muñoz-Amatriaín, and T. J. Close, 2020 Genetic, textual, and archeological evidence of the historical global spread of cowpea (Vigna unguiculata [L.] Walp.). Legume Science 2. https://doi.org/10.1002/leg3.57

Huynh B.-L., T. J. Close, P. A. Roberts, Z. Hu, S. Wanamaker, *et al.*, 2013 Gene pools and the genetic architecture of domesticated Cowpea. Plant Genome 6: lantgenome2013.03.0005.

753   https://doi.org/10.3835/plantgenome2013.03.0005

754 Huynh B.-L., W. C. Matthews, J. D. Ehlers, M. R. Lucas, J. R. P. Santos, *et al.*, 2016 A major

755   QTL corresponding to the Rk locus for resistance to root-knot nematodes in cowpea (Vigna

756   unguiculata L. Walp.). Theor. Appl. Genet. 129: 87–95.

757   https://doi.org/10.1007/s00122-015-2611-0

758 Huynh B.-L., J. D. Ehlers, B. E. Huang, M. Muñoz-Amatriaín, S. Lonardi, *et al.*, 2018 A

759   multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and

760   improvement of cowpea (Vigna unguiculata L. Walp.). Plant J. 93: 1129–1142.

761   https://doi.org/10.1111/tpj.13827

762 IITA, 2023 IITA Cowpea Collection. IITA Cowpea Collection.

763 Janes J. K., J. M. Miller, J. R. Dupuis, R. M. Malenfant, J. C. Gorrell, *et al.*, 2017 The K = 2

764   conundrum. Mol. Ecol. 26: 3594–3602.

765 Jombart T., 2008 adegenet: a R package for the multivariate analysis of genetic markers.

766   Bioinformatics 24: 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

767 Kaplan J., 2020 fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from

768   Categorical Variables

769 Kim S., V. Plagnol, T. T. Hu, C. Toomajian, R. M. Clark, *et al.*, 2007 Recombination and linkage

770   disequilibrium in Arabidopsis thaliana. Nat. Genet. 39: 1151–1155.

771 Kovinich N., A. Saleem, J. T. Arnason, and B. Miki, 2012 Identification of two anthocyanidin

772   reductase genes and three red-brown soybean accessions with reduced anthocyanidin

773   reductase 1 mRNA, activity, and seed coat proanthocyanidin amounts. J. Agric. Food

774   Chem. 60: 574–584. https://doi.org/10.1021/jf2033939

775 Kpoviessi A. D., S. Agbahoungba, E. E. Agoyi, A. Badji, A. E. Assogbadjo, *et al.*, 2022

776   Application of multi-locus GWAS for the detection of bruchid resistance loci in cowpea (

777   *Vigna unguiculata*). Plant Breed. 141: 439–450. https://doi.org/10.1111/pbr.13014

778 Liang Q., M. Muñoz-Amatriaín, S. Shu, S. Lo, X. Wu, *et al.*, 2023 A view of the pan-genome of

779   domesticated Cowpea (Vigna unguiculata [L.] Walp.). Plant Genome e20319.

780   https://doi.org/10.1002/tpg2.20319

781 Lo S., M. Muñoz-Amatriaín, O. Boukar, I. Herniter, N. Cisse, *et al.*, 2018 Identification of QTL

782   controlling domestication-related traits in cowpea (Vigna unguiculata L. Walp). Sci. Rep. 8:

783   6261.

784 Lo S., M. Muñoz-Amatriaín, S. A. Hokin, N. Cisse, P. A. Roberts, *et al.*, 2019 A genome-wide

785   association and meta-analysis reveal regions associated with seed size in cowpea [Vigna

786   unguiculata (L.) Walp]. Theor. Appl. Genet. 132: 3079–3087.

787 Lo S., C. Fatokun, O. Boukar, P. Gepts, T. J. Close, *et al.*, 2020 Identification of QTL for

788   perenniality and floral scent in cowpea (Vigna unguiculata [L.] Walp.). PLoS One 15:

789   e0229167.

790 Lo S., T. Parker, M. Muñoz-Amatriaín, J. C. Berny-Mier Y Teran, J. Jernstedt, *et al.*, 2021

791   Genetic, anatomical, and environmental patterns related to pod shattering resistance in

792   domesticated cowpea [Vigna unguiculata (L.) Walp]. J. Exp. Bot. 72: 6219–6229.

793 Lonardi S., M. Muñoz-Amatriaín, Q. Liang, S. Shu, S. I. Wanamaker, *et al.*, 2019 The genome of

794   cowpea (Vigna unguiculata [L.] Walp.). Plant J. 98: 767–782.

795   https://doi.org/10.1111/tpj.14349

796 Mahalakshmi V., Q. Ng, M. Lawson, and R. Ortiz, 2007 Cowpea [Vigna unguiculata (L.) Walp.]

797   core collection defined by geographical, agronomical and botanical descriptors. Plant

798   Genet. Resour. 5: 113–119. https://doi.org/10.1017/S1479262107837166

799 Marinova K., L. Pourcel, B. Weder, M. Schwarz, D. Barron, *et al.*, 2007 The Arabidopsis MATE

800   transporter TT12 acts as a vacuolar flavonoid/H+ -antiporter active in

801   proanthocyanidin-accumulating cells of the seed coat. Plant Cell 19: 2023–2038.

802   https://doi.org/10.1105/tpc.106.046029

803 McLaren W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, *et al.*, 2016 The Ensembl Variant

804   Effect Predictor. Genome Biol. 17: 122.

805 Mortimore M., B. Singh, F. Harris, and S. Blade, 1997 Cowpea in traditional cropping systems

806 Muñoz-Amatriaín M., H. Mirebrahim, P. Xu, S. I. Wanamaker, M. Luo, *et al.*, 2017 Genome

807   resources for climate-resilient cowpea, an essential crop for food security. Plant J. 89:

808   1042–1054. https://doi.org/10.1111/tpj.13404

809 Muñoz-Amatriaín M., S. Lo, I. A. Herniter, O. Boukar, C. Fatokun, *et al.*, 2021 The UCR

810   Minicore: a resource for cowpea research and breeding. Legume Science 3.

811   https://doi.org/10.1002/leg3.95

812 Murdock G. P., 1959 *Africa : its peoples and their culture history*. McGraw-Hill Book Company,

813   New York.

814 Nap J. P., and T. Bisseling, 1990 Developmental biology of a plant-prokaryote symbiosis: the

815   legume root nodule. Science 250: 948–954. https://doi.org/10.1126/science.250.4983.948

816 Olatoye M. O., Z. Hu, and P. O. Aikpokpodion, 2019 Epistasis Detection and Modeling for

817   Genomic Selection in Cowpea (Vigna unguiculata L. Walp.). Front. Genet. 10: 677.

818 Padulosi S., and N. Q. Ng, 1997 Origin, taxonomy, and morphology of Vigna unguiculata (L.)

819   Walp. Advances in cowpea research.

820 Pasquet R., and S. Padulosi, 2012 Genus Vigna and Cowpea (V. unguiculata [L.] Walp.)

821   taxonomy: current status and prospects, pp. 66–87 in *Innovative research along the*

822   *cowpea value chain*, edited by Boukar O, Coulibaly O, Fatokun CA, Lopez K, Tamo M.

823   International Institute of Tropical Agriculture (IITA).

824 Paudel D., R. Dareus, J. Rosenwald, M. Muñoz-Amatriaín, and E. F. Rios, 2021 Genome-wide

825   association study reveals candidate genes for flowering time in Cowpea (Vigna unguiculata

826   [L.] Walp.). Front. Genet. 12: 667038.

827 Pottorff M., P. A. Roberts, T. J. Close, S. Lonardi, S. Wanamaker, *et al.*, 2014 Identification of

828   candidate genes and molecular markers for heat-induced brown discoloration of seed coats

829   in cowpea [Vigna unguiculata (L.) Walp]. BMC Genomics 15: 328.

830   https://doi.org/10.1186/1471-2164-15-328

831 Ravelombola W., A. Shi, and B.-L. Huynh, 2021 Loci discovery, network-guided approach, and

832   genomic prediction for drought tolerance index in a multi-parent advanced generation

833   intercross (MAGIC) cowpea population. Hortic Res 8: 24.

834 Ravelombola W., A. Shi, B.-L. Huynh, J. Qin, H. Xiong, *et al.*, 2022 Genetic architecture of salt

835   tolerance in a Multi-Parent Advanced Generation Inter-Cross (MAGIC) cowpea population.

836   BMC Genomics 23: 100.

837 R Core Team, 2013 R: A Language and Environment for Statistical Computing

838 Reckling M., J.-M. Hecker, G. Bergkvist, C. A. Watson, P. Zander, *et al.*, 2016 A cropping

839   system assessment framework—Evaluating effects of introducing legumes into crop

840   rotations. Eur. J. Agron. 76: 186–197. https://doi.org/10.1016/j.eja.2015.11.005

841 Remington D. L., J. M. Thornsberry, Y. Matsuoka, L. M. Wilson, S. R. Whitt, *et al.*, 2001

842   Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc.

34

843   Natl. Acad. Sci. U. S. A. 98: 11479–11484. https://doi.org/10.1073/pnas.201394398

844 Santos J. R. P., A. D. Ndeve, B.-L. Huynh, W. C. Matthews, and P. A. Roberts, 2018 QTL

845   mapping and transcriptome analysis of cowpea reveals candidate genes for root-knot

846   nematode resistance. PLoS One 13: e0189185.

