Experimental evolution in maize with replicated divergent selection identifies plant-height associated SNPs

Mila Tost[1,2], Cathy Westhues[1,2], Ginnie Morrison[3], Dietrich Kaufmann[1] and Timothy Beissinger[1,2,*]

[1]Department of Crop Science, Division of Plant Breeding Methodology, University of Goettingen, Carl-Sprengel-Weg 1, Goettingen, 37075, Germany

[2]Center for Integrated Breeding Research, University of Goettingen, Carl-Sprengel-Weg 1, Goettingen, 37075, Germany

[3]University of Missouri, Division of Biological Sciences, 105 Tucker Hall, Columbia, MO, 65211, USA

[*]**Corresponding author:** Timothy Beissinger (timbeissinger@gmail.com)

Running title: Replicated experimental evolution maize

Key words: Experimental evolution, Population genetics, maize, selection signature mapping, plant height

Abstract

Experimental evolution studies are common in agricultural research, where they are often deemed "long term selection". These are often used to perform selection mapping, which involves identifying markers which were putatively under selection based on finding signals of selection left in the genome. A challenge of previous selection mapping studies, especially in agricultural research, has been the specification of robust significance thresholds. This is in large part because long term selection studies are rarely replicated. Usually, significance thresholds in long term selection experiments are based on outliers from an empirical distribution or based on drift simulations. Both of these approaches are prone to missing true positives or including false positives. Under laboratory conditions with model species, replicated selection has been shown to be a powerful tool, especially for the specification of significance thresholds. In this study, we conducted divergent, replicated selection for short and tall plant-height in a random mating maize population under real field conditions. Selection of the 5% tallest or shortest plants was conducted for three generations. Significance thresholds were specified using the false discovery rate for selection (FDRfS) based on a single marker-based statistic leveraging replicated selection ($F_{ST}Sum$). Overall, we found 44 significant SNPs. Thirty-four were located within a region of 1.02 Mb comprising the coding region of the plant-height genes *Dwarf1* and *iAA8*. We applied a window-based analysis and haplotype block analysis to further dissect the pattern of selection in significant regions of the genome. We observed patterns of strong selection in the subpopulations selected for short plant height.

Introduction

The purpose of experimental evolution studies is to understand how populations evolve in response to a selective pressure (Janick, 2010a, 2010b). In agricultural research, they are commonly employed and often deemed "long term selection" studies (Dudley and Lambert, 2010; Moose, Dudley and Rocheford, 2004, Beissinger *et al.,* 2014). For instance, the Illinois long-term selection experiment for maize protein content began in 1896 at the University of Illinois and has been conducted ever since (Dudley and Lambert, 2010; Moose et al. 2004). Experimental evolution studies are also common in model organisms like mice (Firman and Simmons, 2009; Firman *et al.*, 2015), *Drosophila melanogaster* (Teotónio *et al.*, 2009; Burke *et al.*, 2010; Turner *et al.*, 2011; Phillips *et al.*, 2016)*, E. coli* (Herring *et al.*, 2006; Barrick *et al.*, 2009)*, Mimulus guttatus* (monkey flowers) (Kelly *et al.*, 2006; Angert, Bradshaw Jr and Schemske, 2008), and yeast (Payen *et al.*, 2016; Linder *et al.*, 2021). Early experimental evolution studies focused on the evolution of phenotypes alone, but now sequencing technologies allow researchers to study the effect of selection at the DNA level (Burke *et al.,* 2010; Parts *et al.*, 2011; Orozco-terwengel *et al.,* 2012; Remolina *et al.,* 2012).

Many experimental evolution studies in model organisms work with replicates (Teotónio and Rose, 2000; Teotónio *et al.*, 2009; Phillips *et al.*, 2016). These studies are often referred to as Evolve & Resequence (E&R) experiments (Turner *et al.*, 2011). E&R experiments commonly use pooled-based sequencing of the different subpopulations (Turner *et al.*, 2011, Orozco-terWengel *et al.,* 2012; Long *et al.*, 2015; Barghi *et al.*, 2017). In agricultural species, experimental evolution studies that leverage replicated selection populations are exceedingly rare (Dudley and Lambert, 2010, Beissinger *et al.*, 2014; Kumar *et al.*, 2021). On the other hand,

replication plays a very important role in E&R studies because it can help to identify putatively selected sites and differentiate those from sites caused by drift (Baldwin-Brown, Long and Thornton, 2014; Kofler and Schlötterer, 2014; Turner and Miller, 2012; Linder *et al.*, 2021). This assumes that selection is repeatable, while randomly acting forces like drift are not (Teotónio and Rose, 2000; Turner and Miller, 2012; Burke, Liti and Long, 2014).

Genetic drift is a complicating factor in experimental evolution studies because it can be difficult to differentiate from selection (Krimbas and Tsakas, 1971; Turner *et al.*, 2011; Beissinger *et al.*, 2014). This is a recurring challenge in unreplicated selection mapping studies, because without replicating the evolutionary process, no amount of sampling can estimate the variability associated with genetic drift (Krimbas and Tsakas, 1971; Turner *et al.*, 2011; Hirsch *et al.*, 2014; Barghi, Hermisson and Schlötterer 2020, Kumar *et al.*, 2021). The variability introduced by genetic drift depends on population size and the number of generations of selection (Johannson *et al*., 2010, Turner and Miller, 2012; Kessner and Novembre, 2015). Replicated selection populations can help differentiate neutral loci subject to drift alone, from putatively selected loci (Turner and Miller, 2012; Kofler and Schlötterer, 2014; Barghi *et al.,* 2019; Linder *et al.,* 2021). Replicates can be analyzed to calculate a null expectation and to derive an approximation of the number of false positive observations in the experiment.

In previous experimental evolution studies on crops, significance thresholds for identifying signals of selection were frequently derived based on outlier quantiles of the empirical distribution (Beissinger *et al.*, 2014; Hirsch *et al.*, 2014; Lorenz *et al.*, 2015; Wang *et al.*, 2018;

Ayalew *et al.*, 2020). This assumes that selection is a locus-specific force; putatively selected loci should be found at the tail of the empirical distribution (Akey *et al.*, 2002; Akey, 2009). Drift is assumed to act uniformly genome-wide, so drifting loci dictate the bulk of the empirical distribution (Akey *et al.*, 2002; Akey, 2009). However, it has been recognized for a very long time that outlier-based thresholds are arbitrary (Akey *et al.*, 2002; Akey, 2009).

Another common approach for specifying significance thresholds for selection is to conduct drift simulations (Kumar *et al.*, 2021; Turner *et al.*, 2011). This approach is an improvement over arbitrary outlier thresholds, but it is impossible to know with certainty that simulations of genetic drift fully mimic the complex demographic scenarios experienced by a real-world population, including an experimental one.

For this study, we performed replicated divergent selection for plant height on a random mating maize population for three generations. Plant height is an ideal model trait because it is highly heritable and highly polygenic (Pfeiffer *et al*., 2014, Mazaheri *et al.,* 2019). We replicated selection on four subpopulations, two of which were selected for tall plant height and two which were selected for short plant height. Each subpopulation was maintained with a population size of approximately 5,000 individuals, making this the largest replicated experimental evolution study conducted on a crop species to date. Every generation, we selected the 5% tallest or shortest plants (~250 plants). The truncation threshold of 5% tallest or shortest plants was calculated based on 96 randomly sampled plants in from each

111 subpopulation (in total 384 plants). The 384 randomly sampled plants from the fourth
112 generation were also used for genotyping-by-sequencing (GBS) to conduct selection mapping.
113
114 We calculated F_{ST} for each identified marker in the genome (Weir and Cockerham, 1984). Then,
115 we calculated a version of F_{ST} leveraging replicated selection. Finally, we generated significance
116 thresholds based on Turner and Miller's (2012) false discovery rate for selection (FDRfS). Finally,
117 we compared these replication-based significance thresholds to those generated from the
118 empirical distribution of F_{ST} and from drift simulations.
119
120 Regardless of the method used to specify significance thresholds, regions corresponding to
121 selection can be quite large and contain neutral linked loci (Barghi and Schlötterer, 2019;
122 Franssen, Barton and Schlötterer, 2016; Tobler *et al.*, 2014). Single marker-based statistics like
123 F_{ST} are prone to identify neutral loci that have hitchhiked along with those that are under
124 selection (Franssen, Barton and Schlötterer, 2016; Tobler *et al.*, 2014). Several experimental
125 evolution studies have additionally investigated reconstructed haplotype blocks to disentangle
126 the complex pattern of linkage (Franssen, Barton and Schlötterer, 2016; Michalak *et al.*, 2018).
127 These investigations showed that often a large number of putatively significant sites
128 corresponds to only a small number of significant haplotype blocks (Franssen, Barton and
129 Schlötterer, 2016; Michalak *et al.*, 2018). Therefore, we applied a cubic smoothing spline
130 approach to identify window boundaries for selection regions (Beissinger *et al.*, 2015). This
131 procedure may also reduce the inflation of observations due to linked loci (Barghi and

Schlötterer, 2019). Additionally, we phased our data and investigated the haplotype blocks under selection.

