bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Genetic diversity analysis and characterization of Ugandan sorghum**

2 Subhadra Chakrabarty[1], Raphael Mufumbo[1], Steffen Windpassinger[1], David Jordan[2], Emma
3 Mace[2], Rod J. Snowdon[1], Adrian Hathorn[2]

4

5 [1]Department of Plant Breeding, Justus Liebig University, Giessen, Germany

6 [2]Queensland Alliance for Agriculture and Food Innovation, University of Queensland,
7 Warwick, QLD, 4370, Australia

8

9 * Corresponding author

10

11 **Abstract**

12 The National Genebank of Uganda houses diverse and rich *Sorghum bicolor* germplasm
13 collection. This genetic diversity resource is untapped, under-utilized and has not been
14 systematically incorporated into sorghum breeding programs. In this study, we characterized
15 the germplasm collection using whole genome SNP markers. Discriminant analysis of
16 principal components (DAPC) was implemented to study racial ancestry of the accessions in
17 comparison to a global sorghum diversity set and characterize sub-groups and admixture in
18 the Ugandan germplasm. Genetic structure and phylogenetic analysis was conducted to
19 identify distinct genotypes in the Ugandan collection and relationships among groups.
20 Furthermore, in a case study for identification of potentially useful adaptive trait variation for
21 breeding, we performed genome-wide association studies for juvenile cold tolerance.
22 Genomic regions potentially involved in adaptation of Ugandan sorghum varieties to cooler
23 climatic conditions were identified that could be of interest for expansion of sorghum
24 production into temperate latitudes. The study demonstrates how genebank genomics can
25 potentially facilitate effective and efficient usage of valuable, untapped germplasm
26 collections for agronomic trait evaluation and subsequent allele mining.

27

28 **Keywords**: Sorghum bicolor; genetic diversity; population structure; cold tolerance;
29 temperate climate adaptation; genome wide association study; genebank;

30

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

31 **1. Introduction**

32 *Sorghum bicolor* [L.] Moench (sorghum) is the fifth most important cereal crop globally and
33 shows remarkable diversity, including five different races, their intermediates and several
34 crop forms classified as grain, forage, sweet and broomcorn types (Hariprasanna and Patil
35 2015). Sorghum has extraordinary untapped variation in grain type, plant type, adaptability,
36 productive capacity and underutilized genetic potential (Lost Crops of Africa 1996). Because
37 of its wide adaptability to drought and heat, sorghum's importance is expected to increase
38 with the changing global climate and an ongoing increase in the use of marginal lands for
39 agriculture (Paterson et al. 2009).

40 In Eastern Africa sorghum is traditionally grown as a food-fodder crop by small holder
41 farmers in low-input agricultural systems spanning highland, lowland and semi-arid cropping
42 regions. Uganda is one of only three countries in which all of the five basic races and ten
43 intermediate races of *S. bicolor* are endemic (Reddy et al. 2002). The country's broad
44 sorghum diversity reflects the variety of environments where the crop is grown, mainly on
45 marginal agricultural lands ranging from extremely arid and semi-arid zones in eastern and
46 northern Uganda to cool highlands in south-western regions. The Uganda National Genebank
47 houses a large collection of *S. bicolor* accessions, including a vast range of landraces whose
48 diversity has yet to be capitalized for use in breeding. This germplasm has not yet been fully
49 characterised and evaluated, limiting its utilisation to date in sorghum improvement programs
50 in Uganda and elsewhere. The considerable geographical and topological diversity of Uganda
51 makes this genetic resource a potentially interesting reservoir for genetic analysis and
52 diversity for adaptive traits of interest for sorghum breeding. For example, cool highland
53 areas in the southwestern Uganda are potential sources of diversity for cold-tolerance traits
54 that could help improve sorghum adaptation in temperate cropping regions.

55 Effective and efficient management of germplasm from a genebank collection is an essential
56 prerequisite for farmers and breeders to identify, extract and exploit the extensive diversity.
57 Genome-wide characterization of untapped genetic resources using genome sequencing
58 technologies provides new opportunities to sustainable breeding and efficient usage of
59 material.

60 In this present study, the diverse Uganda National Genebank *S. bicolor* collection,
61 representing different agro-ecological zones of the country, was investigated using whole
62 genome SNP markers and population genetic analysis. The primary objective was to

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

63 genetically characterize Ugandan sorghum germplasm in the context of global sorghum
64 diversity. As a case study for the value of this resource in adaptive breeding, we furthermore
65 used the available genome data, in association with phenotypic data for juvenile cold
66 tolerance traits, to identify genomic regions enriched with genetic variants associated to low
67 temperature adaptation. The results demonstrate how genebank genomics can help facilitate
68 discovery of economically or biologically important plant diversity and genes as a
69 prerequisite for crop genetic improvement and climatic adaptation.

70

71 **2. Materials and Methods**

72 **Plant materials**

73 A total of 3333 diverse Ugandan *S. bicolor* germplasm accessions (UG set) collected by the
74 Plant Genetic Resources Centre at the Uganda National Genebank were used in this study
75 (Table S1). This germplasm collection represents the entire sorghum diversity from all eco-
76 geographical regions of Uganda. For racial composition analysis, the collection was
77 compared to a global sorghum germplasm collection of 1033 genotypes (global set) which
78 was previously described by (Tao et al. 2020).

79 **DNA extraction and genotyping**

80 DNA was extracted from seeds by Diversity Arrays Technology Pty Ltd.
81 (www.diversityarrays.com) and genotyped using DArTseq, an efficient genotyping-by-
82 sequencing (GBS) platform which enables discovery of genome-wide markers through
83 genome complexity reduction using restriction enzymes. Subsequently, sorghum reference
84 genome version v3.1.1 (McCormick et al. 2018) was used for sequence alignment and single
85 nucleotide polymorphism (SNP) calling.