847   https://doi.org/10.1371/journal.pone.0189185

848 Saunders, 1960 Inheritance in the cowpea (Vigna sinensis Endb.) II: Seed coat colour pattern;

849   flower, plant and pod colour. S. Afr. J. Agric. Ext. (SAJAE).

850   https://doi.org/10.10520/aja05858860_201

851 Shin D. H., M. Cho, M. G. Choi, P. K. Das, S.-K. Lee, *et al.*, 2015 Identification of genes that

852   may regulate the expression of the transcription factor production of anthocyanin pigment 1

853   (PAP1)/MYB75 involved in Arabidopsis anthocyanin biosynthesis. Plant Cell Rep. 34:

854   805–815. https://doi.org/10.1007/s00299-015-1743-7

855 Singh B. B., and International Institute of Tropical Agriculture, 2014 *Cowpea : the food legume*

856   *of the 21st century*. Crop Science Society of America, Madison, WI.

857 Sodedji F. A. K., S. Agbahoungba, E. E. Agoyi, M. K. Kafoutchoni, J. Choi, *et al.*, 2021 Diversity,

858   population structure, and linkage disequilibrium among cowpea accessions. Plant Genome

859   14: e20113. https://doi.org/10.1002/tpg2.20113

860 Spillman W. J., 1911 Inheritance of the ``Eye'' in Vigna. Am. Nat. 45: 513–523.

861 Spillman W. J., and W. J. Sando, 1930 Mendelian factors in the cowpea (Vigna species). Pap.

862   Mich. Acad. Sci. Arts Lett. 11: 249–283.

863 Tarawali, S., Singh, B., Peters, M. & Blade, S., 1997 Cowpea haulms as fodder, pp. 313–325 in

864   *Advances in Cowpea Research.*, edited by B. B. Singh D. R. M. R. A. K. E. D. IITA.

865 United Nations Statistics Division, Standard country or area codes for statistical use (M49).

866   United Nations Statistics Division.

867 Vaillancourt R. E., and N. F. Weeden, 1992 CHLOROPLAST DNA POLYMORPHISM

868   SUGGESTS NIGERIAN CENTER OF DOMESTICATION FOR THE COWPEA, VIGNA

869   UNGUICULATA (LEGUMINOSAE). Am. J. Bot. 79: 1194–1199.

870   https://doi.org/10.1002/j.1537-2197.1992.tb13716.x

871 Weng Y., J. Qin, S. Eaton, Y. Yang, W. S. Ravelombola, *et al.*, 2019 Evaluation of Seed Protein

872   Content in USDA Cowpea Germplasm. HortScience 54: 814–817.

873 Wu X., T. Sun, W. Xu, Y. Sun, B. Wang, *et al.*, 2021 Unraveling the Genetic Architecture of Two

874   Complex, Stomata-Related Drought-Responsive Traits by High-Throughput Physiological

875   Phenotyping and GWAS in Cowpea (Vigna. Unguiculata L. Walp). Front. Genet. 12:

876   743758. https://doi.org/10.3389/fgene.2021.743758

877 Xie D.-Y., S. B. Sharma, N. L. Paiva, D. Ferreira, and R. A. Dixon, 2003 Role of anthocyanidin

878   reductase, encoded by BANYULS in plant flavonoid biosynthesis. Science 299: 396–399.

879   https://doi.org/10.1126/science.1078540

880 Xiong H., A. Shi, B. Mou, J. Qin, D. Motes, *et al.*, 2016 Genetic Diversity and Population

881   Structure of Cowpea (Vigna unguiculata L. Walp). PLoS One 11: e0160941.

882   https://doi.org/10.1371/journal.pone.0160941

883 Xiong H., J. Qin, A. Shi, B. Mou, D. Wu, *et al.*, 2018 Genetic differentiation and diversity upon

884   genotype and phenotype in cowpea (Vigna unguiculata L. Walp.). Euphytica 214: 4.

885   https://doi.org/10.1007/s10681-017-2088-9

886 Xu W., C. Dubos, and L. Lepiniec, 2015 Transcriptional control of flavonoid biosynthesis by

887   MYB-bHLH-WDR complexes. Trends Plant Sci. 20: 176–185.

888   https://doi.org/10.1016/j.tplants.2014.12.001

889 Yao S., C. Jiang, Z. Huang, I. Torres-Jerez, J. Chang, *et al.*, 2016 The Vigna unguiculata Gene

890   Expression Atlas (VuGEA) from de novo assembly and quantification of RNA-seq data

891   provides insights into seed maturation mechanisms. Plant J. 88: 318–327.

892 Zhou X., and M. Stephens, 2012 Genome-wide efficient mixed-model analysis for association

893   studies. Nat. Genet. 44: 821–824. https://doi.org/10.1038/ng.2310

894 Zhou Z., Y. Jiang, Z. Wang, Z. Gou, J. Lyu, *et al.*, 2015 Resequencing 302 wild and cultivated

895   accessions identifies genes related to domestication and improvement in soybean. Nat.

896   Biotechnol. 33: 408–414. https://doi.org/10.1038/nbt.3096

897

A

Region	Accessions
West Africa	935
East Africa	267
South Asia	158
Central Africa	144
Southern Africa	127
North America	110
North Africa	74
Latin America & the Caribbean	68
West Asia	29
Southern Europe	26
Eastern Europe	13
Southeast Asia	12
East Asia	8
Australia & New Zealand	7
Melanesia	6
Northern Europe	3
Western Europe	1
Micronesia	1

Germplasm
- Trad. cultivar
- Breeding
- Weedy
- Wild
- Unknown

B H_e

C Prop. SNPs seg.

D Number of sites — minor allele frequency:
(0, 0.05], (0.05, 0.1], (0.1, 0.15], (0.15, 0.2], (0.2, 0.25], (0.25, 0.3], (0.3, 0.35], (0.35, 0.4], (0.4, 0.45], (0.45, 0.5]

E Dist. (x 10^4)

Region
- West Africa
- Central Africa
- North Africa
- East Africa
- Southern Africa
- Asia
- Americas
- Europe
- Oceania