Material and methods

Plant materials and experimental design

The Shoepeg population is a maize Southern dent landrace population (Doebley *et al.*, 1988), which was collected as open-pollinating variety in 1960 in Missouri, USA. We ordered Shoepeg material from the National Plant Germplasm System as accession PI 269743 (U.S. National Plant Germplasm System, 2021). Seeds were greenhouse-propagated to generate our starting population for selection. The formation of the Shoepeg experimental evolution population and results preceding the first generation of selection are described in Gyawali *et al.* (2019). To summarize, in 2016 the Shoepeg population was divided into four subpopulations, each consisting of approximately 5000 plants which were allowed to open-pollinate. In each subpopulation, plant height was measured (Figure 1). In total, 384 (96 x 4) individual plants out of approximately 20,000 (5000 x 4) plants were measured across all four subpopulations (Figure 1). Based on these plant height measurements, a truncation threshold corresponding to approximately the 5% shortest or tallest plants was calculated for each subpopulation (Figure 1) (Gyawali *et al.*, 2019). Two subpopulations were selected for short plants, and two were selected for tall plants. The ~5% shortest or tallest plants from each subpopulation were harvested as females to generate the next generation (Figure 1) (Gyawali *et al.*, 2019).

153 This procedure was repeated in 2017 and 2018 (Figure 1). Ultimately, we generated four
154 experimentally evolved subpopulations, two comprised of short-selected plants and two
155 comprised of tall-selected plants (Figure 1). The populations were grown again in 2020 and plant
156 height was measured for 96 randomly-chosen representative plants in each subpopulation (in
157 total 384 plants). Additionally, tissue was collected from the same 384 plants for genotyping
158 (Figure 1).
159

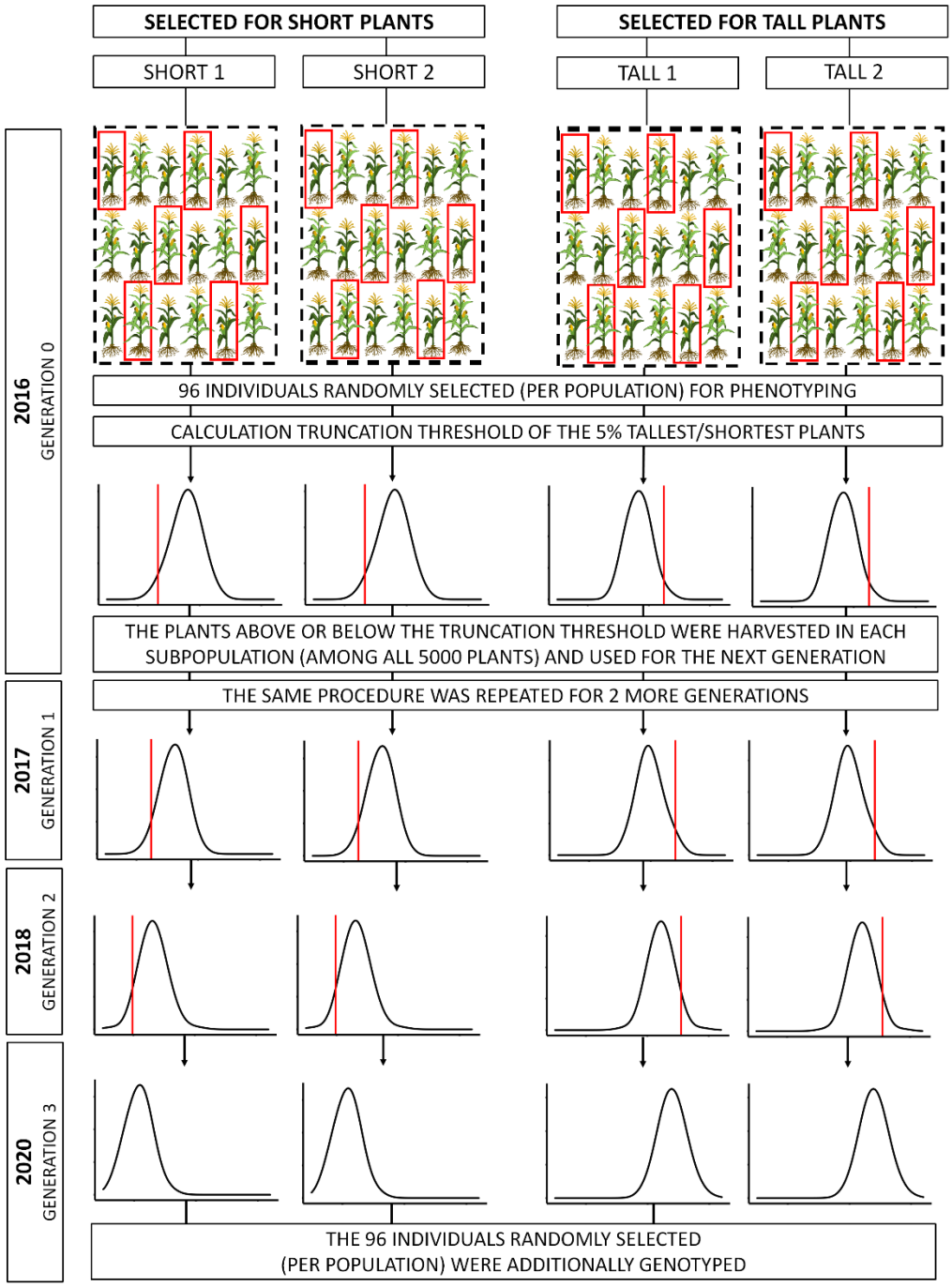


Figure 1: Simplified scheme for the development of subpopulations selected in divergent directions with their replicates for the three generations of selection. The subpopulations short 1 and 2 were selected for short plant height and the subpopulations tall 1 and 2 were selected for tall plant height.

Every year after the first, the subpopulations were grown in isolated fields at least 200 m apart and away from any other maize to avoid contamination via cross-pollination (Nieh *et al.*, 2014; Gyawali *et al.*, 2019). In 2016, 2017 and 2018, the subpopulations were grown in Columbia,

168 Missouri, USA (Gyawali *et al.*, 2019). In 2020 the subpopulations were grown in Göttingen,
169 Germany. In Missouri, each field had an approximate size of 0.1 ha (Gyawali *et al.*, 2019). The
170 row-spacing was 91 cm and the space between the plants was 15 cm (Gyawali *et al.*, 2019). In
171 Göttingen, the field size for three of the fields was 0.25 ha. One field was only 0.125 ha due to
172 fewer seeds caused by a lower-than-expected harvest from one of the subpopulations in 2018.
173 The distance between the seeds in Göttingen was approximately 33.3 cm and the row-spacing
174 was approximately 1 m. The fields in Goettingen experienced noteworthy but non-devastating
175 crow damage.
176
177 Plant height measurements were taken after flowering, from the ground up to the ligule of the
178 flag leaf. Phenotypic data was analyzed using R version 4.0.3 (R Core Team, 2021). For
179 phenotypic data analysis, data from 2016, 2017, 2018 and 2020 was used. The realized
180 heritability from the breeder's equation (h^2_{bs}) was calculated as $h^2_{bs} = \frac{R}{S}$ (Lush, 1937), where S is
181 the selection differential and R is the response to selection. The selection differential was
182 calculated as the difference between the mean of the population and the mean of the selected
183 portion (Lush, 1937). The response to selection was calculated as difference between the mean
184 of the previous generation and the mean of the next generation (Lush, 1937). This was done
185 between all subsequent generations.
186
187 Flowering time was measured because in maize it tends to be correlated with plant height
188 (Peiffer *et al.*, 2014; Mazaheri *et al.,* 2019). This could lead to indirect selection for flowering

189 time, thereby effecting the ability of plants to mate randomly. The possibility of this impacting
190 the effective population size is addressed below.
191
192 **Genotyping, quality control and filtering**
193 In 2020, the 96 randomly chosen plants from each subpopulation in generation four (384
194 individual plants in total) were genotyped using a genotyping-by-sequencing (GBS) approach.
195 Leaf tissue from those 384 plants was freeze-dried and shipped to the University of Wisconsin
196 Biotechnology center, Madison, WI, USA for DNA extraction, library preparation, and
197 sequencing via one lane on a NovaSeq6000 machine. Each subpopulation was sequenced on its
198 own sequencing run. Total sequencing yield from each run ranged from 125,000,000 to
199 300,000,000 fragments. Fragments had read lengths between 73 and 151 base pairs. GBS was
200 conducted using paired-end sequencing with the restriction enzyme *Ape*KI (Elshire *et al.*, 2011).
201 For SNP variant calling the GB-eaSy pipeline was used (Wickland *et al.*, 2017, Tange, 2011).
202 Reads were aligned to the 5th version of the B73 reference genome (Maize GDB, 2021) with the
203 Burrows Wheeler Aligner (Li, 2013). In total, 15,390,958 SNPs were called by the GB-eaSy
204 pipeline (Wickland *et al.*, 2017). In the GB-eaSy pipeline, the SNPs were filtered for a minimum
205 mapping and base quality of 30 with the BCFtools software (Danecek *et al.*, 2021). After filtering,
206 8,302,778 SNPs were kept. The mapping quality ranged from 30 to 60, and was on average
207 57.44. The raw read depth ranged from 161 to 4,999, and was on average 1,705 reads. The
208 average distance between adjacent markers was 257.67 bp.
209

210 Subsequent SNP filtering was conducted in R version 4.0.3 (R Core Team, 2021). Six individual
211 samples with a marker coverage below 0.1 were removed from the analysis. The 10% quartile of
212 marker coverage 0.6042, and 90% quartile was 0.7014 (supplementary figure S1). Markers with
213 an average sample read depth below 1 or above 10 were removed. This filtering threshold was
214 determined based on the distribution of read depth across all samples and markers
215 (supplementary figure S2). Furthermore, we required that every SNP marker was observed at
216 least 40 times in each subpopulation. Monomorphic and multi-allelic (>2) markers were
217 removed from the data set. The final data set contained 4,029,092 SNP markers, which were
218 used for the downstream analysis.
219