86 **SNP data filtration**

87 A total of 40,290 SNP markers were reported for the global and UG sets. Firstly all
88 nonspecific markers and those belonging to supercontigs were removed. The remaining
89 34,469 were used for further analysis. For racial ancestry analysis, an extremely high
90 stringency was then applied to remove markers which were duplicated or exhibited greater
91 than 1% missing data and genotypes showing greater than 1% missing data. The global
92 diversity set comprised conversion lines containing introgressed chromosome regions from
93 the Sorghum Conversion Program conducted by Texas Agricultural Experiment Station

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

94 (Thurber et al. 2013). Hence, to reduce disparity with UG samples in the co-analysis with the
95 global set, all markers from the specific genomic regions impacted by the conversion
96 program were excluded as follows: Sb06: all markers, Sb07: all markers beyond 40Mb, Sb09:
97 all markers beyond 46Mb. Markers with minor allele frequency (MAF) less than 0.01 were
98 also excluded. This set was then imputed using Beagle 5.1 (Browning, Zhou, and Browning
99 2018) to infer the remaining missing data values. A total of 2,331 common markers between
100 the UG and global set were used for the racial ancestry analysis.

101 **Population structure and genetic diversity study**

102 In order to understand the racial classification and the population structure of the UG
103 sorghum collection, principal component analysis (PCA) and Discriminant Analysis of
104 Principal Components (DAPC) were implemented using the R package Adegenet (2.1.3)
105 (Jombart, Devillard, and Balloux 2010). To avoid bias caused by the large size of the UG set,
106 we initially used only a representative subset of the UG set to assign racial groupings in
107 comparison to the global collection. Hence, prior to racial group identification of the UG set
108 in comparison to the global set, K-mean clustering was performed to select UG genotypes
109 representative for all identified groups. Ten clusters were identified using the 'find.clusters'
110 function. From each of these clusters, ten randomly-selected samples were combined with the
111 global dataset for DAPC co-analysis. To validate the racial assignment of the UG set groups,
112 the DAPC co-analysis was repeated three times, each time a different random selection of 10
113 genotypes from each identified group. The SNP data was converted to the genlight object bit-
114 level genotype coding scheme using the function 'vcfR2genlight' of the vcfR tool
115 (https://github.com/knausb/vcfR). After an initial transformation using the PCA analysis,
116 clusters were subsequently identified using discriminant analysis (DA).

117 To describe the population structure of the UG germplasm and evaluate the distribution of
118 racial groups in relation to geographical origins of accessions, the filtered marker set (34,466)
119 was pruned to exclude SNPs in strong LD using PLINK software (Purcell et al. 2007) SNPs
120 were pruned with a window of 50 SNPs, step size of 5 makers and r^2 threshold of 0.5. A total
121 of 12,742 markers for all UG lines were used to analyse population structure using
122 Discriminant analysis (DA). To elaborate the genetic relationship among the accessions a
123 pairwise distance matrix was established using the tool VCF2Dis (https://github.com/BGI-
124 shenzhen/VCF2Dis), which was then converted to a neighbour-joining phylogenetic tree
125 using the R package ape (5.5) (Paradis, Claude, and Strimmer 2004) and visualized with R
126 package ggtree (3.0.4) (Yu 2020). Weir and Cockerham's Fst (wcFst) (Weir and Cockerham

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

127 1984) was calculated among the UG subpopulations using the function '--weir-fst-pop' of
128 vcftools (0.1.17) to identify patterns of genetic differentiation.

129 **Phenotyping and association mapping**

130 For evaluating juvenile survival under cold stress, two field trials at Gross Gerau (GG),
131 Germany (spring 2019 and 2020) and one climate chamber experiment were conducted on a
132 UG subset of 444 (field trials) and 255 (climate chamber) accessions representing all agro-
133 ecological zones (Table S1) For the field experiments (Table S2) , all genotypes were sown in
134 micro-plots consisting of single rows (2.5 x 0.7 m) using an alpha lattice block design with
135 two replications. Being cold stress experiments, sowing times were several weeks earlier than
136 normal for sorghum in that area. Even though the mean soil temperatures were relatively high
137 during the course of the experiments (13.8 and 16.2 °C, respectively) and allowed for a
138 satisfying emergence, in both years several cold nights (up to -1.5 °C) implied strong stress
139 on the seedlings. Around four weeks after emergence, when the last cold event lay 7 days
140 back, the number of surviving plants was scored per plot. For subsequent GWAS, the alpha-
141 lattice adjusted mean value of both years was used. For the climate chamber experiment
142 (CC), 16 seedlings per genotype were established in 12 x 12 x 12 cm pots. Experiments were
143 designed as randomized complete block design with four replications (Table S3) and the
144 number of surviving plants per pots was scored as a measure for juvenile cold stress
145 tolerance.

146 After eliminating markers and genotypes with more than 25% missing data points, the
147 dataset was imputed with using Beagle 5.1 (Browning, Zhou, and Browning 2018). It was
148 further corrected for MAF lower than 5%. A total of 4,099 markers were used for genome
149 wise association study (GWAS) implemented in R package GenABEL (1.8-0) (Aulchenko,
150 de Koning, and Haley 2007) for juvenile cold tolerance traits.

151 Population stratification was adjusted by incorporating two frequently used models: principal
152 coordinate and genomic kinship matrix corrections principle component and kinship analysis
153 (Stich et al. 2008) . The correlations between each marker and trait can be well identified by
154 these models after variation due to population structure has been controlled.

155 To reduce the type II error rate and in order to classify a marker–trait association as
156 significant, a threshold of $-\log10$ (p value) ≥ 3.0 was defined (Gabur et al. 2019). Linkage
157 disequilibrium across the entire genome was calculated using the squared allele frequency

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

158 correlations (r^2) between each pair of SNPs. Haplotype blocks were calculated using an LD
159 threshold of $r^2 > 0.7$ to define blocks, as described by (Gabriel et al. 2002).

160 Candidate genes were selected based on *Sorghum bicolor* reference genome *v3.1.1* hosted by
161 Phytozome 12 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Sbicolor).
162 Orthologous genes within selected haploblocks were identified by homology comparisons of
163 the genomic sequences in maize and rice. Protein sequence alignments were conducted by
164 using blastp option of DIAMOND (Buchfink, Xie, and Huson 2015).