Figure 1. Genetic diversity of the IITA Core Cowpea Collection
A) Region of origin and germplasm status for genotyped accessions. Labels beside each bar indicate the number of accessions per region. B) Average expected heterozygosity per region of origin calculated using all sites. C) Proportion of SNPs segregating per region. D) Folded allele frequency spectrum by germplasm status. E) Heatmap of pairwise genetic distances. Row labels indicate germplasm status and column labels indicate region of origin.
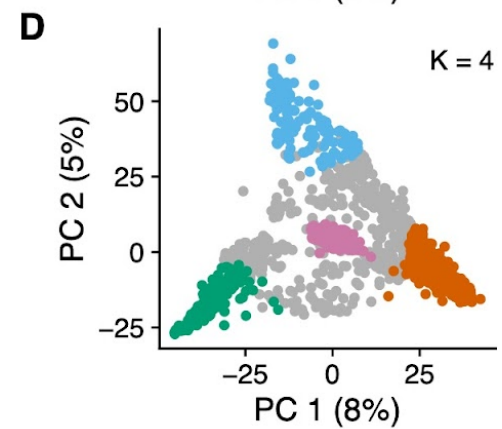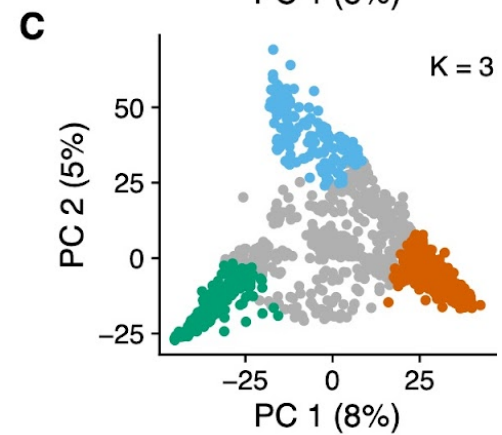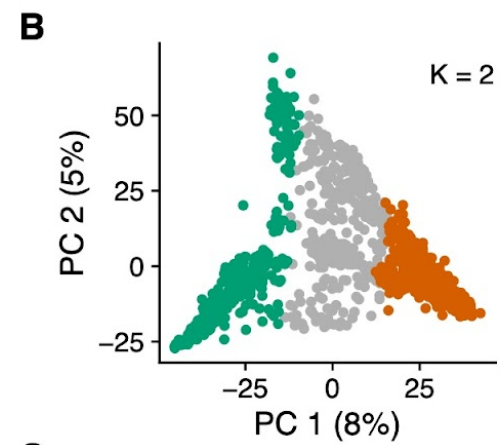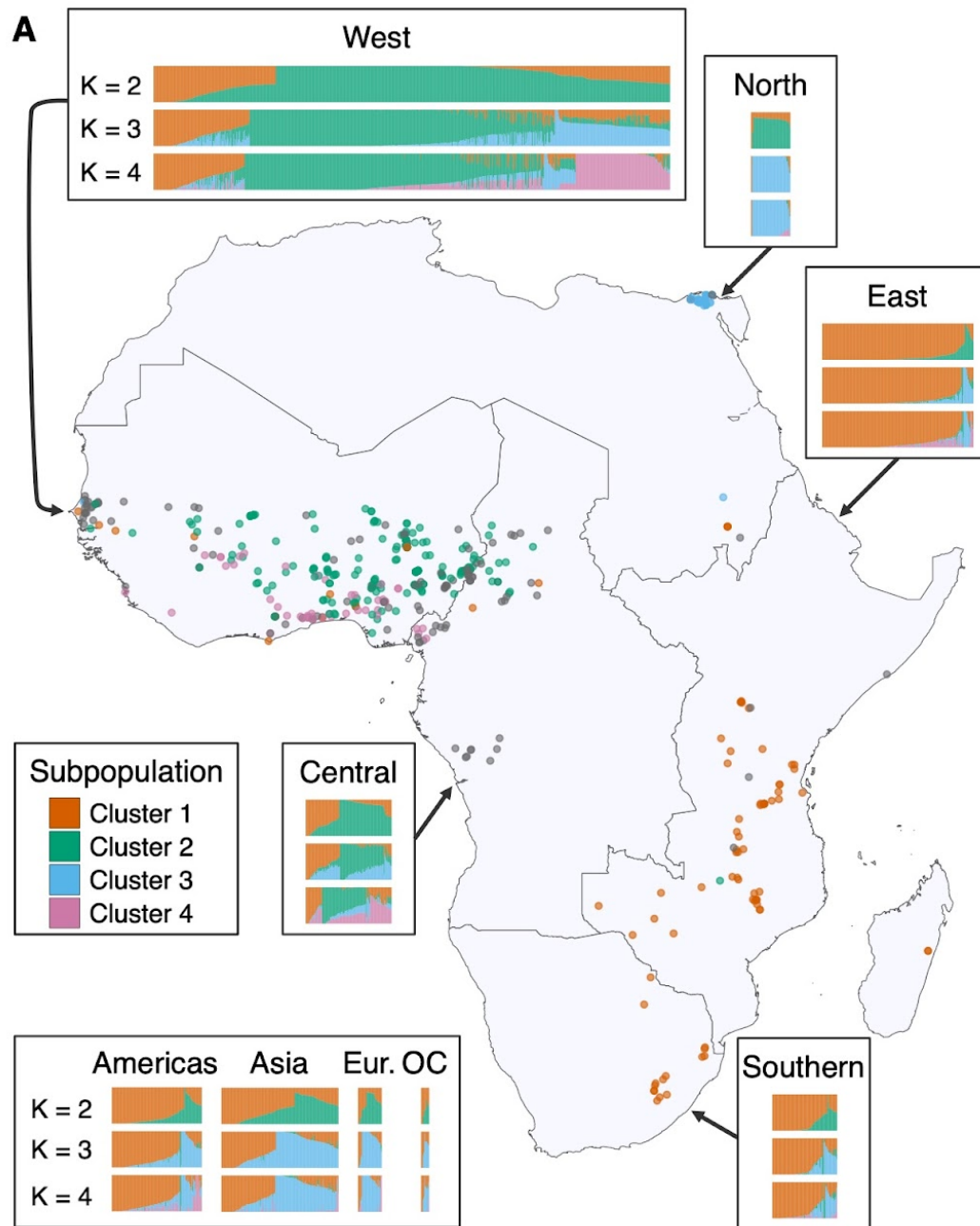
Figure 2. Population structure in global cowpea germplasm
A) Inferred ancestry for all genotyped individuals from the core collection and map of collection sites for 545 traditional cultivars with known collection sites in Africa. Accessions are grouped by subregion according to the United Nations geoscheme and each accession with a known collection site is colored according to subpopulation assignment when modeling K = 4 subpopulations. Accompanying barplots specify the proportions of ancestry attributed to each subpopulation per individual when modeling K = 2, K = 3, and K = 4 subpopulations, respectively. Eur. = Europe, OC = Oceania. B - D) Principal components analysis of allele frequencies with accessions colored by subpopulation for K = 2 - 4.
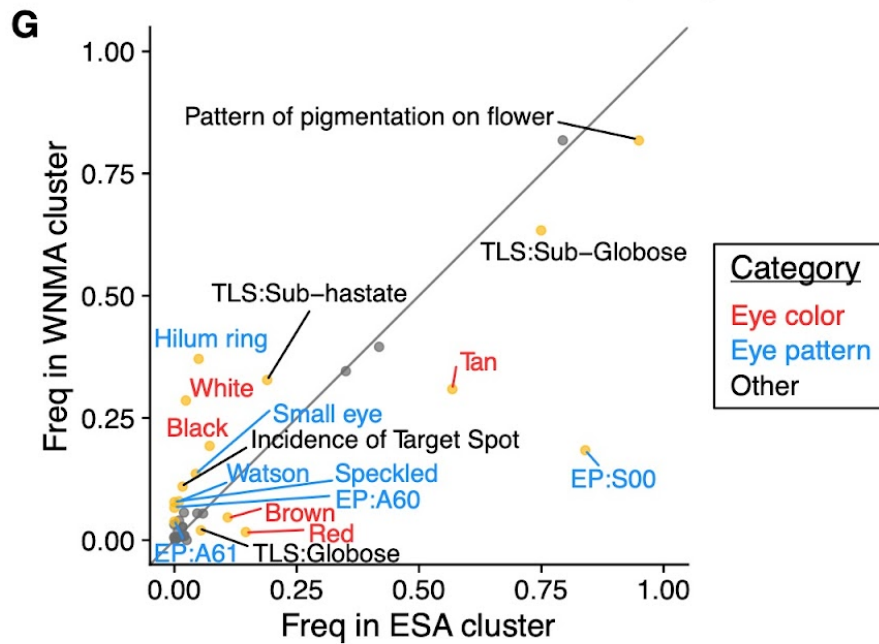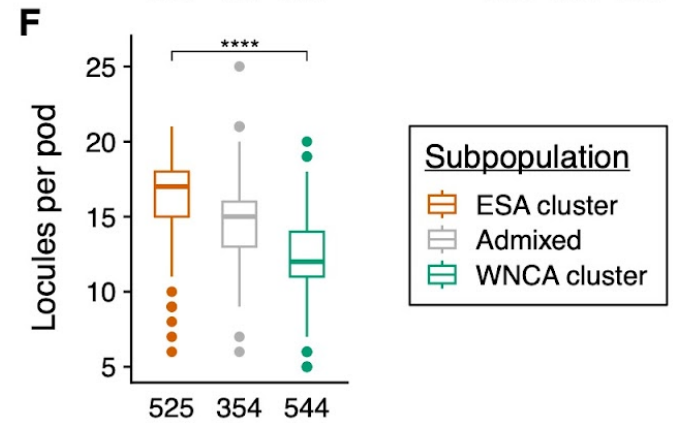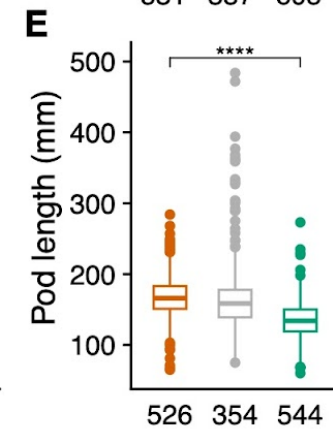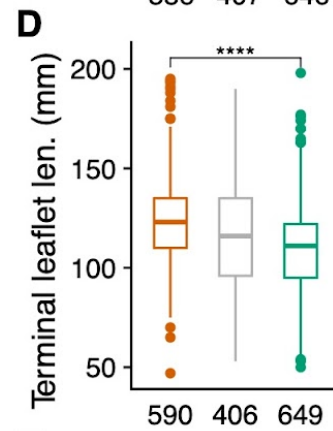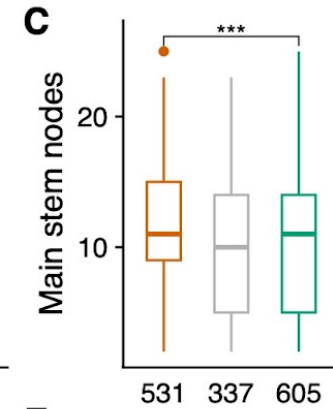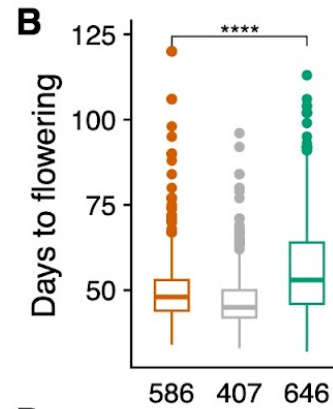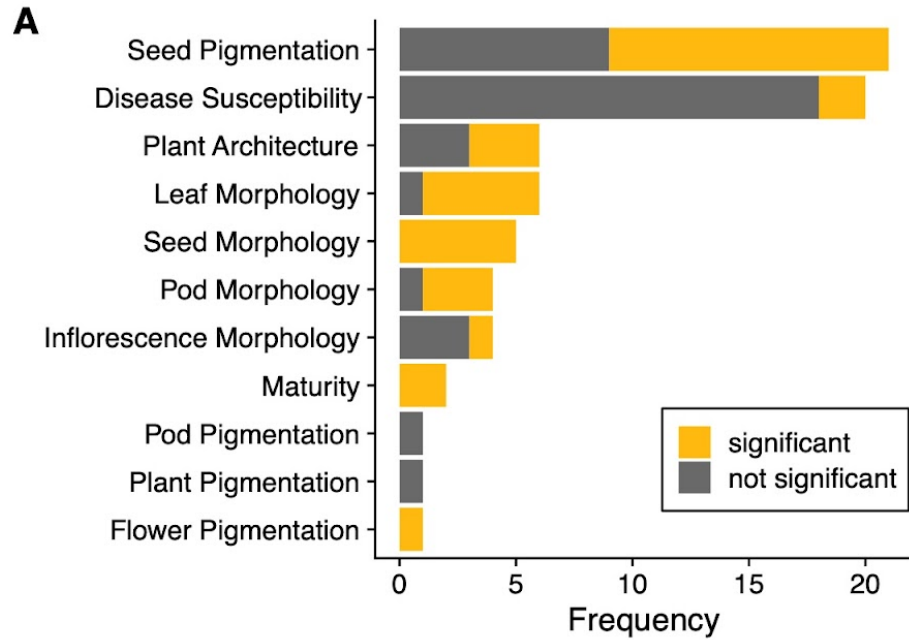
Figure 3. Phenotypic distributions in subpopulation clusters
A) Number of phenotypes by category, colored by the number of phenotypes with statistically significant difference between the ESA and WNCA clusters. B-F) Tukey's boxplots of representative phenotype distributions by subpopulation cluster. Numbers below each boxplot denote N. Additional phenotypes are presented in Figure S6. G) Qualitative trait frequencies in the two subpopulation clusters. Grey reference line indicates equal frequency in both clusters. Points are colored according to whether there is a significant difference in trait frequency (goldenrod) between the two subpopulation clusters. Labels correspond to phenotype names and are colored by category. Abbreviations: EP- eye pattern, TLS- terminal leaf shape.
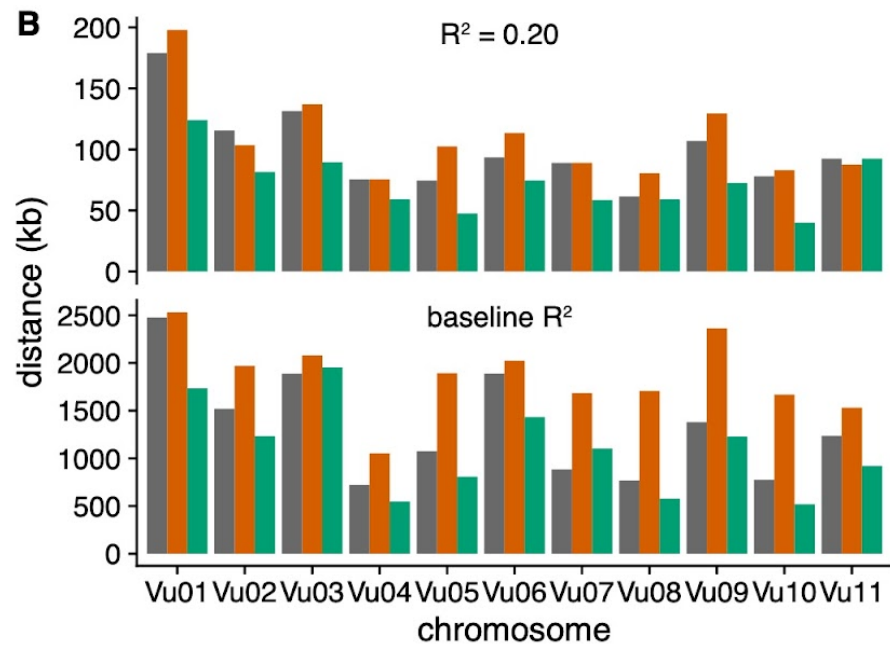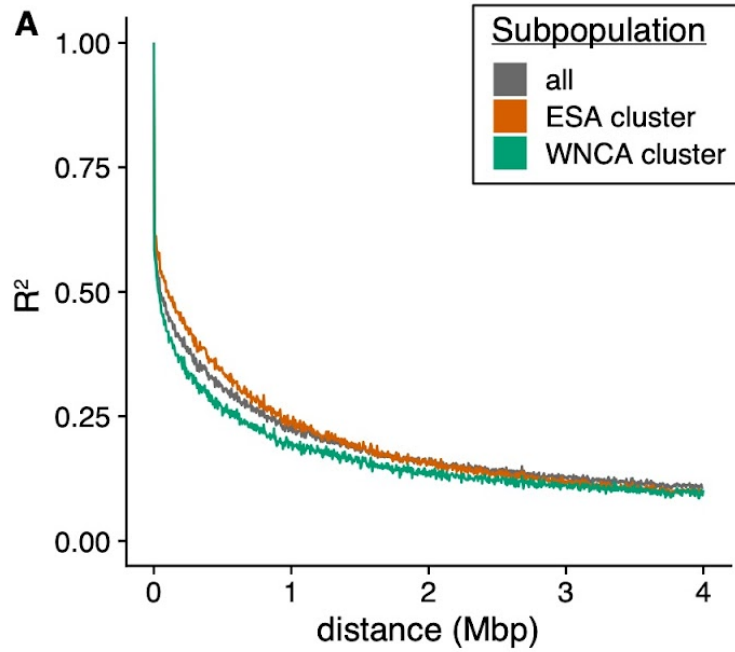
Figure 4. Linkage disequilibrium in the cowpea genome.
A) Mean pairwise linkage disequilibrium (LD) between SNPs in 500 bp windows. B) LD per chromosome. Minimum distance to R^2 of 0.20 (top) and to chromosome-specific background level of LD (bottom).