220 **Allele frequency estimation**

221 The allele frequency at each SNP was calculated in R according to its maximum-likelihood
222 estimate, which is computed by dividing the number of observations of a particular allele by the
223 number of observations in total (Beissinger *et al.*, 2014). The script is available at
224 (https://github.com/milaleonie/ExpEvo_with_replicated_selection).
225

226 **Estimating effective population size**

227 Effective population size was calculated using the known demographic parameters of the
228 populations and based off the formula, $N_e = \frac{4 * N_f * N_m}{N_f + N_m}$, where N_m is the number of male and N_f
229 the number of female individuals (Crow and Kimura, 1970). In every generation the ~5%
230 shortest or tallest plants were harvested, amounting to ~250 plants. Under the assumption of
231 complete random mating, all 5,000 plants could contribute as male parents for the next

232 generation. In field trials, the assumption of random mating of all 5,000 plants is violated
233 because of the effects of assortative mating (Allard, 1999). The number of male parents
234 contributing to the next generation is reduced due to varying flowering dates and limited spatial
235 pollen dispersion (Allard, 1999). Therefore, we evaluated the flowering dates of the 96 randomly
236 chosen plants from each subpopulation to approximate the number of maximal simultaneously
237 flowering tassels and silks in the entire population. We looked for the quantile in which the
238 majority of silks is mature and receptive, because it was assumed that silks remain receptive up
239 to five days after silk emergence (Nieh *et al.*, 2014). We calculated the number of flowering
240 tassels during this time interval and projected it onto the entire subpopulation. This procedure
241 is explained in more detail in supplementary figure S3.
242
243 **Calculation of linkage disequilibrium (LD) decay**
244 The extent of linkage disequilibrium (LD) was estimated based on 100,000 SNP markers with
245 TASSEL v5 using the squared correlation between markers, r^2, over a window size of 2,000 bp
246 (Bradbury *et al*., 2007). We filtered the individuals out with a marker coverage below 0.6. For
247 this analysis, we required that every SNP marker was observed at least 80 times in each
248 subpopulation. This filtering resulted in 1,243,604 SNP markers. Additionally, we thinned the
249 data set by randomly sampling 10,000 markers per chromosome. LD decay was modeled with a
250 nonlinear regression model as expected value r^2 under drift-recombination equilibrium as
251 $E(r^2) = \frac{1}{(1+C)}$ and $C = 4Nc$ with N as the number of individuals at each site and C as the
252 recombination coefficient between sites (Remington *et al*., 2001). We also created a pairwise LD
253 heatmap with the R package LDheatmap (Shin *et al*., 2006).

254

255 **Scan for selection signatures**

256 Our scan for selection was based on a statistic derived from F_{ST}. F_{ST} was calculated according to

257 $F_{ST} = \frac{s^2}{\mu(p)*\mu(1-p)+\frac{s^2}{r}}$, where s^2 is the sample variance of the allele frequency between

258 populations, $\mu(p)$ is the mean allele frequency between the subpopulations and r is the number

259 of populations (Weir and Cockerham, 1984). The F_{ST} statistic was calculated between the

260 divergently selected subpopulations (Short 1 vs Tall 1; Short 2 vs Tall 2). Additionally, the F_{ST}

261 statistic was calculated between the subpopulations selected in the same direction (Short 1 vs

262 Short 2; Tall 1 vs Tall 2).

263

264 **Calculation of *F_{ST}Sum***

265 The sum of F_{ST} was calculated between the two non-redundant comparisons of divergently

266 selected subpopulations (Short 1 vs Tall 1; Short 2 vs Tall 2) and subpopulation selected in the

267 same direction (Short 1 vs Short 2; Tall 1 vs Tall 2). Assessing divergence among replicates can

268 help in the distinction between putatively selected loci and false positives (Remolina *et al.,*

269 2012; Kofler and Schlötterer, 2014)

270

271 **Window-based analysis**

272 We implemented a window-based analysis to assess linked selection. A cubic smoothing spline

273 was applied to the single-marker based F_{ST}*Sum* values. From the fitted spline, inflection points

274 were calculated and used as window boundaries for newly defined regions. This analysis was

275 performed with the R package GenWin (Beissinger *et al.*, 2015). Briefly, the GenWin package

returns window boundaries that we used to define regions with linked markers that may be analyzed together (Beissinger *et al*., 2015). However, those regions are of variable sizes, so we used the W statistic (*Wstat*), as defined by Beissinger *et al.* (2015), to correct for this. Wstat is defined according to $Wstat = \frac{(X-\mu)}{\sqrt{\frac{s^2}{n}}}$, where X is the mean $F_{ST}Sum$ of a window, μ is the mean $F_{ST}Sum$ of the entire data set, s^2 is the sample variance of $F_{ST}Sum$ values across the entire data set and n is the number of markers. *Wstat* can therefore be used to compare regions, despite each consisting of a different number of markers (Beissinger *et al*., 2015).

Calculation of significance thresholds

Significance thresholds for selection were calculated in three ways: 1) based on the empirical distribution; 2) based on drift simulations; and 3) based on the FDRfS (Turner and Miller, 2012).

1) The significance thresholds based on the empirical distribution were calculated by taking the 99.9th and the 99.99th percentile of the entire distribution of the $F_{ST}Sum$ values (Akey, 2009; Beissinger *et al.*, 2014; Hirsch *et al.*, 2014; Ma *et al.*, 2015; Cesconeto *et al.*, 2017; Wang *et al.*, 2018; Ayalew *et al.*, 2020).
2) To calculate the significance thresholds based on drift simulations, we first simulated drift 1,000,000 times based on the demographic parameters of the replicated populations. Next, we calculated $F_{ST}Sum$ for the simulated allele frequencies. The 99.9999th percentile of all values generated by the drift simulations was used as the significance thresholds (Kumar *et al.*, 2021). The drift simulations are described in more detail below.

297 3) High $F_{ST}Sum$ values calculated between divergently selected subpopulations are
298 observed when a site shows the same pattern of differentiation in both comparisons.
299 Sites with high $F_{ST}Sum$ calculated between subpopulations selected in the same direction
300 are false positive observations for selection for plant height. By dividing the number of
301 sites diverged between the subpopulations selected in the same direction by the number
302 of sites diverged between the divergently selected subpopulation for a certain $F_{ST}Sum$,
303 we obtained the false discovery rate for selection (FDRfS) for $F_{ST}Sum$. The FDRfS was
304 calculated for all possible $F_{ST}Sum$ values (Turner and Miller, 2012). We identified the
305 $F_{ST}Sum$ which gave an FDRfS <5%. The corresponding $F_{ST}Sum$ value was used as our
306 significance threshold. The FDRfS was also used with *Wstat* to generate window-based
307 values. The significance threshold was chosen as the *Wstat* value that excluded all
308 comparisons between subpopulations selected in the same direction.
309

310 **Simulations of drift**

311 Drift was simulated with an R function
312 (https://github.com/milaleonie/ExpEvo_with_replicated_selection), which was implemented in
313 R version 4.0.3 (R Core Team, 2021). The function was implemented using a wrapper-script
314 based on that previously published by Kumar *et al.* (2021), which enabled the implementation of
315 the drift simulator over a large set of markers (Kumar *et al.*, 2021). We conducted 1,000,000
316 simulations. For each simulation, initial allele frequencies were set based on the allele frequency
317 spectrum observed in generation 0 (Gyawali *et al.*, 2019). Drift was simulated with a population
318 size of 5,000 individuals with 250 female and 5,000 male individuals for three generations. We

319 assumed that every ear contributed ~500 kernels. Every kernel could have been pollinated by
320 one of the male parents, which resulted in much higher harvest than 5,000 kernels. Therefore,
321 5,000 kernels were randomly drawn from the entire harvest to represent the seeds planted for
322 the next generation.
323
324 **Haplotype estimation and haplotype block calculation**
325 The data were phased using fastPHASE version 1.4.8 (Sheet and Stephens, 2006) for further
326 dissection. We phased the genomic data of the region from 9.437 to 10.457 Mb on chromosome
327 3 with 10 iterations of the expectation-maximization (EM) algorithm (Sheet and Stephens,
328 2006). The model was supplied with labels indicating the different subpopulations. Haplotype
329 blocks were calculated with the R package HaploBlocker (Pook *et al.*, 2019). Additionally, we
330 conducted the same haplotype block analysis in 39 randomly selected regions across the entire
331 genome, with the same physical length and marker density like the region from 9.437 to 10.457
332 Mb on chromosome 3. We also investigated the region from 9.437 to 10.457 Mb on
333 chromosome 3 in generation 0 in the haplotype block analysis.
334
335 ## Results
336 **Phenotypic evaluation**
337 We observed notable and significant plant height differences between the subpopulations
338 selected in opposite directions (Figure 2). After three generations of selection, the mean plant
339 height difference between subpopulations selected in the opposite direction was 66.55 cm (P ~
340 0***). The mean plant height decreased in the subpopulations selected for short plant height by

22.92 cm (P ~ 0***) and increased in the subpopulations selected for tall plant height by 30.61 cm (P ~ 0***). Throughout the experiment, the response to selection was always negative in the subpopulations selected for short plants and always positive in the subpopulations selected for tall plants (Figure 2).