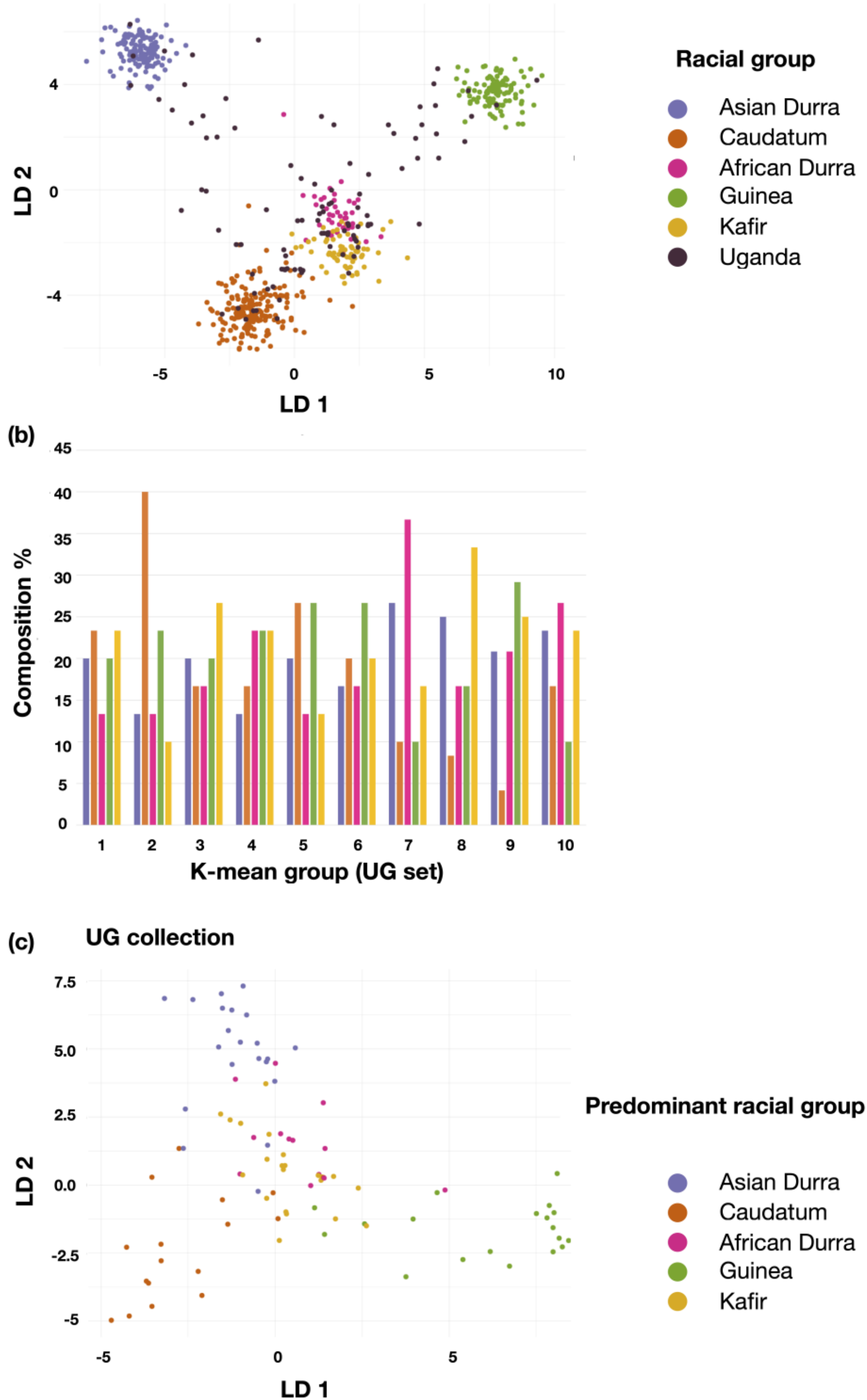
165

166 **3. Results**

167 **Genetic diversity and population structure analysis**

168 Appropriate conservation and effective utilization of novel UG germplasm can be facilitated
169 by understanding the underlying variation and genetic diversity. DAPC between the global
170 and the UG set revealed the presence of all racial groups for the latter (Figure 1). The
171 majority of the accessions could not be directly classified to any particular racial cluster, but
172 rather showed intermediate classifications between two racial groups, indicating the presence
173 of admixture. This was further confirmed by Kmean grouping (Figure 1b). All racial groups
174 were represented in each of the ten identified Kmean clusters, however cluster two showed a
175 strong overrepresentation of caudatum race.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