A Full coat 2_31919
   2_01960

B Small eye

C Large eye

D Holstein 2_15395

E Watson

1 cm

Chromosome
Vu01 Vu02 Vu03 Vu04 Vu05 Vu06 Vu07 Vu08 Vu09 Vu10 Vu11

F 2_01960
   Vigun09g139900
   Vu09 position (Mbp)

G 2_31919
   Vigun10g165400
   Vu10 position (Mbp)

H 2_15395
   Vigun10g163900
   Vu10 position (Mbp)

R^2
1.0
0.8
0.6
0.4
0.2
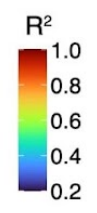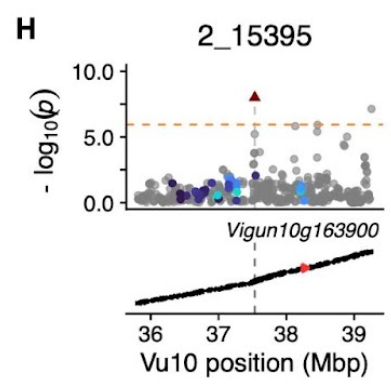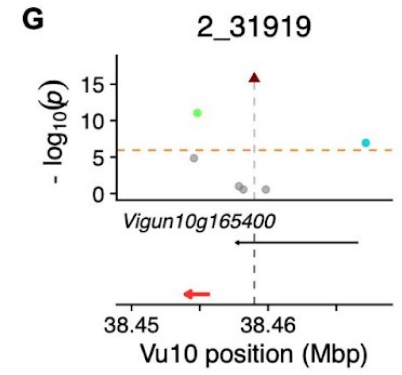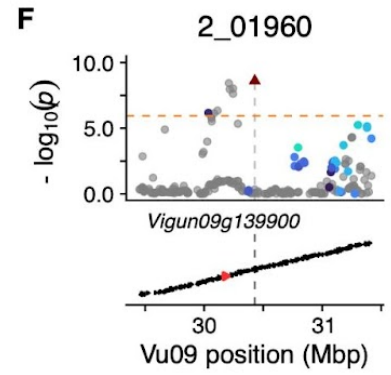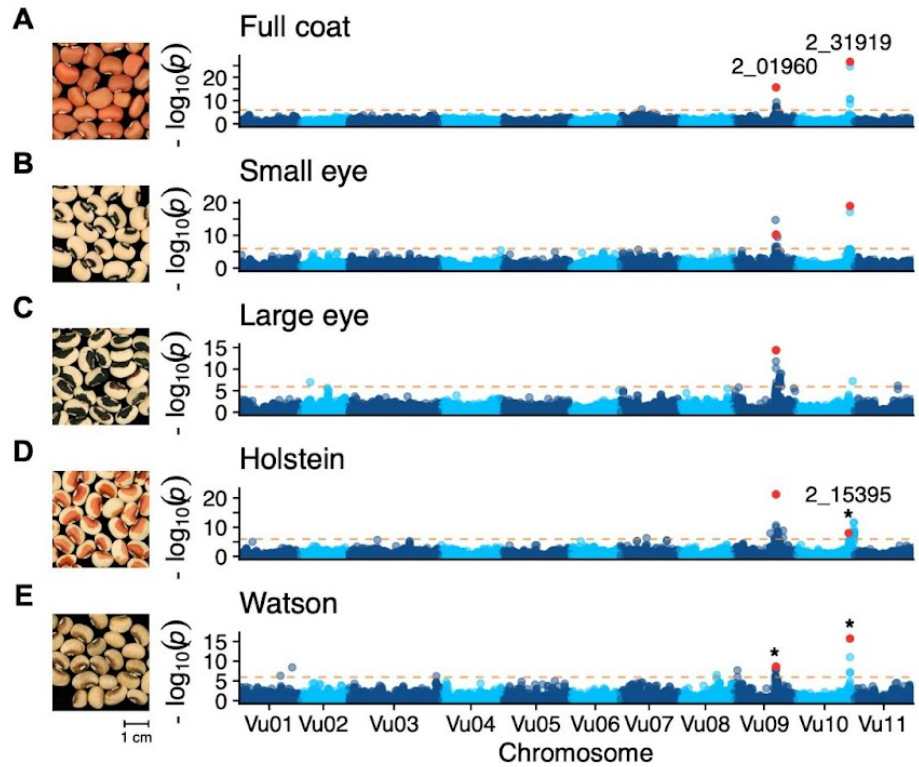
Figure 5. The genetic basis of seed coat pigmentation pattern