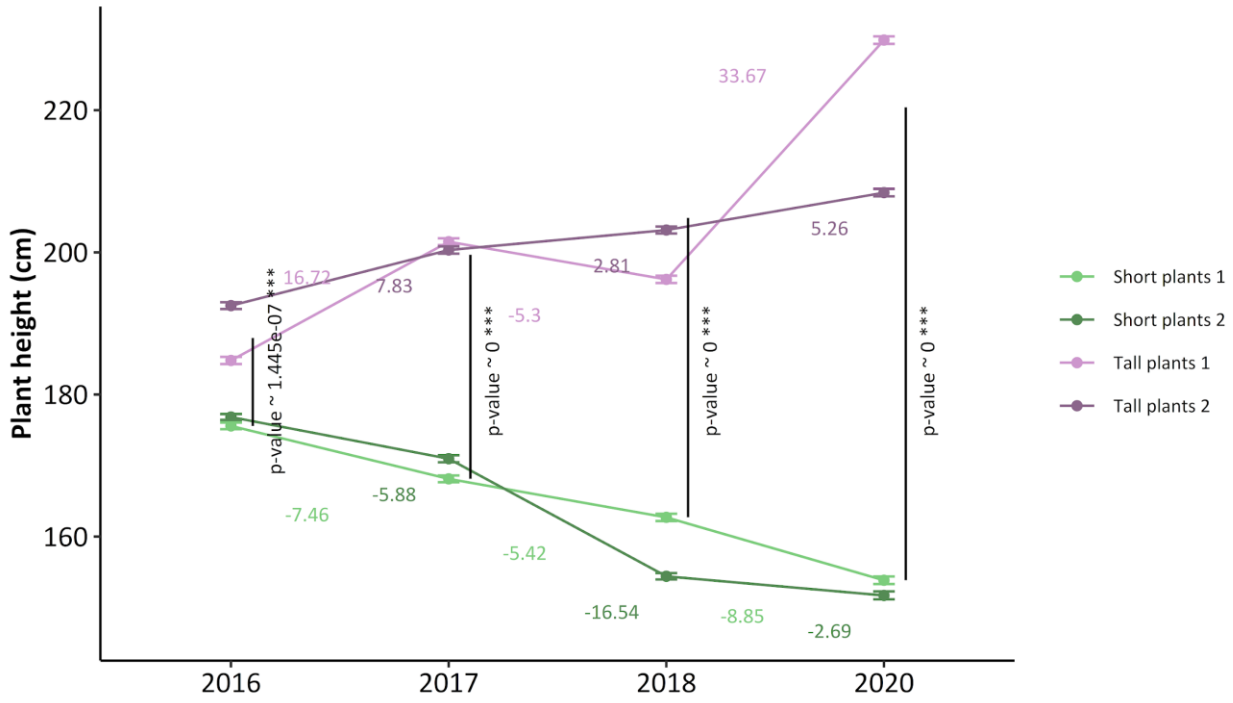


Figure 2: Mean plant height in each generation of selection. Error bars indicate a confidence interval for our estimation of means.

Realized heritability was computed by using the breeder's equation (h^2_{bs}) (Lush, 1937), and was estimated to be 0.3. Since each population was evaluated in a different environment (i.e. the year and location where it was selected), this application of the breeder's equation should be considered only approximate. Plant height and flowering time were correlated in the subpopulations short 1, short 2, tall 1, and tall 2 with correlation coefficients of 0.2183 (P ~ 0.0376**), -0.0384 (P ~ 0.7165), 0.2709 (P ~ 0.0107**) and 0.4572 (P ~ 0***), respectively.

356

357 **Effective population size and inbreeding**

358 The effective population size under the assumption of complete random mating was 952
359 individuals. Due to differences in flowering time between plants, the assumption of random
360 mating is violated. Based on observed flowering times, we estimated that ~ 59.38% and ~
361 82.29% in the short and ~ 56.25% and ~ 38.54% in the tall subpopulations had the potential to
362 pollinate one another. When we re-estimated effective population size based on this number of
363 male plants, the effective population size was estimated ~ 922 and ~ 943 individuals in the short
364 subpopulations and ~ 918 and ~ 885 individuals in the tall subpopulations.

365

366 **Test for selection**

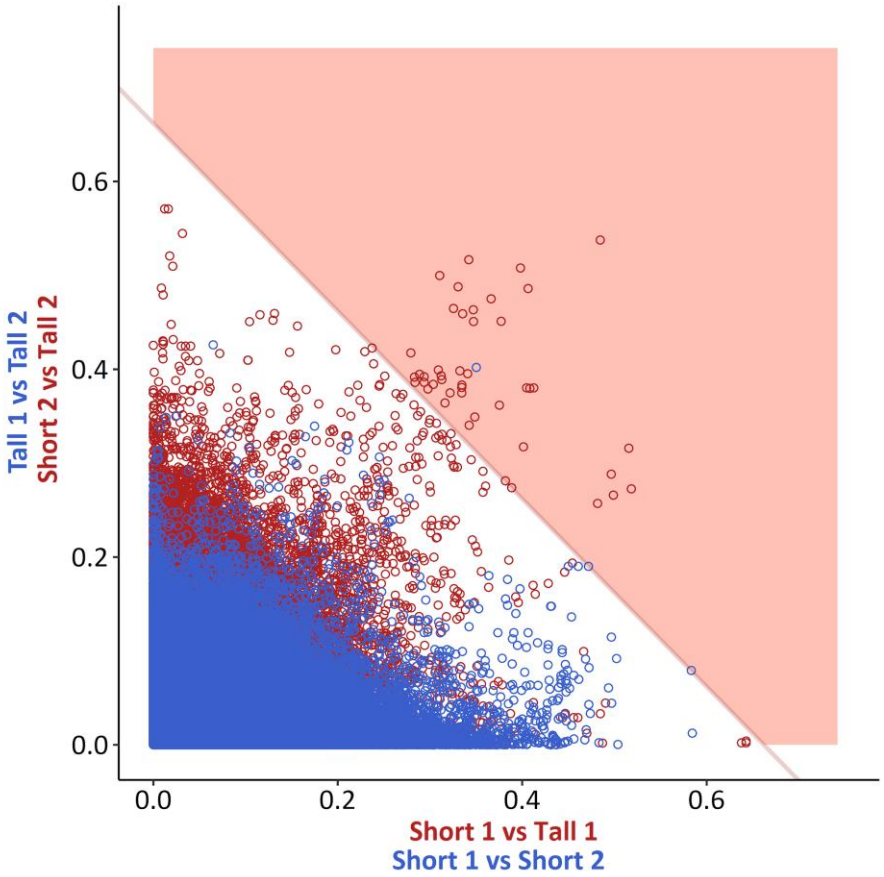
367 To visualize the distribution of F_{ST} values observed between subpopulations selected in the
368 same and opposite directions, we plotted the F_{ST} values calculated between the different
369 comparisons against each other (Figure 3). This type of plot, but based on allele frequency
370 differences, was previously named a "Sauron plot" by Turner and Miller (2012). We created this
371 plot with F_{ST}, which resulted in a "Quarter-Sauron" plot, because F_{ST} ranges from 0 to 1, whereas
372 allele frequency differences range from -1 to 1. A "Sauron" plot based on the allele frequency
373 differences, as was originally implemented by Turner and Miller (2012) is provided as
374 supplemental file S4. This plot depicts how the FDRfS was computed and provides a visualization
375 of the scope of drift and selection.

376

377 In the plot, blue points correspond to the sum of F_{ST} values calculated between the
378 subpopulations selected in the same direction for each marker, while red points correspond to

379 the sum of F_{ST} values calculated between the subpopulations selected in opposite directions for
380 each marker (Figure 3). Therefore, the FDRfS can be calculated for every area of the plot, which
381 corresponds to a certain $F_{ST}Sum$ value (Figure 3). The region of the plot where the FDRfS is <5%
382 is indicated by the semi-transparent red region (Figure 3). In this region, 44 SNPs with $F_{ST}Sum$
383 >0.662 were identified for subpopulations selected in opposite directions, while only two SNPs
384 displayed a $F_{ST}Sum$ >0.662 between subpopulations selected in the same direction (Figure 3).
385 This leads to an exact FDRfS of 4.5%. The FDRfS for every possible $F_{ST}Sum$ and the number of
386 markers observed at this value are available in supplementary figure S5. $F_{ST}Sum$ was only
387 calculated for comparisons without missing data, to make it less sensitive to outliers.

388



389

390 **Figure 3: "**Quarter-Sauron" plot of genetic differentiation for $F_{ST}Sum$ observed between the subpopulations
391 selected in the same direction (blue) and in opposite directions (red). Each dot represents one SNP. The semi-
392 transparent red region corresponds to a false discovery rate for selection (FDRfS) <5%.
393

394 **Comparison of significance thresholds and candidate gene identification**

395 We also evaluated significance thresholds based on outlier thresholds and drift simulations.
396 Thresholds based on the 99.9th and the 99.99th percentile of the empirical distribution were
397 exceeded, respectively, by 1,422 and 143 $F_{ST}Sum$ values observed at single SNP markers (Figure
398 4A). In total, 93 and 10 regions exceeded the thresholds based on the 99.9th and the 99.99th
399 percentile of the empirical distribution of *Wstat* values (Figure 4B). Significance thresholds
400 based on drift simulations were exceeded by 9,998 $F_{ST}Sum$ values observed at single SNP
401 markers (Figure 4A). Significance thresholds based on the FDRfS were exceeded by the $F_{ST}Sum$
402 observed at 44 single SNP markers (Figure 4A) and by *Wstat* observed across 3 different regions
403 (Figure 4B). $F_{ST}Sum$ and *Wstat* values calculated between the subpopulations selected in the
404 same direction are depicted in Figure 4C and Figure 4D.