176

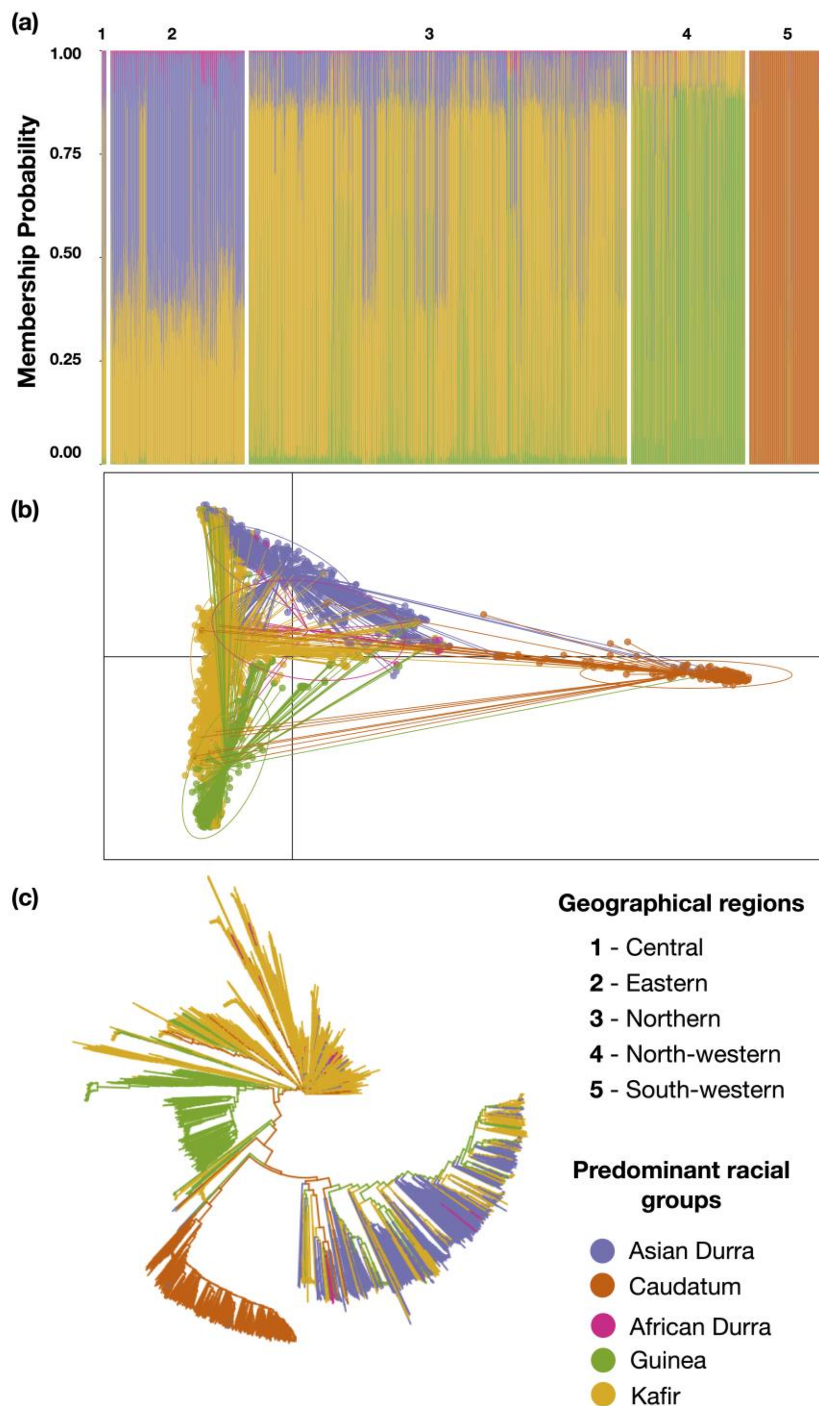
177 **Figure 1**: Scatterplot and composition of discriminant coefficients of the DAPC analysis for
178 racial group determination of UG samples based on global diversity set. (a) DA loadings
179 (LD) displaying UG clustering in comparison to global racial groups. (b) Imputed racial
180 composition of ten identified Ugandan Kmean clusters. The x axis represents the clusters
181 whereas the y axis indicates the racial composition in percentage for each cluster. (c) DA
182 loading of Ugandan representative accessions based on races. Each dot represents an
183 individual and the colour code is displayed in the index.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

184 Results from Kmean clustering and DAPC using the three random samples from UG groups
185 co-analysed with the global reference collection confirmed the racial distribution and
186 presence of admixture across the UG diversity set (Figure 2, Figure S1, Table S4).

187 To better understand the relationships among the Ugandan accessions based on their imputed
188 racial classification and the geographical origin, a DAPC analysis was performed only for the
189 UG set (Figure 2). Genotypes from the central, eastern, northern and north-western
190 geographical regions were found to be predominantly admixtures, whereas the southwestern
191 region (dominated by genotypes of highland origin) showed a distinct genetic pattern with a
192 predominance of caudatum race (Figure 2a,b). Four hundred PCs (27.4% of variance
193 conserved by first three PCs) and four discriminant eigenvalues were retained in the DAPC
194 analysis in order to capture maximum variance. Overall, diversity in the UG collection is
195 defined more by racial grouping rather than geographical grouping (Figure 2a,b) A
196 phylogenetic tree confirmed the divergence of south-western accessions (Figure 2c).

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



197

198 **Figure 2**: DAPC analysis of the Ugandan genebank sorghum collection. (a) Barplot, (b)
199 Scatter plot and (c) phylogenetic tree analysis of subpopulations. Colours represent
200 predominant racial group assignment of each accession. The first two linear discriminants
201 (LDs) are represented by the axes of the scatterplot. In (a) each vertical bar represents one
202 individual accession, coloured based on the identified racial classification. Circles in (b)
203 represent identified clusters corresponding to racial groups.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

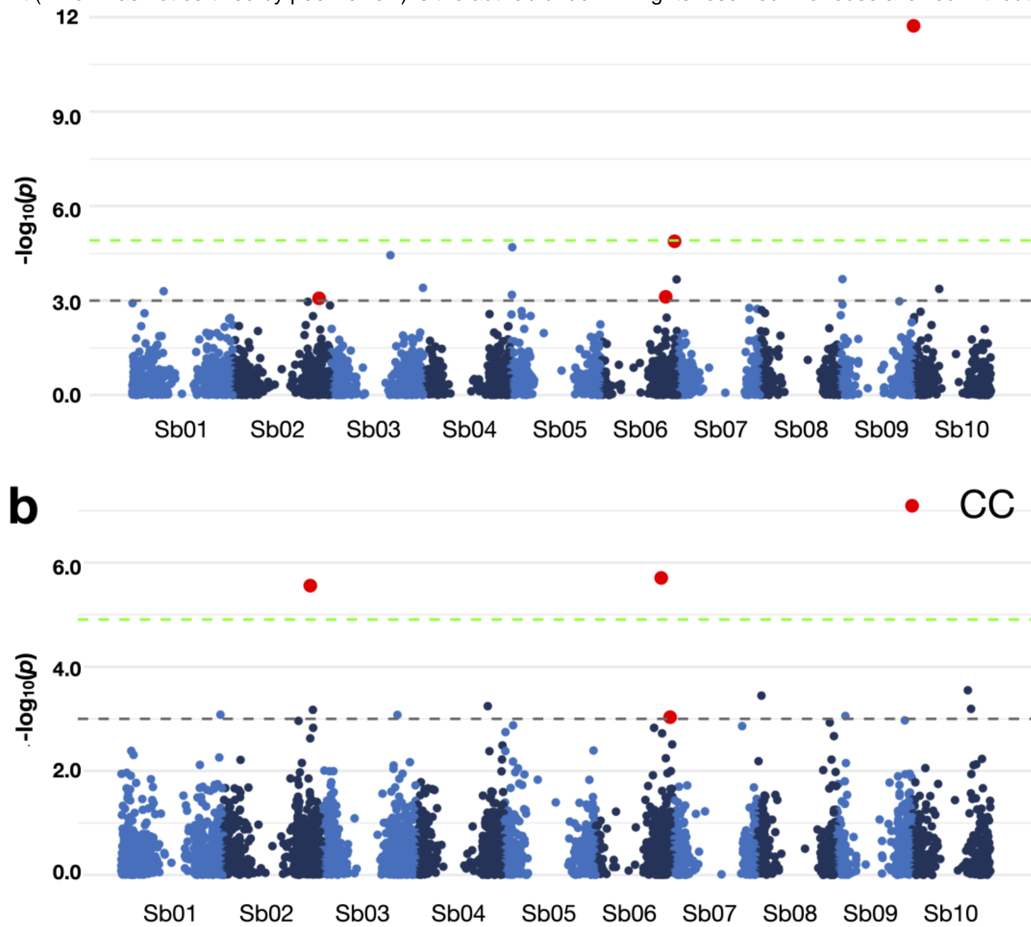
204 Due to significant adaptability differences between the lowland and highland races from the
205 Ugandan diversity set, genetic diversity among the sub groups within the population were
206 tested. Fst revealed overall genetic variance of 0.30 (ranging from -0.002 to 0.95), indicating
207 presence of significant genetic structuring or population subdivision within the germplasm.