A-E) Manhattan plots from genome-wide association mapping of seed coat pigmentation pattern phenotypes. Orange line indicates Bonferroni corrected significance threshold. Tag SNPs are highlighted in red and labeled. Tag SNP 2_15395 is unique to Holstein, highlighted SNP on Vu10 in Watson and Full coat correspond to 2_31919. Photographs from UCR Cowpea Minicore as follows: Self colored (TVu-3842/MinicoreUCR_247), Small eye/Eye 1 (TVu-16278/MinicoreUCR_197), Large eye/Eye 2 (TVu-8775/MinicoreUCR_318), Holstein (TVu-7625/MinicoreUCR_290), Watson(TVu-2933/MinicoreUCR_232). F-H) Local linkage equilibrium for tag SNPs (triangle) for peaks indicated by * in A-E. Candidate genes are indicated in red and labeled.
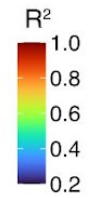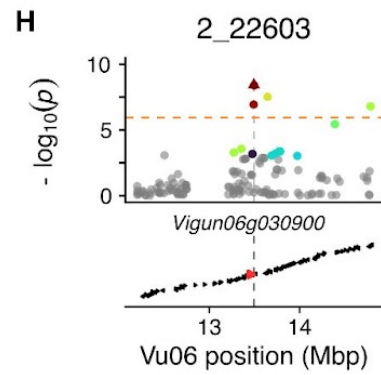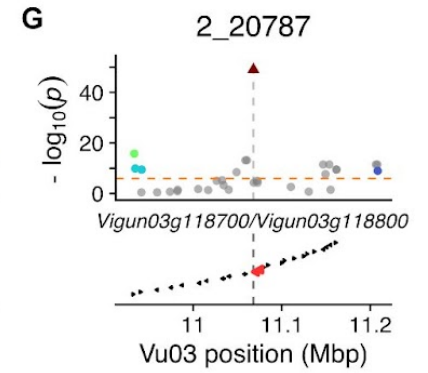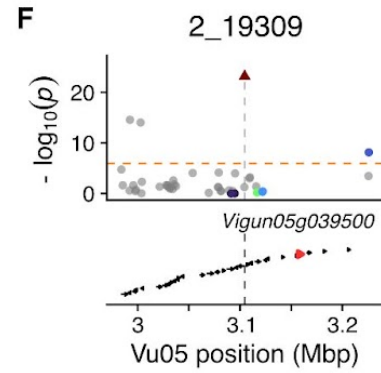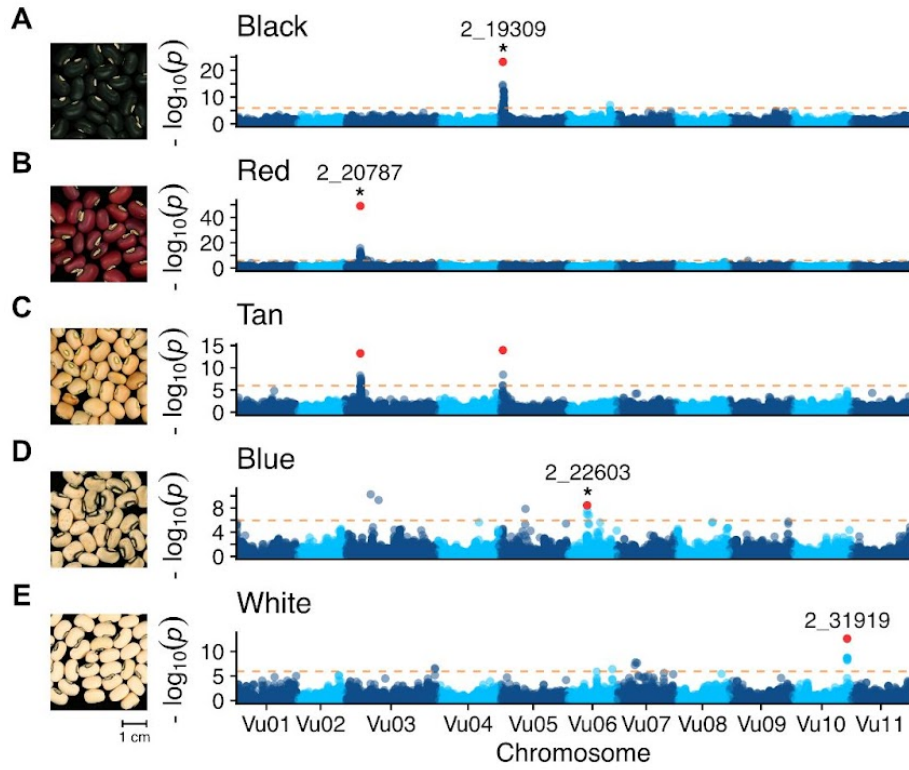
Figure 6. The genetic basis of seed coat color

A-E) Manhattan plots from genome-wide association mapping of seed color phenotypes. Orange line indicates Bonferroni corrected significance threshold. Tag SNPs are highlighted in red and labeled. Photographs from UCR Cowpea Minicore as follows: Black (TVu-13968/MinicoreUCR_139), Red (TVu-15426/MinicoreUCR_173), Tan (TVu-1036/MinicoreUCR_108), Blue (TVu-4669/MinicoreUCR_262),White (TVu-14401/MinicoreUCR_153). F-H) Local linkage equilibrium for tag SNPs (triangle) for peaks indicated by * in A-E. Candidate genes are indicated in red and labeled.

Phenotype	Genotype at patterning loci	Genotype at color loci
Full coat	$Pat9.1^G$, $Pat10.1^T$, $Pat10.2^c$	
Small eye (Eye 1)	$Pat9.1^A$, $Pat10.1^c$, $Pat10.2^T$ or $Pat10.2^c$	
Large eye (Eye 2)	$Pat9.1^A$, $Pat10.1^T$, $Pat10.2^c$	
Holstein	$Pat9.1^A$, $Pat10.1^T$, $Pat10.2^c$	
Watson	$Pat9.1^G$, $Pat10.1^c$, $Pat10.2^T$ or $Pat10.2^c$	
Black		$Col3.1^G$, $Col5.1^c$, $Col6.1^T$
Red		$Col3.1^T$, $Col5.1^c$ or $Col5.1^T$, $Col6.1^T$
Tan		$Col3.1^G$, $Col5.1^T$, $Col6.1^T$
Blue		$Col3.1^G$, $Col5.1^c$, $Col6.1^c$
White	$Pat10.1^c$	$Col3.1^G$, $Col5.1^c$ or $Col5.1^T$, $Col6.1^T$

Table 1. Genetic model of seed pigmentation pattern and color.
Proposed genotypes at pigmentation patterning and color quantitative trait loci. QTL correspond to SNPs as follows: *Pat9.1* = 2_01960, *Pat10.1* = 2_31919, *Pat10.2* = 2_15395, *Col3.1* = 2_20787, *Col5.1* = 2_19309, Col6.1 = 2_22603.