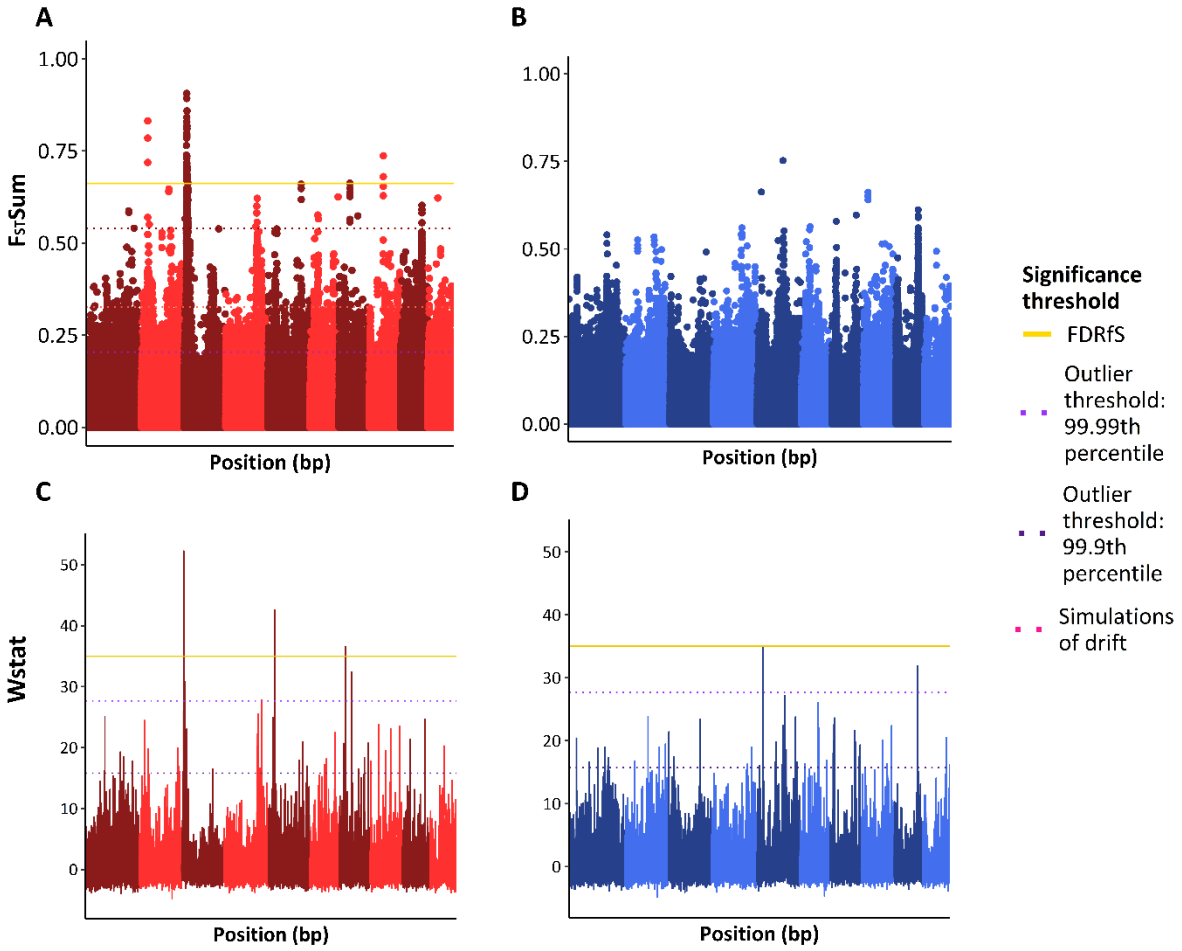


Figure 4: Differentiation along the entire genome expressed by F_{ST} at each SNP marker between the subpopulations selected in opposite directions (A) and the same direction (B) and expressed by *Wstat* values across regions (C, D) with the significance thresholds based on the 99.9th percentile of the empirical distribution (light purple), based on the 99.99th percentile of the empirical distribution (dark purple), drift simulations (pink), and by the false discovery rate for selection (FDRfS) (yellow).

Candidate plant height SNPs

All three regions identified based on *Wstat* overlapped with coding regions in the B73 v5 annotated gene models (Maize GDB, 2021). For two of the regions, gene models were associated with annotated gene IDs (Table 1). For the gene *impb24 - importin beta24*, no information was found based on a literature review (Table 1). The gene *pip2e* belongs to the plasma membrane intrinsic protein subfamily PIP2, which has been described as associated with salt stress (Fox *et al*., 2021). The gene model *Zm00001eb122360*, which coding region was

419 located within the identified region on chromosome 3, did not have an associated candidate
420 gene (Table 1).
421
422 Twenty-seven out 44 SNPs identified by F_{ST}*Sum* were located within or 85 kb from B73 v5 (Maize
423 GDB, 2021) annotated gene models. Many markers correspond to the same gene models and
424 appear to be located in the same region (Table 1). On chromosome 3 we found 34 significant
425 markers, which were located in a region from 9.437 to 10.457 Mb (Table 1). All 34 markers in
426 that region show very similar allele frequency changes and resulting differentiation statistic
427 (Table 1). The plant height candidate genes *Dwarf1/GA 3-oxidase 1* and the *iAA8 - Aux/iAA-*
428 *transcription factor 8* are located in this region.
429
430 The plant height candidate genes *Dwarf1* and the *iAA8* have already been shown to be directly
431 involved in plant height differences in many studies (Teng *et al.*, 2013; Peiffer *et al.*, 2014; Liu,
432 Fernie and Yan, 2020; Ramos Báez *et al.*, 2020). Therefore, these candidate genes were used as
433 ground truth to test the method. Other identified gene models did not show a direct connection
434 to plant height based on the literature review.
435
436 Four markers were located directly within the coding region of the *iAA8* gene (Table 1). Two
437 markers were found near the *Dwarf1* gene and were located ~ 85 kb upstream from this gene
438 (Table 1). The window-based region with the highest *Wstat* value was also located on
439 chromosome 3 within the previously described region ranging from 9.437 to 10.457 Mb (Table
440 1). This 0.12 Mb region is ~ 138 kb away from *iAA8,* and is ~ 518 kb away from *Dwarf 1* (Table 1).

441 The region is also 31 kb away from a significant marker identified in the Gyawali *et al.,* (2019)
442 study (Table 1). Other single SNP markers were also within or close to the coding region of other
443 annotated gene models, but for these no previous evidence for their impact on plant height was
444 found based on a literature review (Table 1).

445

446 **Table 1:** Significant SNP markers and regions with their associated gene models and candidate genes.

Chr	Position (bp)	F_{ST}*Sum* value	Region (bp)	*Wstat* value	Nr of SNPs	Maize gene annotation and candidate gene nearby	Distance to gene (bp)
2	19993292	0.7187				nothing	
2	19993690	0.8316				nothing	
2	19993788	0.7849				nothing	
3	8055976	0.7648				**Zm00001eb121570:** IDP7380	within the coding region
3	8055983	0.7909					
3	8055985	0.7388					
3	9435824	1.0224				**Zm00001eb122210:** No candidate gene associated	within the coding region
3	9435914	0.9061					
3	9474493	0.8283				**Zm00001eb122220:** No candidate gene associated	within the coding region
3	9529777	0.8925				nothing	
3	9529813	0.8413				nothing	
3	9529816	0.7980				nothing	
3	9529823	0.8103				nothing	
3	9529845	0.7946				nothing	
3	9529863	0.7903				nothing	
3	9544069	0.7926				**Zm00001eb122230:** No candidate gene associated	within the coding region
3	9544090	0.7879					
3	9544092	0.7849					
3	9545220	0.7307					
3	9548992	0.6977				**Zm00001eb122250:** No candidate gene associated	within the coding region
3	9781288	0.7179				**Zm00001eb122280:** pco120183	within the coding region
3	9781314	0.7361					within the coding region
3	9784253	0.7092				**Zm00001eb122290:** No candidate gene associated	within the coding region
3	9784255	0.6966					
3	9784270	0.7150					
3	9792757	0.6828				nothing	
3	9830677	0.7018				**Zm00001eb122300:** No candidate gene associated	within the coding region
3	9909752	0.6876				**Zm00001eb122350:** No candidate gene associated	within the coding region
3			9911000-9923000	52.24	62	**Zm00001eb122360:** No candidate gene associated	overlaps with the coding region
3	9921856	0.6766					
3	9921871	0.6833					

3	9921927	0.6749					
3	9921995	0.8183					
3	9922080	0.7083					
3	9922083	0.7052					
3	10053938	0.6698				**Zm00001eb122410:** iAA8 - Aux/iAA-transcription factor 8	within the coding region
3	10053961	0.6796					
3	10053964	0.6857					
3	10062671	0.6967					
3	10355793	0.8104				**Zm00001eb122500:** Dwarf1 - dwarf plant1	
3	10355846	0.8589					
3	15011305	0.6630				nothing	
5			30103000-30115000	42.61	22	**Zm00001eb221820:** impb24 - importin beta24	overlaps with the coding region
7	79405796	0.6624				nothing	
7			42687000-42725000	36.55	24	**Zm00001eb306380:** pip2e - plasma membrane intrinsic protein2	overlaps with the coding region
8	93026326	0.7371				nothing	
8	93026386	0.6803				nothing	

Haplotype block analysis

We investigated the region from 9.437 to 10.457 Mb on chromosome 3 in more detail by dissecting the haplotypes in the population (Figure 5). The haplotype blocks located within and around the coding region of the gene *Dwarf1* indicate allele fixation in almost all subpopulations (Figure 5B). Only in subpopulation tall 2 did we observe three different haplotype blocks around and within the coding region of *Dwarf1* (Figure 5B). But, two of the haplotype blocks contained only one haplotype each (Figure 5B). The majority of observed haplotypes belonged to a single block (Figure 5B). Within and approximately 10 kb from the coding region of the gene *iAA8*, we observed three different haplotype blocks in the two short subpopulations and four in the two tall subpopulations (Figure 5B). The majority of haplotypes identified in the short subpopulations were located in one block (Figure 5B). The other two haplotype blocks contained only one haplotype each (Figure 5B). In the short plant subpopulations, we observed larger haplotype blocks (Figure 5B) than in the tall subpopulations. The block size was on average