208 **Phenotypic variation for juvenile cold stress survival, association mapping and**
209 **haplotype analysis**

210 A genome-wide association study (GWAS) was conducted for identification of regions of
211 interest associated to juvenile survival under cold stress in two temperate-climate field
212 environments and one controlled-environment climate chamber test. Highly significant
213 differences for cold tolerance among genotypes (p=0.000***, Table S5) were found in all
214 experiments. Comparing lowland- and highland genotypes as groups by a one-way ANOVA,
215 highland genotypes showed a superior cold tolerance in all experiments as expected
216 (p=0.015* for GG19, p=0.000*** for GG20, p=0.000*** for CC).

217 Significant marker-trait associations to juvenile survival under cold stress, consistent across
218 all test environments, were identified on chromosomes Sb02, Sb06, and Sb09 (Figure 3 a, b).
219 We compared these selected genomic regions with previously curated QTL in sorghum QTL-
220 Atlas (Mace et al. 2019), based on physical position (v3.0) and filtered using category
221 resistance abiotic and subcategory cold tolerance. A sum of nineteen overlapping QTL
222 involved in juvenile cold tolerance were identified (Table S6) in these regions. This result
223 affirms the important role of these genomic regions towards cold adaptation.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



224 **Fig 3**: Association mapping of survival under cold stress under (a) field condition (Gross
225 Gerau; GG) and (b) climate chamber (CC). The grey dotted horizontal line indicates a
226 threshold of genome-wide cut-off at -log10(p) > 3.0 while the green line indicates the
227 Bonferroni threshold at -log10(p) > 4.9. The selected associated markers which overlapped
228 under both conditions are marked in red.

229 Comparing orthologous gene segments of sorghum, maize and rice, six genes involved in
230 various cold acclimatization and tolerance were identified (Table S7). For example, a F-box
231 protein-encoding gene, *Sobic.006G245200* was identified at a distance of 9kb from the
232 associated marker Sb06-58472351. This gene family has been shown to be upregulated under
233 cold stress in rice, *Brassica rapa* and pepper (Jain et al. 2007; Rameneni et al. 2018;
234 Venkatesh et al. 2020).

235

236 **Discussion**

237 Plant genetic resources like the sorghum collection of the National Genebank of Uganda are
238 extremely important public germplasm resources for local breeders in crop centres of origin
239 and the agricultural and crop research community worldwide. The substantial variation we
240 identified in the Ugandan *S. bicolor* germplasm reflects the highly diverse environments
241 where sorghum grows in Uganda. Accessions with adaptation to drought and heat stress from

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

242 the arid regions of Northern and Eastern Uganda, and germplasm from the cold highlands of
243 the South-Western Kigezi region, carry potentially useful genetic diversity for heat and cold
244 stress adaptation traits. The latter are of major interest for breeding programs aiming to
245 expand sorghum production into temperate climates in North America, Asia and Europe,
246 whereas heat and drought stress tolerance are becoming increasingly important for global
247 sorghum production in the face of climate change. The rich genetic resource of the Ugandan
248 genebank sorghum collection has not yet been fully characterised and evaluated, limiting its
249 utilisation to date in sorghum improvement programs in Uganda and elsewhere.

250 Genetic structure has been previously documented in sorghum in smaller germplasm
251 collections (Westengen et al. 2014; Mofokeng et al. 2014; Cuevas and Prom 2020). This
252 study reports genetic diversity in a substantial population of 3333 accessions from a key
253 sorghum centre of origin, comparing it to previously described sorghum diversity by co-
254 analysis with a global diversity set (Tao et al. 2020).