Supplemental Figures for

The pattern of genetic variability in a core collection of 2,021 cowpea accessions
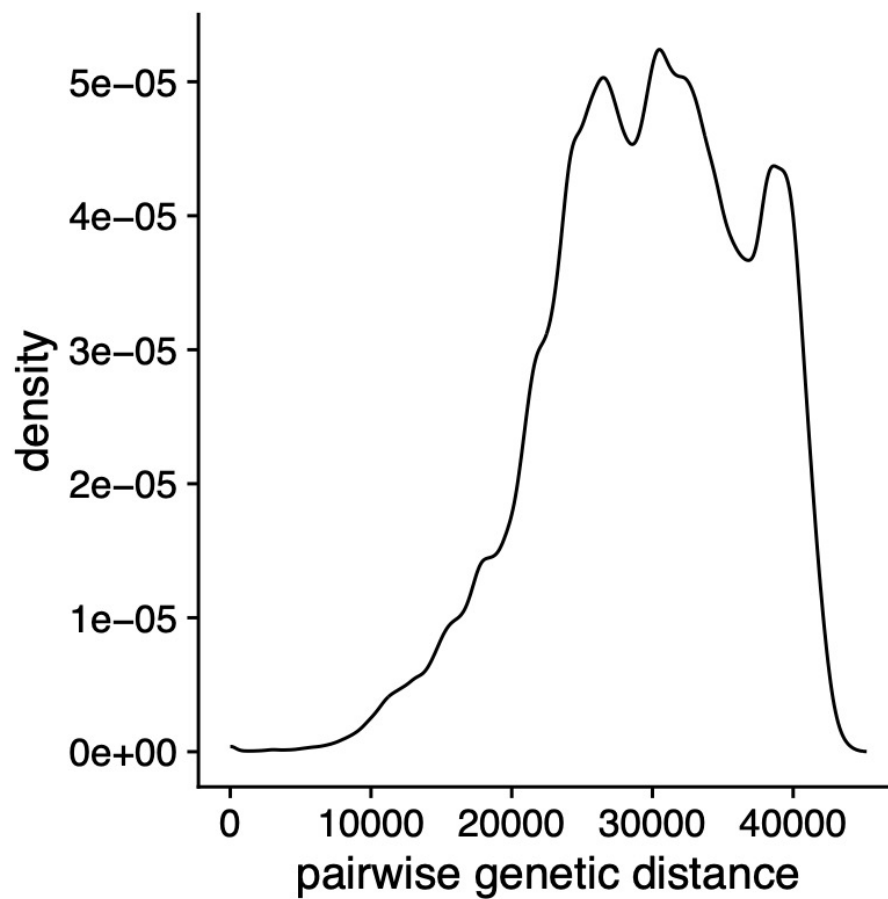
Figure S1. Distribution of pairwise genetic distances between accessions.
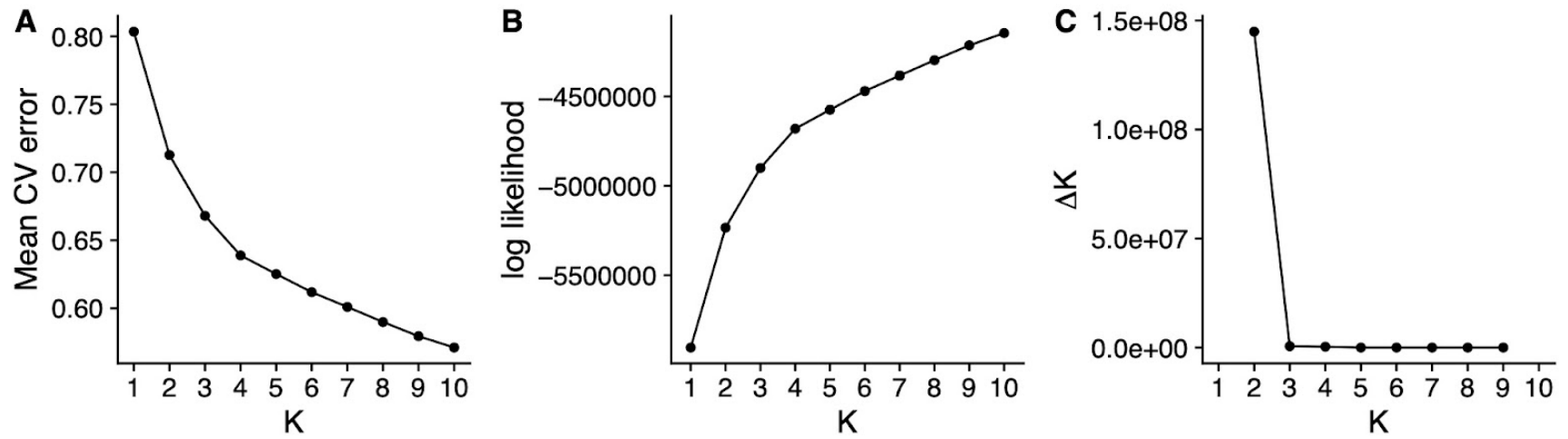
Figure S2. Admixture statistics per K.
A) Mean cross-validation error. B) Mean log-likelihood. C) Delta K.
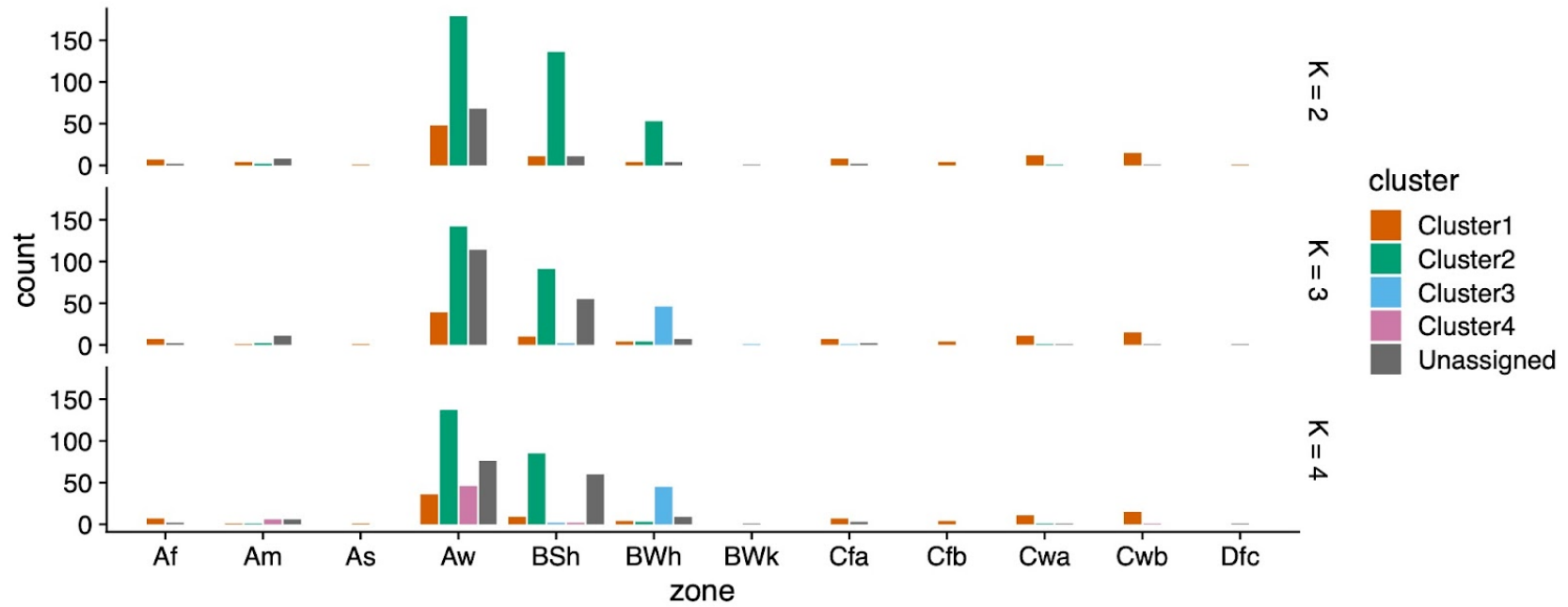
Figure S3. Climate zone distribution by subpopulation cluster for accessions with collection coordinates.
Climate designations correspond to the Köppen climate classification.
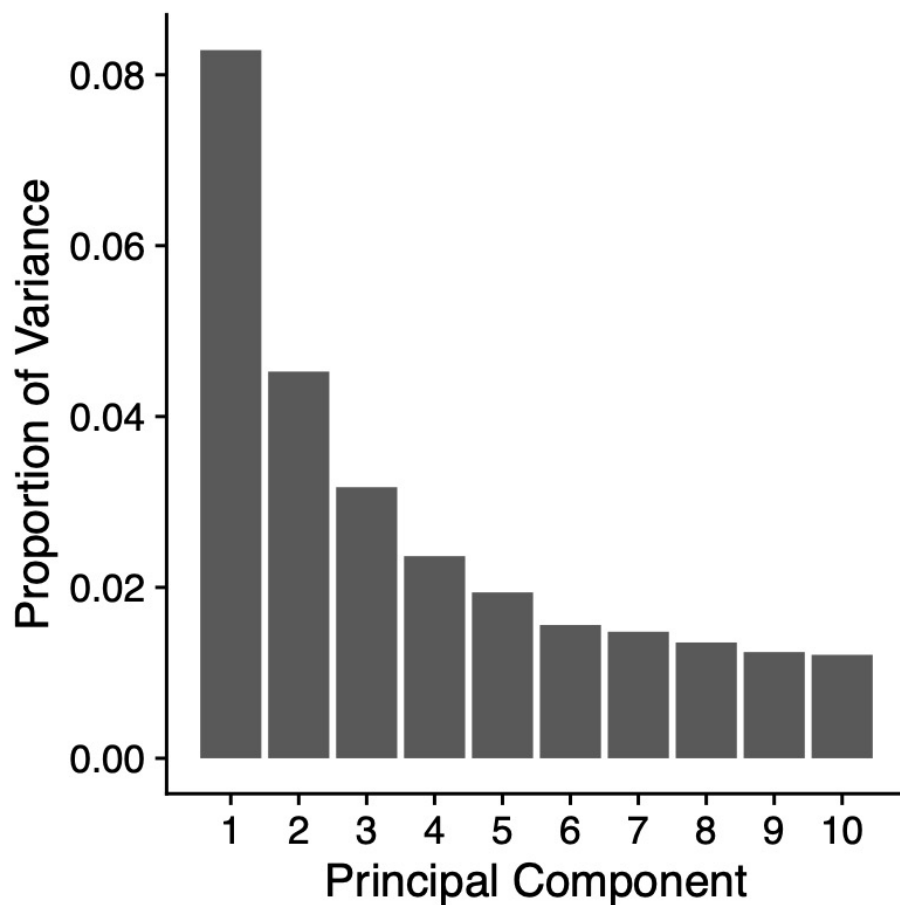