461 646.4 bp in subpopulation short 1 and 809 bp in subpopulation short 2. In subpopulation tall 1
462 and tall2, we observed average block sizes of 160.9 bp and 160.97 bp, respectively (Figure 5B).
463
464 To supplement our haplotype investigation, we looked at LD across putatively selected regions
465 in the different subpopulations. A pairwise LD heatmap indicates strong LD across the entire
466 region from 9.437 to 10.457 Mb in the subpopulations selected for short plant height (Figure
467 5C). We observed a low LD across the entire region from 9.437 to 10.457 Mb in the
468 subpopulations selected for tall plant height (Figure 5C).
469

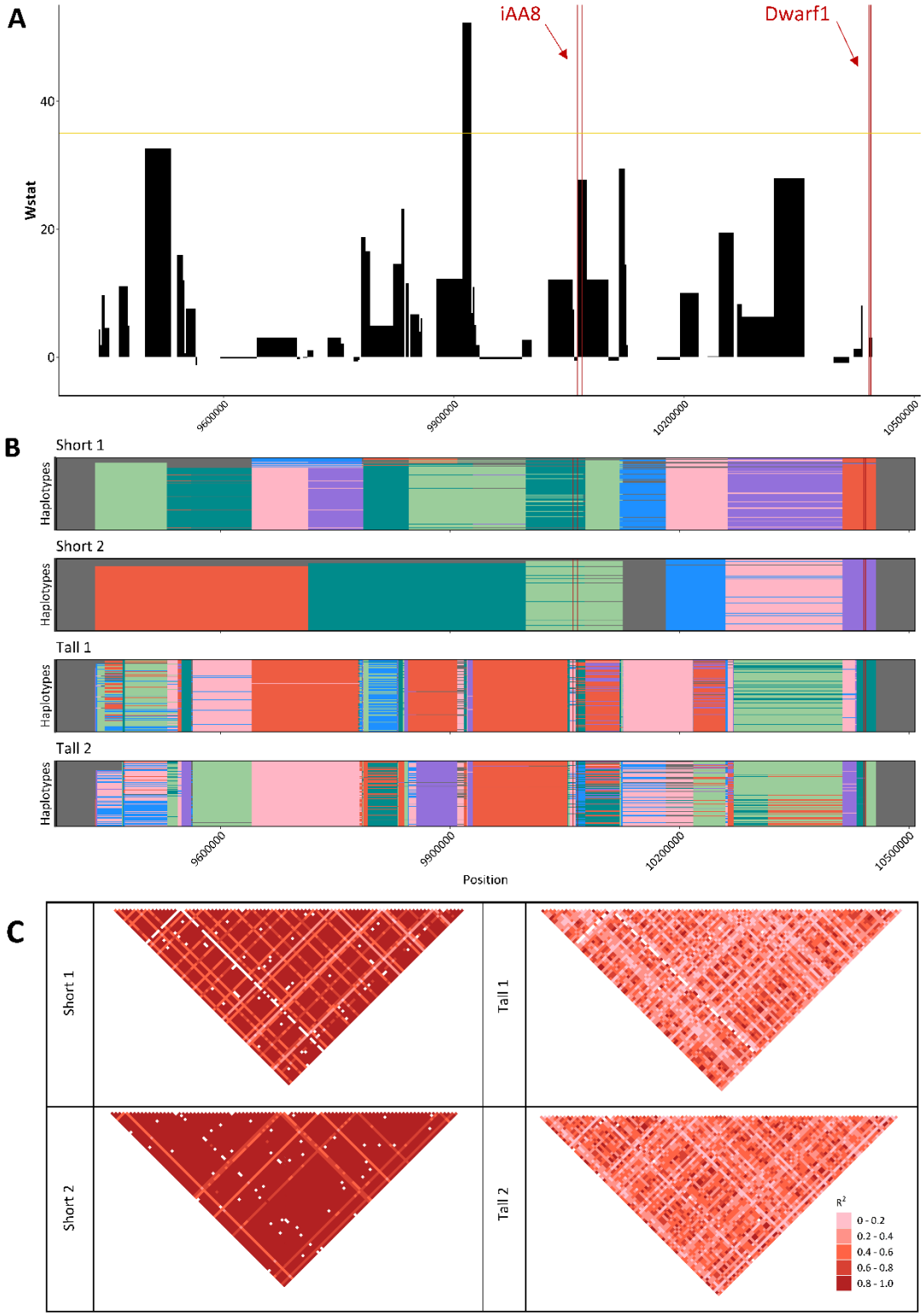


Figure 5: The region on chromosome 3 (from 9.437 Mb to 10.457 Mb) with the window boundary-based regions and their corresponding *Wstat* (A), the haplotype blocks (B) and pairwise LD heatmap for the different subpopulations(C).

Discussion

In contrast to most agriculture long-term selection experiments, our study included replicated selection populations. This allowed us to directly compare the commonly used approaches to significance threshold specification leveraging replications. Our results demonstrate that significance thresholds specified using evolutionary replicates can widely vary from thresholds based on empirical outliers or simulations.

In our study we used $F_{ST}Sum$ values to specify significance thresholds, which can only be calculated when replicates are available. If we had used conventional F_{ST} without leveraging replications, ~2,733 and ~274 would have passed the outlier thresholds of the 99.9th and 99.99th percentile. This is 2-fold the number of markers compared to thresholds calculated on $F_{ST}Sum$. In some studies, empirical outlier thresholds as low as the 99th percentile have been used (Wang *et al.*, 2018), which may comprise many false positives. Overall in this experiment, the number of markers identified using the FDRfS approach corresponds to the 99.99891st percentile of the empirical distribution. This percentile of the empirical distribution is much higher than the used percentiles of the empirical distribution used in most other studies (Wang *et al.*, 2018; Ayalew *et al.*, 2020).

FDRfS based on $F_{ST}Sum$ is mathematically equivalent to calculating an FDRfS value based on mean F_{ST} values. However, $F_{ST}Sum$ can be plotted in a much more interpretable way, so it is the statistic we chose to work with.

497 Despite our best efforts to include all relevant factors in our drift simulations, including
498 biological factors like mating scheme, setting initial allele frequencies based on the observed
499 allele frequency distribution, and including technical factors such as the sequencing procedure,
500 we were not able to identify a significance threshold which excluded false positive observations
501 (supplementary figure S6). Significance thresholds based on the FDRfS applied to $F_{ST}Sum$, on the
502 other hand, identified fewer markers than outlier-based or simulation-based significance
503 thresholds. Because these thresholds are based on replications, markers exceeding them are
504 expected to correspond to true signals of selection.
505
506 Previous studies applied significance thresholds based on the FDRfS on statistics like Chi-Square
507 (Linder *et al.*, 2021) and allele frequency differences (Turner and Miller, 2012). We adapted the
508 calculation of the FDRfS to $F_{ST}Sum$, a version of F_{ST} which leverages replications. For comparison,
509 we also explored calculating FDRfS from allele frequency differences (DiffStat) as was done by
510 Turner and Miller (2012). The results are available in supplementary figures S5 and S7. We
511 identified more significant markers using the $F_{ST}Sum$ approach at an FDRfS <5%. This was also
512 observed by Kofler and Schlötterer (2014).
513
514 Most of the 44 markers identified based on $F_{ST}Sum$ correspond to same region under selection.
515 On chromosome 3, we found 34 significant markers, all of which were located in a region
516 ranging from 9.437 to 10.457 Mb. All 34 markers show very similar allele frequency changes.
517 Because linkage leads to selection operating on regions instead of individual markers, we used
518 the R package GenWin (Beissinger *et al*., 2015) to calculate *Wstat* values based on a cubic

519 smoothing spline applied to $F_{ST}Sum$. This allowed us to define window boundaries of regions
520 which were putatively under selection. When we compared *Wstat* to $F_{ST}Sum$ values, we
521 observed a few significant regions, rather than many significant markers within regions.
522 Furthermore, we observed smaller differences between outlier-based and replication-based
523 significance thresholds when we calculated the thresholds for *Wstat*. The region identified by
524 *Wstat* on chromosome 3 from 9.911 to 9.923 Mb was very close to a marker from previous
525 study by Gyawali *et al.* (2019).
526
527 The plant height candidate gene *iAA8* is located within the putatively selected region from 9.437
528 to 10.457 Mb on chromosome 3 (Ramos Báez *et al.*, 2020). The plant height gene *Dwarf1* is also
529 located in this region (Teng *et al.*, 2013; Peiffer *et al.*, 2014; Liu, Fernie and Yan, 2020).
530
531 *Dwarf1* is involved in the expression of gibberellin 3-oxidase (Teng *et al.*, 2013). High expression
532 of the gene leads to high levels of bioactive gibberellin and results in taller plants (Teng *et al.*,
533 2013). When the *Dwarf1* gene is not expressed, plants exhibit the *Dwarf1* phenotype, which is
534 characterized by dwarfism (Teng *et al.*, 2013). Multiple gene expression studies have previously
535 reported that this gene has a large effect on plant height (Teng *et al.*, 2013; Liu, Fernie and Yan,
536 2020). In one experiment, this gene explained 32.3% of the phenotypic variance for plant height
537 (Teng *et al.*, 2013). The *iAA8* gene regulates auxin biosynthesis, transport, and signaling
538 (Matthes *et al.*, 2019; Ramos Báez *et al.*, 2020). Auxin is a plant hormone and plays a key role in
539 plant growth and development (Ramos Báez *et al.*, 2020). The *iAA8* gene was also observed to
540 regulate maize inflorescence architecture (Galli *et al.*, 2015; Ramos Báez *et al.*, 2020). Both