255 The DAPC method is a good alternative to other population structure analysis software such
256 as STRUCTURE because of its ability to deal with large datasets. Clustering of genotypes
257 presented in this study provide interesting leads for increasing diversity in breeding programs
258 and germplasm utilization. Overall, results from DAPC, Kmean and phylogenetic tree
259 analysis were in agreement and provided evidence that the global racial diversity was well-
260 covered in the UG germplasm. However, considerable admixture between racial groups was
261 identified, particularly between genotypes from lowland areas in Central, Eastern and
262 Northern Uganda, indicating frequent gene flow between these regions. In contrast, the south-
263 western region appeared genetically distinct and was comprised predominantly of Caudatum
264 germplasm. Because this region includes cool-temperature highland areas it may contain
265 interesting adaptive diversity for juvenile or reproductive cold tolerance. Population structure
266 and phylogenetic analysis revealed the distinctness of this group compared to the other eco-
267 geographical regions, presumably due to the distinctness of the highland cultivation
268 environment and a relatively low exchange of germplasm between highly and lowland
269 farmers. Similarly, (Mekbib 2008) reported the existence of caudatum and its intermediate
270 races in the highlands of Ethiopia. This confers with the theory that the spread and
271 diversification of crops to different locations can lead to new variants, a process influenced
272 by genotype by environment interactions and geographical isolation (Pressoir and Berthaud
273 2004). Based on the principles of artificial selection (Yamasaki et al. 2007), farmers select
274 cultivars with preferred traits and thus increase their proportion compared to other cultivars.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

275 According to (Akatwijuka et al. 2016), sorghum from south-western Uganda tend to have
276 semi-compact elliptic panicles, a well-known characteristic of caudatum and its intermediate
277 races.

278 Sorghum has a high potential for adaptation to a wide range of environmental conditions.
279 Besides yield and other agronomic traits, the improvement of cold tolerance at juvenile and
280 reproductive stages (Schaffasz2019(a,b); S. Chakrabarty et al. 2021) is a major breeding
281 objective for sorghum temperate cropping regions. Early seedling vigour is critical for crop
282 establishment in any environment (Xie et al. 2014) and vigorous germination and growth
283 under low temperatures is essential for early establishment and weed competition in
284 temperate climates. Improving cold tolerance in the early juvenile stage allows higher yield
285 potential and better maturity. Sharifi (2008) observed that survival percentage, germination,
286 and chlorophyll content were good indicators of early cold tolerance in rice. The yield of
287 sorghum is highly temperature dependent, especially between sowing and flowering time
288 (Craufurd et al. 1999; Fiedler et al. 2014). Hence, breeding for juvenile cold tolerance is of
289 utmost importance, especially for temperate European climates. In this study, juvenile
290 survival under low temperature was studied for multi environments. In contrast to emergence
291 and juvenile biomass under cold conditions which have been extensively studied in several
292 publications (e. g. Burow et al. 2011, Fiedler et al. 2014, Schaffasz et al. 2019), the trait
293 juvenile survival has received much less attention so far. Though, it is of utmost importance,
294 because a satisfying emergence is worthless if the seedlings later succumb to cold stress.
295 Promising hotspots for cold tolerance during juvenile development were identified. Since
296 population stratification was accounted for before performing GWAS, we have reduced the
297 likelihood that of genetic background effects are generating spurious associations. We also
298 learned that multiple QTL identified in previous studies (Mace et al. 2019) and genes known
299 to be involved in cold stress endurance were physically co-located with our QTL, suggesting
300 the importance of our selected genomic regions in this regard.

301 The results from this study therefore pinpoint interesting plant materials and genome regions
302 containing alleles which can be mined for improvement of cold temperature adaptation.
303 Given the complex genetics underlying juvenile cold tolerance (Burow et al. 2011; Bekele et
304 al. 2014; Fiedler et al. 2014) the implementation of genomic prediction appears best suited to
305 improve selection. The data generated in this study represent a promising first step towards
306 the implementation of genomic prediction to enhance cold adaptation traits in sorghum
307 breeding.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

308

309 **Conclusions**

310 To our knowledge, this is the first extensive study of the unique and large sorghum
311 germplasm collection conserved at the National Genebank of Uganda. The results indicate
312 immense genetic and racial diversity of predominantly admixed accessions, and a unique,
313 relatively genetically isolated resource of caudatum germplasm from the south-western
314 Ugandan highlands. Despite its highly complex nature, juvenile chilling stress is critical for
315 temperate climate adaptation. The discovery of important genomic regions for these traits and
316 high phenotypic variability associated with these regions can potentially enable breeders to
317 enhance early stage chilling tolerance in sorghum. The genomic and phenotypic data
318 collected in this study provide an objective criterion for the selection of accessions for genetic
319 diversity preservation and management, utilization in breeding programs and genetic
320 relationship analysis with other germplasm collections. The results provide important new
321 insight for adaptive crops breeding to support the expansion and stability of sorghum
322 production in the face of increasing abiotic stress constraints and climatic change.

323 **Acknowledgement**

324 We want to thank Mario Tolksdorf/JLU Experimental Station Gross-Gerau, Germany,
325 Annette Plank and Birgit Keiner (JLU Giessen) for the help in greenhouse.

326 **Conflict of interest**

327 The authors declare no conflicts of interest.

328 **Author contributions**

329 SC generated the data, conducted the data analysis wrote the manuscript. RM curated the
330 material and assisted in data collection. SW planned and oversaw field trials and data
331 collection and assisted in data analysis. DJ and EM provided material and assisted in data
332 analysis. RJS conceived the study and edited the manuscript. AH provided ideas and assisted
333 in data analysis.

334

335 **References**

336 Akatwijuka, R, P R Rubaihayo, and T L Odong. 2016. "Genetic Diversity among Sorghum
337 Landraces of Southwestern Highlands of Uganda." *African Crop Science Journal* 24 (2):

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

338 179–90.

339 Aulchenko, Yurii S, Dirk-Jan de Koning, and Chris Haley. 2007. "Genomewide Rapid
340 Association Using Mixed Model and Regression: A Fast and Simple Method For
341 Genomewide Pedigree-Based Quantitative Trait Loci Association Analysis." *Genetics*
342 177 (1): 577–85. https://doi.org/10.1534/genetics.107.075614.

343 Bekele, Wubishet A, Karin Fiedler, Amukelani Shiringani, Daniel Schnaubelt, Steffen
344 Windpassinger, Ralf Uptmoor, Wolfgang Friedt, and Rod J Snowdon. 2014.
345 "Unravelling the Genetic Complexity of Sorghum Seedling Development under Low-
346 temperature Conditions." *Plant, Cell & Environment* 37 (3): 707–23.

347 Browning, Brian L, Ying Zhou, and Sharon R Browning. 2018. "A One-Penny Imputed
348 Genome from Next-Generation Reference Panels." *The American Journal of Human*
349 *Genetics* 103 (3): 338–48. https://doi.org/10.1016/j.ajhg.2018.07.015.