Figure S4. Proportion of variance explained by the first ten principal components.
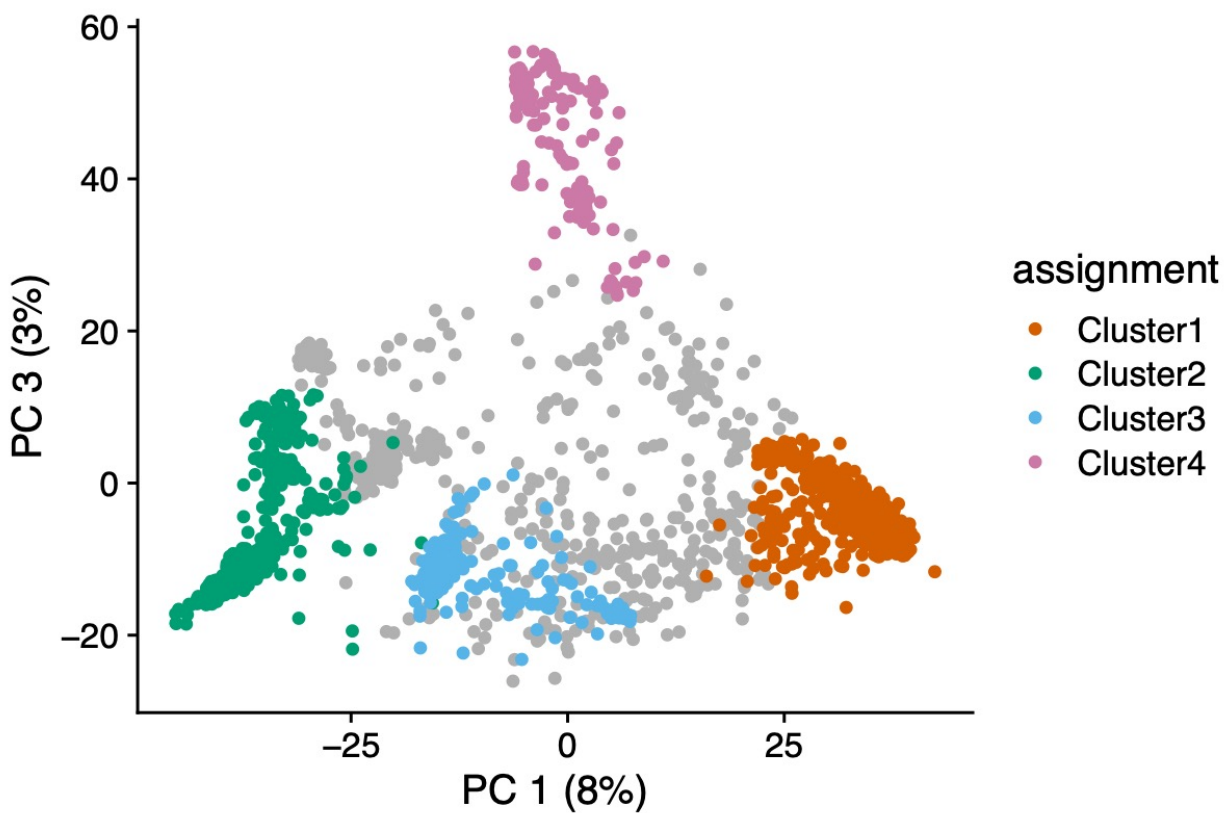
Figure S5. PC 1 vs. PC 3 (K = 4).
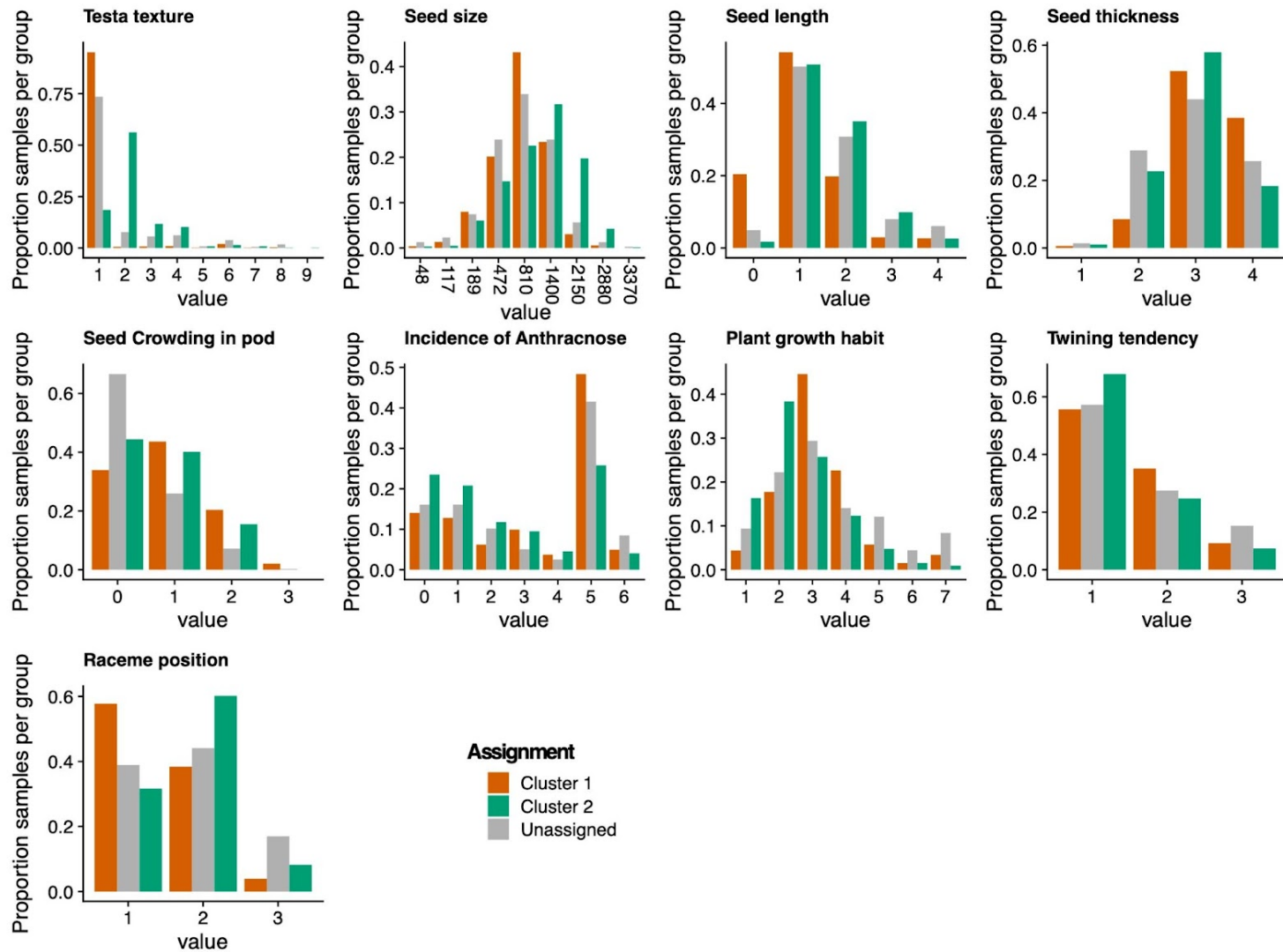
Figure S6. Distributions of discrete phenotypes that vary between Cluster 1 and Cluster 2 (K = 2). All phenotypes shown had a statistically significant difference between the two subpopulation clusters (Mann Whitney Wilcoxon adjusted p-value < 0.05)**.**
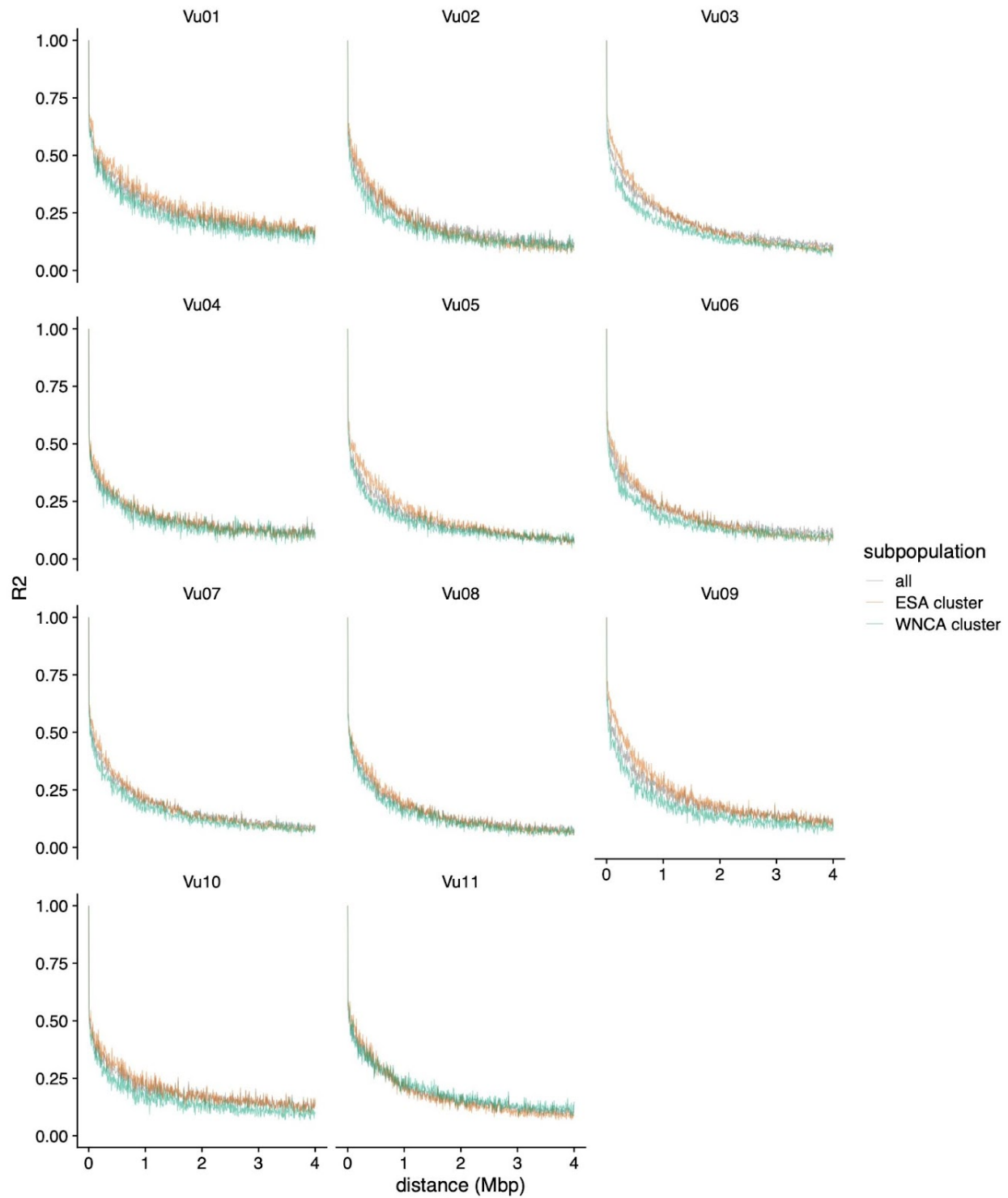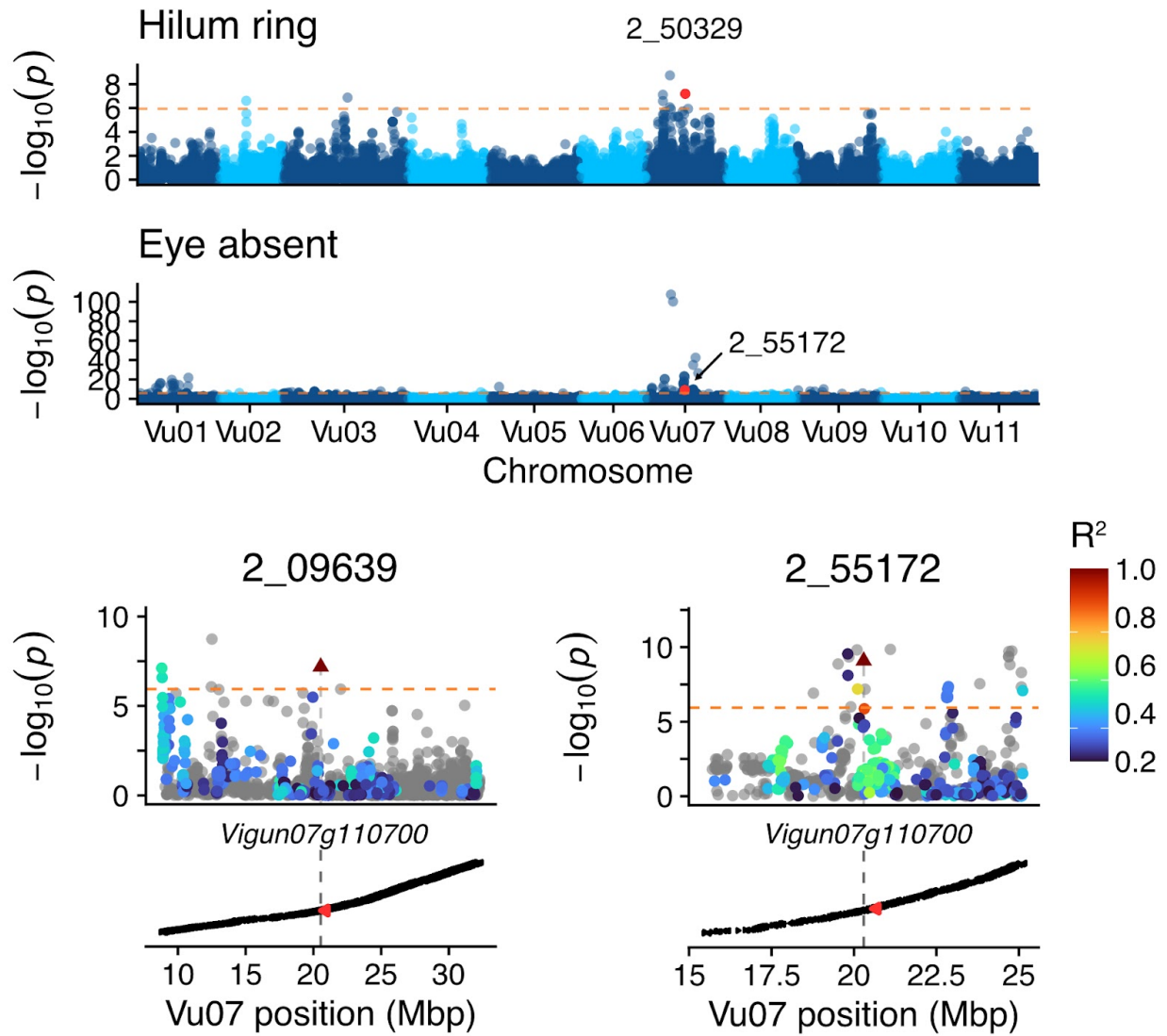
Figure S7. Linkage disequilibrium decay by chromosome.