541 genes appear to be causal for plant height differences based on the literature (Teng *et al.*, 2013;
542 Peiffer *et al.*, 2014; Liu, Fernie and Yan, 2020; Ramos Báez *et al.*, 2020). Therefore, these
543 candidate genes were used as ground truth to test the method.
544
545 In the haplotype block analysis, all alleles located in the coding region of *Dwarf1* show limited
546 genetic variation in all subpopulations except subpopulation tall 2. In subpopulation tall 2, the
547 remaining variation appears to be limited. Within and around 10 kb of the coding region for the
548 gene *iAA8*, we observe more genetic variation in the subpopulations selected for tall plants than
549 in the subpopulations selected for short plants. Overall, we observe larger blocks and
550 consequently fewer blocks in our candidate gene region in the subpopulations selected for short
551 plants. The large differences in block sizes and number of blocks may indicate stronger selection
552 in the subpopulations selected for short plants.
553
554 This could theoretically be a result of more inbreeding in the short subpopulations, which we
555 tested by evaluating haplotype blocks in 39 random regions of the genome. The random regions
556 were of equal size and marker density to the candidate gene region. We observe on average 31
557 blocks in short 1, 30 blocks in short 2, 31 blocks in tall 1 and 34 blocks in tall 2. The blocks have
558 an average block size of 402.8 bp in short 1, 427.36 bp in short 2, 409.01 bp in tall 1 and 404.76
559 bp in tall 2. This indicates that the tall subpopulations show similar levels of inbreeding to the
560 short populations. Results from the calculated haplotype blocks from the 39 random regions are
561 available in supplementary figure S8 and S9. Exemplary haplotype block figures from two
562 random regions are available in supplementary figures S10 and S11.

563

564 Overall, a slow decline in LD was observed with increasing physical distance across all
565 chromosomes (Supplemental S12). R^2 of 0.1 was observed at a distance of 2 Mb (Supplemental
566 S12). But when we consider the subpopulations separately in the pairwise LD heatmap, we
567 observe very different results (Figure 5C). In the pairwise LD heatmap calculated for each
568 subpopulation, strong LD was observed in the subpopulations short 1 and short 2 across the
569 entire region ranging from 9.437 to 10.457 Mb (Figure 5C). In the subpopulations tall 1 and tall
570 2, we observe low LD in the entire region (Figure 5C).

571

572 The observation of strong LD in the short plant populations supports the argument that strong
573 selection may act on the entire candidate gene region in theses populations (Franssen, Barton
574 and Schlötterer, 2016; Barghi and Schlötterer, 2019). The strength of selection and the
575 experimental design influence hitchhiking (Baldwin-Brown, Long and Thornton, 2014; Kofler and
576 Schlötterer, 2014). We applied a strong selective regime in our experiment in comparison to
577 previous studies (Kofler and Schlötterer, 2014; Long *et al*., 2015). Strong selection leads to more
578 hitchhiking and a lower mapping resolution (Kofler and Schlötterer, 2014), which we observe in
579 the short subpopulations. However, strong LD may also persist due to more inbreeding in the
580 short subpopulations. But we only observe strong LD in the short subpopulations in our
581 candidate gene region. When we calculated the pairwise LD heatmap for the different
582 subpopulations in two random regions (supplementary figures S10 and S11), there was no
583 similarly strong difference in LD between the subpopulations as we observed for our candidate
584 gene region (Figure 5)

585

586 It is not only the strength of selection that can influence the number of variants selected in
587 experimental evolution studies. The number of selected variants also depends highly on
588 experimental design factors such as the population size, the duration of the experiment, the
589 applied selective regime (Kofler and Schlötterer, 2014; Kessner and Novembre, 2015), and the
590 species (Tobler *et al*., 2014; Barghi *et al.*, 2017). Divergent selection can increase the power of
591 the experiment in comparison to a single direction experiment (Kessner and Novembre, 2015).
592 We observe differences regarding the resolution between the subpopulations selected in
593 divergent directions. Even though the same strength of selection is applied in the divergently
594 selected subpopulations, differences between the subpopulations can be observed (Moose,
595 Dudley and Rocheford, 2004). But, these differences indicate that there is an increase in power
596 due to divergent selection (Kessner and Novembre, 2015).

597

598 Previous simulation studies investigating experimental evolution have demonstrated that
599 replication is one of the most powerful tools to increase power (Kofler and Schlötterer, 2014). In
600 fact, Kofler and Schlötterer (2014) observed that replication is more important for detecting
601 variants than population size (Kofler and Schlötterer, 2014; Kessner and Novembre, 2015).
602 Regions under selection might become smaller due to more recombination over time, and
603 subsequently we may find selection on additional regions as generations progress. This
604 phenomenon is visible in our "Quarter-Sauron" and "Sauron" plots (supplemental file S4)
605 (supplemental file S4). It was observed that the blue points do not appear completely random.
606 We observed that 682 markers diverged at F_{ST} of 0.3 between the subpopulations selected for

607 short plant height. There was no divergence shown in the subpopulations selected for tall plant
608 height at the same positions (supplemental file S4). At these markers, the frequency of the
609 reference allele was much higher in the subpopulation short 1. This could have been caused by
610 selection for specific environmental conditions (Dudley, 2007; Beissinger *et al.,* 2014). The
611 clustering of markers could have been caused by genetic hitchhiking of linked markers (Tobler *et*
612 *al*., 2014; Michalak *et al.*, 2018; Barghi and Schlötterer, 2019).
613

614 **Conclusion**

615 In conclusion, our research demonstrates that replicated selection is a powerful tool in
616 agricultural experimental evolution studies. We observed selection across replicates, even
617 though we only selected for three generations. Moreover, we saw that a significance threshold
618 based on the FDRfS is more reliable than existing methods for significance threshold
619 identification. The application of a cubic smoothing spline to single marker-based statistics like
620 $F_{ST}Sum$ to return window boundaries of regions which are putatively under selection helped
621 collapse signal from multiple hitchhiking markers. Haplotype block analysis additionally helped
622 to investigate the regions putatively under selection. We observe very different patterns of
623 selection in divergently selected subpopulations, which might be due to selection. After more
624 generations of selection, and a reduction of linkage between putatively selected loci and neutral
625 loci, we may draw more certain conclusions about selection for the plant height candidate
626 genes. Furthermore, we might be able to identify significant markers close to our plant height
627 candidate genes and novel gene loci without any associated gene models.
628

629

630 **Availability of data and materials**

631 All the data and statistics about the present study has been included in the current manuscript
632 in the form of figure and tables. Raw data are publicly available at figshare;
633 https://figshare.com/projects/Shoepeg_ExpEvo2020/139981. Our analysis code is available
634 publicly on github: https://github.com/milaleonie/ExpEvo_with_replicated_selection.

635

636 **Acknowledgements**

637 Abiskar Gyawali contributed substantially to this study but died before submission. Journal
638 policy prohibits him from being listed as a co-author because he could not approve the final
639 manuscript. We are grateful for Abi's contributions and miss him tremendously. Kate Guill was
640 instrumental in conducting early-generation seed increase and field management in Missouri.

641

642 **Funding**

643 This work was supported by startup funds provided by the University of Goettingen. The first
644 two generations of selection were funded by the USDA Agricultural Research Service. We
645 utilized the computational resources of the University of Goettingen's GWDG.

646

647 **Conflict of interest**

648 The author declares that there is no conflict of interest.

References

Aggrey S, Ankra-Badu GA, and Marks H. Effect of long-term divergent selection on growth characteristics in Japanese quail. Poultry Science. 2003; 82(4): 538–542.

Akey JM, G Zhang, K Zhang, L Jin, and MD Shriver. Interrogating a High-Density SNP Map for Signatures of Natural Selection. Genome Research. 2002;12(12): 1805–1814.

Akey JM. Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Research. 2009;19(5): 711–722.

Allard RW. Principles of plant breeding. 2nd ed. New York: J. Wiley. 1999.

Angert AL, Bradshaw HD Jr., and Schemske DW. Using Experimental Evolution to Investigate Geographic Range Limits in Monkeyflowers. Evolution. 2008;62(10): 2660–2675.

Ayalew H, ME Sorrells, BF Carver, PS Baenziger, XF Ma. Selection signatures across seven decades of hard winter wheat breeding in the Great Plains of the United States. The Plant Genome. 2020;13(3).

Baldwin-Brown JG, A Long and KR Thornton. The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. Molecular Biology and Evolution. 2014; 31(4): 1040-55.

Barghi N, R Tobler, V Nolte, C Schlötterer. Drosophila simulans: A Species with Improved Resolution in Evolve and Resequence Studies. G3 Genes|Genomes|Genetics. 2017; 7(7): 2337–2343.

Barghi N, R Tobler, V Nolte, AM Jakšić, F Mallard, KA Otte, M Dolezal,T Taus, R Kofler,C Schlötterer. Genetic redundancy fuels polygenic adaptation in Drosophila. PLOS Biology. 2019; 17(2): 1–31.

Barghi N, C Schlötterer. Shifting the paradigm in Evolve and Resequence studies: From analysis of single nucleotide polymorphisms to selected haplotype blocks. Molecular Ecology. 2019; 28: 521–524.

Barghi N, J Hermisson, C Schlötterer. Polygenic adaptation: a unifying framework to understand positive selection. Nature reviews. 2020; 21: 769–781.