350 Buchfink, Benjamin, Chao Xie, and Daniel H Huson. 2015. "Fast and Sensitive Protein
351 Alignment Using DIAMOND." *Nature Methods* 12 (1): 59–60.

352 Burow, Gloria, John J Burke, Zhanguo Xin, and Cleve D Franks. 2011. "Genetic Dissection
353 of Early-Season Cold Tolerance in Sorghum (Sorghum Bicolor (L.) Moench)."
354 *Molecular Breeding* 28 (3): 391–402. https://doi.org/10.1007/s11032-010-9491-4.

355 Chakrabarty, S., N. Kravcov, A. Schaffasz, R.J. Snowdon, B. Wittkop, and S. Windpassinger.
356 2021. "Genetic Architecture of Novel Sources for Reproductive Cold Tolerance in
357 Sorghum." *Frontiers in Plant Science* 12. https://doi.org/10.3389/fpls.2021.772177.

358 Chakrabarty, Subhadra, Natalja Kravcov, André Schaffasz, Rod J. Snowdon, Benjamin
359 Wittkop, and Steffen Windpassinger. 2021. "Genetic Architecture of Novel Sources for
360 Reproductive Cold Tolerance in Sorghum." *Frontiers in Plant Science* 0 (November):
361 2574. https://doi.org/10.3389/FPLS.2021.772177.

362 Craufurd, P Q, V Mahalakshmi, F R Bidinger, S Z Mukuru, J Chantereau, P A Omanga, A
363 Qi, et al. 1999. "Adaptation of Sorghum: Characterisation of Genotypic Flowering
364 Responses to Temperature and Photoperiod." *Theoretical and Applied Genetics* 99 (5):
365 900–911. https://doi.org/10.1007/s001220051311.

366 Cuevas, Hugo E, and Louis K Prom. 2020. "Evaluation of Genetic Diversity, Agronomic
367 Traits, and Anthracnose Resistance in the NPGS Sudan Sorghum Core Collection."
368 *BMC Genomics* 21 (1): 88. https://doi.org/10.1186/s12864-020-6489-0.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

369 Fiedler, Karin, Wubishet A Bekele, Ria Duensing, Susann Gründig, Rod Snowdon, Hartmut
370 Stützel, Arndt Zacharias, and Ralf Uptmoor. 2014. "Genetic Dissection of Temperature-
371 Dependent Sorghum Growth during Juvenile Development." *Theoretical and Applied*
372 *Genetics* 127 (9): 1935–48. https://doi.org/10.1007/s00122-014-2350-7.

373 Gabriel, Stacey B, Stephen F Schaffner, Huy Nguyen, Jamie M Moore, Jessica Roy, Brendan
374 Blumenstiel, John Higgins, Matthew DeFelice, Amy Lochner, and Maura Faggart. 2002.
375 "The Structure of Haplotype Blocks in the Human Genome." *Science* 296 (5576): 2225–
376 29.

377 Gabur, Iulian, Harmeet Singh Chawla, Rod J Snowdon, and Isobel A P Parkin. 2019.
378 "Connecting Genome Structural Variation with Complex Traits in Crop Plants."
379 *Theoretical and Applied Genetics* 132 (3): 733–50. https://doi.org/10.1007/s00122-018-
380 3233-0.

381 Hariprasanna, K, and J V Patil. 2015. "Sorghum: Origin, Classification, Biology and
382 Improvement BT - Sorghum Molecular Breeding." In , edited by R Madhusudhana, P
383 Rajendrakumar, and J V Patil, 3–20. New Delhi: Springer India.
384 https://doi.org/10.1007/978-81-322-2422-8_1.

385 Jain, Mukesh, Aashima Nijhawan, Rita Arora, Pinky Agarwal, Swatismita Ray, Pooja
386 Sharma, Sanjay Kapoor, Akhilesh K Tyagi, and Jitendra P Khurana. 2007. "F-Box
387 Proteins in Rice. Genome-Wide Analysis, Classification, Temporal and Spatial Gene
388 Expression during Panicle and Seed Development, and Regulation by Light and Abiotic
389 Stress." *Plant Physiology* 143 (4): 1467–83. https://doi.org/10.1104/pp.106.091900.

390 Jombart, Thibaut, Sébastien Devillard, and François Balloux. 2010. "Discriminant Analysis
391 of Principal Components: A New Method for the Analysis of Genetically Structured
392 Populations." *BMC Genetics* 11 (1): 94. https://doi.org/10.1186/1471-2156-11-94.

393 "Lost Crops of Africa." 1996. In , 127–44. Washington, D.C: National Academies Press.
394 https://www.nap.edu/read/2305/chapter/10#131.

395 Mace, Emma, David Innes, Colleen Hunt, Xuemin Wang, Yongfu Tao, Jared Baxter, Michael
396 Hassall, Adrian Hathorn, and David Jordan. 2019. "The Sorghum QTL Atlas: A
397 Powerful Tool for Trait Dissection, Comparative Genomics and Crop Improvement."
398 *Theoretical and Applied Genetics* 132 (3): 751–66.

399 McCormick, Ryan F, Sandra K Truong, Avinash Sreedasyam, Jerry Jenkins, Shengqiang

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

400 Shu, David Sims, Megan Kennedy, et al. 2018. "The Sorghum Bicolor Reference
401 Genome: Improved Assembly, Gene Annotations, a Transcriptome Atlas, and Signatures
402 of Genome Organization." *The Plant Journal* 93 (2): 338–54.
403 https://doi.org/https://doi.org/10.1111/tpj.13781.

404 Mekbib, Firew. 2008. "Farmers' Breeding of Sorghum in the Center of Diversity, Ethiopia: I.
405 Socioecotype Differentiation, Varietal Mixture and Selection Efficiency." *Journal of*
406 *New Seeds* 9 (1): 43–67. https://doi.org/10.1080/15228860701879299.