Figure S8. Genome-wide association mapping manhattan plots and focal loci for additional eye pattern phenotypes.
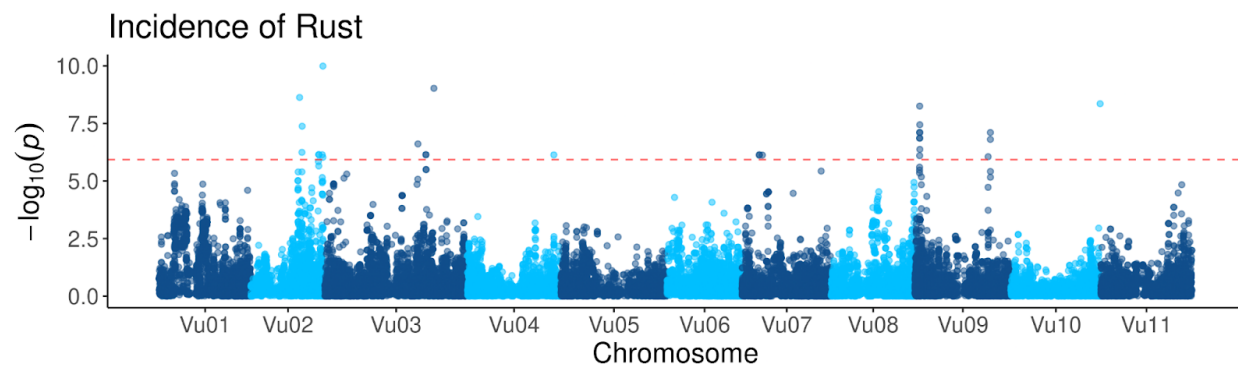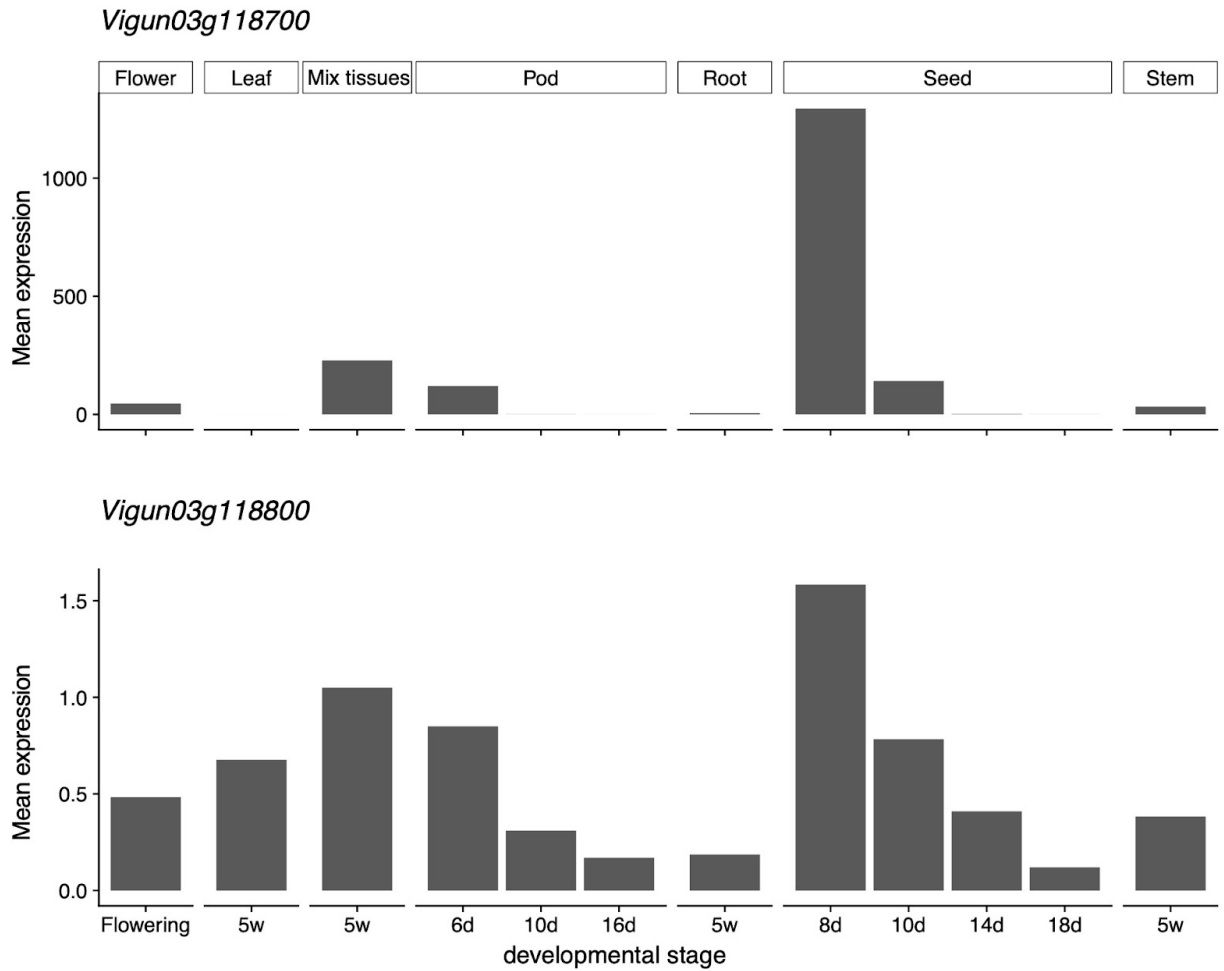
Figure S9. GWAS for rust resistance

Figure S10. Expression of 2 candidate genes for Tan and Red seed coat color across developmental stages and tissues.