407 Mofokeng, Alina, Hussein Shimelis, Pangirayi Tongoona, and Mark Laing. 2014. "A Genetic
408 Diversity Analysis of South African Sorghum Genotypes Using SSR Markers." *South*
409 *African Journal of Plant and Soil* 31 (3): 145–52.
410 https://doi.org/10.1080/02571862.2014.923051.

411 Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. "APE: Analyses of
412 Phylogenetics and Evolution in R Language." *Bioinformatics* 20 (2): 289–90.

413 Paterson, Andrew H, John E Bowers, Rémy Bruggmann, Inna Dubchak, Jane Grimwood,
414 Heidrun Gundlach, Georg Haberer, et al. 2009. "The Sorghum Bicolor Genome and the
415 Diversification of Grasses." *Nature* 457 (7229): 551–56.
416 https://doi.org/10.1038/nature07723.

417 Pressoir, G, and J Berthaud. 2004. "Population Structure and Strong Divergent Selection
418 Shape Phenotypic Diversification in Maize Landraces." *Heredity* 92 (2): 95–101.
419 https://doi.org/10.1038/sj.hdy.6800388.

420 Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A R Ferreira,
421 David Bender, Julian Maller, Pamela Sklar, Paul I W De Bakker, and Mark J Daly.
422 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based
423 Linkage Analyses." *The American Journal of Human Genetics* 81 (3): 559–75.

424 Rameneni, Jana Jeevan, Vignesh Dhandapani, Parameswari Paul, Sangeeth Prasath Devaraj,
425 Su Ryun Choi, So Young Yi, Seongmin Hong, Sang Heon Oh, Man-Ho Oh, and Yong
426 Pyo Lim. 2018. "F-Box Genes in Brassica Rapa: Genome-Wide Identification,
427 Structural Characterization, Expressional Validation, and Comparative Analysis." *Plant*
428 *Molecular Biology Reporter* 36 (3): 500–517. https://doi.org/10.1007/s11105-018-1083-
429 1.

430 Reddy, V G; Rao, N K; Reddy, B V S;, and K E P Rao. 2002. "Sorghum Research Reports

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

431 Geographic Distribution of Basic and Inter- Sorghum Germplasm," 15–17.

432 Schaffasz, André, Steffen Windpassinger, Wolfgang Friedt, Rod Snowdon, and Benjamin
433 Wittkop. 2019. "Sorghum as a Novel Crop for Central Europe: Using a Broad Diversity
434 Set to Dissect Temperate-Adaptation." *Agronomy* 9 (9): 535.

435 Schaffasz, André, Steffen Windpassinger, Rod Snowdon, and Benjamin Wittkop. 2019.
436 "Reproductive Cold Stress Tolerance in Sorghum F1 Hybrids Is a Heterotic Trait."
437 *Agronomy* 9 (9): 508.

438 Sharifi, P. 2008. "Inheritance of Cold Tolerance in Rice at the Germination Stage." *Asian*
439 *Journal of Plant Sciences*.

440 Stich, Benjamin, Jens Möhring, Hans-Peter Piepho, Martin Heckenberger, Edward S Buckler,
441 and Albrecht E Melchinger. 2008. "Comparison of Mixed-Model Approaches for
442 Association Mapping." *Genetics* 178 (3): 1745–54.

443 Tao, Yongfu, Xianrong Zhao, Xuemin Wang, Adrian Hathorn, Colleen Hunt, Alan W
444 Cruickshank, Erik J van Oosterom, Ian D Godwin, Emma S Mace, and David R Jordan.
445 2020. "Large-Scale GWAS in Sorghum Reveals Common Genetic Control of Grain Size
446 among Cereals." *Plant Biotechnology Journal* 18 (4): 1093–1105.
447 https://doi.org/https://doi.org/10.1111/pbi.13284.

448 Thurber, Carrie S, Justin M Ma, Race H Higgins, and Patrick J Brown. 2013. "Retrospective
449 Genomic Analysis of Sorghum Adaptation to Temperate-Zone Grain Production."
450 *Genome Biology* 14 (6): 1–13.

451 Venkatesh, Jelli, Min-Young Kang, Li Liu, Jin-Kyung Kwon, and Byoung-Cheorl Kang.
452 2020. "F-Box Family Genes, LTSF1 and LTSF2, Regulate Low-Temperature Stress
453 Tolerance in Pepper (Capsicum Chinense)." *Plants* .
454 https://doi.org/10.3390/plants9091186.

455 Weir, B S, and C Clark Cockerham. 1984. "Estimating F-Statistics for the Analysis of
456 Population Structure." *Evolution* 38 (6): 1358–70. https://doi.org/10.2307/2408641.

457 Westengen, Ola T, Mark Atam Okongo, Leo Onek, Trygve Berg, Hari Upadhyaya, Siri
458 Birkeland, Siri Dharma Kaur Khalsa, Kristoffer H Ring, Nils C Stenseth, and Anne K
459 Brysting. 2014. "Ethnolinguistic Structuring of Sorghum Genetic Diversity in Africa and
460 the Role of Local Seed Systems." *Proceedings of the National Academy of Sciences* 111
461 (39): 14100 LP – 14105. https://doi.org/10.1073/pnas.1401646111.

bioRxiv preprint doi: https://doi.org/10.1101/2022.01.31.478463; this version posted February 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

462 Xie, Lixia, Zhengwei Tan, Yuan Zhou, Rongbao Xu, Laibao Feng, Yongzhong Xing, and
463 Xiaoquan Qi. 2014. "Identification and Fine Mapping of Quantitative Trait Loci for
464 Seed Vigor in Germination and Seedling Establishment in Rice." *Journal of Integrative*
465 *Plant Biology* 56 (8): 749–59.

466 Yamasaki, Masanori, Stephen I Wright, and Michael D McMullen. 2007. "Genomic
467 Screening for Artificial Selection during Domestication and Improvement in Maize."
468 *Annals of Botany* 100 (5): 967–73.

469 Yu, Guangchuang. 2020. "Using Ggtree to Visualize Data on Tree-Like Structures." *Current*
470 *Protocols in Bioinformatics* 69 (1): e96. https://doi.org/https://doi.org/10.1002/cpbi.96.

471