Barrick JE, DS Yu, SH Yoon, H Jeong, TK Oh, D Schneider, RE Lenski, JF Kim. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature. 2009; 461(7268): 1243–1247.

Beissinger TM, GJM Rosa, SM Kaeppler, D Gianola, N de Leon. Defining window-boundaries for genomic analyses using smoothing spline techniques. Genetics Selection Evolution. 2015; 47(30):1-9.

Beissinger TM, CN Hirsch, B Vaillancourt, S Deshpande, K Barry, CR Buell, SM Kaeppler, D Gianola, N de Leon. A Genome-Wide Scan for Evidence of Selection in a Maize Population Under Long-Term Artificial Selection for Ear Number. Genetics. 2014; 196(3): 829–840.

Burke MK, JP Dunham, P Shahrestani, KR Thornton, MR Rose, AD Long. Genome-wide analysis of a long-term evolution experiment with Drosophila. Nature. 2010; 467(7315): 587–590.

Burke MK. How does adaptation sweep through the genome? Insights from long-termselection experiments. Proceedings of the Royal Society B: Biological Sciences. 2012; 279(1749): 5029–5038.

684 Burke MK, Liti G, Long AD. Standing Genetic Variation Drives Repeatable Experimental Evolution in
685 Outcrossing Populations of Saccharomyces cerevisiae. Molecular Biology and Evolution. 2014;31(12):
686 3228–3239.

687 Bradbury PJ, ZW Zhang, DE Kroon, TM Casstevens, Y Ramdoss, ES Buckler. TASSEL: software for
688 association mapping of complex traits in diverse samples. Bioinformatics. 2007; 23(19):2633-2635.

689 Cesconeto RJ, S Joost, CM McManus, SR Paiva, JA Cobuci, J Braccini. Landscape genomic approach to
690 detect selection signatures in locally adapted Brazilian swine genetic groups. Ecology and Evolution.
691 2017; 7(22): 9544–9556.

692 Crow JF, M Kimura. An introduction to population genetics theory. 1970. Harper & Row Publishers.

693 Danecek P, JK Bonfield, J Liddle, J Marshall, V Ohan, MO Pollard, A Whitwham, T Keane, SA McCarthy, RM
694 Davies, H Li. Twelve years of SAMtools and BCFtools. GigaScience. 2021; 10(2): giab008.

695 Deardorff ER, KA Fitzpatrick, GVS Jerzak, PY Shi, LD Kramer, GD Ebel. West Nile Virus Experimental
696 Evolution in vivo and the Trade-off Hypothesis. PLoS Pathogens. 2011; 7(11): e1002335.

697 Doebley J, JD Wendel, JSC Smith, CW Stuber, MM Goodman. The origin of cornbelt maize: the isozyme
698 evidence. Economic Botany. 1988; 42(1): 120 – 131.

699 Dudley JW, RJ Lambert. 100 Generations of Selection for Oil and Protein in Corn. 2003. Wiley & Sons.

700 Elshire RJ, JC Glaubitz, Q Sun, JA Poland, K Kawamoto, ES Buckler, SE Mitchell. A Robust, Simple
701 Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE. 2011; 6(5), p. e19379.

702 Firman RC, F Garcia-Gonzalez, E Thyer, S Wheeler, Z Yamin, M Yuan, LW Simmons. Evolutionary change
703 in testes tissue composition among experimental populations of house mice. Evolution. 2015; 69(3):
704 848–855.

705 Firman RC, LW Simmons. Experimental evolution of sperm quality via postcopulatory sexual selection in
706 house mice. Evolution. 2009; 64-5: 1245–1256.

707 Fox AR, F Scochera, T Laloux, K Filik, H Degand, P Morsomme, K Alleva, F Chaumont. Plasma membrane
708 aquaporins interact with the endoplasmic reticulum resident VAP27 proteins at ER–PM contact sites
709 andendocytic structures. New Phytologist. 2020; 228:973-988.

710 Franssen SU, NH Barton, C Schlötterer. Reconstruction of Haplotype-Blocks Selected during Experimental
711 Evolution. Molecular Biology and Evolution. 2016; 34(1): 174–184.
712
713 Galli M, Q Liua, B L. Moss, S Malcomber, W Lia, C Gainese, S Federicia, J Roshkovana, R Meeley, JL
714 Nemhauserb, A Gallavottia. Auxin signaling modules regulate maize inflorescence architecture.
715 Proceedings of the National Academy of Sciences. 2015; 112(43): 13372–13377.

716 Gyawali A, V Shrestha, KE Guill, S Flint-Garcia, TM Beissinger. Single-plant GWAS coupled with bulk
717 segregant analysis allows rapid identification and corroboration of plant-height candidate SNPs. BMC
718 Plant Biology. 2019; 19(1):412.

719 Herring CD, A Raghunathan, C Honisch, T Patel, M Kenyon Applebee, AR Joyce, TJ Albert, FR Blattner, D
720 van den Boom, CR Cantor, BØ Palsson. Comparative genome sequencing of Escherichia coli allows
721 observation of bacterial evolution on a laboratory timescale. Nature Genetics. 2006; 38(12): 1406–1412.

722 Hirsch CN, SA Flint-Garcia, TM Beissinger, SR Eichten, S Deshpande, K Barry, MD McMullen, JB Holland,
723 ES Buckler, N Springer, CR Buell, N de Leon, SM Kaeppler. Insights into the Effects of Long-Term Artificial
724 Selection on Seed Size in Maize. Genetics. 2014;198(1): 409–421.

725 Janick, J. Plant Breeding Reviews: Long-Term Selection: Maize, Volume 24, Part 1. 2010a. John Wiley &
726 Sons, Inc.

727 Janick, J. (2010b) Plant Breeding Reviews, Volume 24, Part 2 Long-term Selection: Crops, Animals, and
728 Bacteria. 2010b. John Wiley & Sons, Inc.

729 Johansson AM, ME Pettersson, PB Siegel, Ö Carlborg. Genome-Wide Effects of Long-Term Divergent
730 Selection. PLoS Genetics. 2010; 6(11), e1001188.

731 Kelly SA, PP Czech, JT Wight, KM Blank, T Garland Jr. Experimental evolution and phenotypic plasticity of
732 hindlimb bones in high-activity house mice. Journal of Morphology. 2006; 267(3): 360–374.

733 Kessner D, J Novembre. Power Analysis of Artificial Selection Experiments Using Efficient Whole Genome
734 Simulation of Quantitative Traits. Genetics. Genetics, 2015; 199, 991–1005.

735 Kofler R, C Schlötterer. A Guide for the Design of Evolve and Resequencing Studies. Molecular Biology
736 and Evolution. 2014;31(2): 474–483.

737 Krimbas CB and S Tsakas. The Genetics of Dacus oleae. V. Changes of Esterase Polymorphism in a Natural
738 Population Following Insecticide Control-Selection or Drift? Evolution. 1971; 25(3): 454–460.

739 Kumar R, A Gyawali, GD Morrison, CA Saski, DJ Robertson, DD Cook, N Tharayil, RJ Schaefer, TM
740 Beissinger, RS Sekhon. Genetic Architecture of Maize Rind Strength Revealed by the Analysis of
741 Divergently Selected Populations. Plant and Cell Physiology. 2021; 62(7): 1199–1214.

742 Lenski RE. Experimental evolution and the dynamics of adaptation and genome evolution in microbial
743 populations. The ISME Journal. 2017; 11(10): 2181–2194.

744 Levy SF, JR Blundell, S Venkataram, DA Petrov, DS Fisher, G Sherlock. Quantitative evolutionary dynamics
745 using high-resolution lineage tracking. Nature. 2015; 519(7542): 181–186.

746 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013.

747 Linder RA, B Zabanavar, A Majumder, H Chiao-Shyan Hoang, VG Delgado, R Tran, VT La, SW Leemans, AD
748 Long. Adaptation in outbred sexual yeast is repeatable, polygenic, and favors rare haplotypes. Molecular
749 Biology and Evolution, 2022; 39(12): 1–25.

750 Liu J, AR Fernie, J Yan. The Past, Present, and Future of Maize Improvement: Domestication, Genomics,
751 and Functional Genomic Routes toward Crop Enhancement. Plant Communications. 2020;1(1),100010.

752 Long A, G Liti, A Luptak, O Tenaillon. Elucidating the molecular architecture of adaptation via evolve and
753 resequence experiments. Nature Reviews Genetics. 2015; 16(10): 567–582.

754 Lorenz AJ, TM Beissinger, R Rodrigues Silva, N de Leon. Selection for Silage Yield and Composition Did
755 Not Affect Genomic Diversity Within the Wisconsin Quality Synthetic Maize Population.G3
756 Genes|Genomes|Genetics. 2015; 5(4): 541–549.

757 Lush JL. Animal Breeding Plans. Ames, Ia., Collegiate Press. 1937.

758 Ma Y, X Ding, S Qanbari, S Weigend, Q Zhang, H Simianer. Properties of different selection signature
759 statistics and a new strategy for combining them. Heredity. 2015; 115(5): 426–436.

760 Maize GDB. Maize genetics and genomics database. 2021. www.maizegdb.org/.

761 Marks HL. Long-Term Selection for Body Weight in Japanese Quail Under Different Environments. Poultry
762 Science. 1996; 75(10): 1198–1203.

763 Matthes MS, NB Best, JM Robil, S Malcomber, A Gallavotti, P McSteen. Auxin EvoDevo: Conservation and
764 Diversification of Genes Regulating Auxin Biosynthesis, Transport, and Signaling. Molecular Plant. 2019;
765 12(3): 298–320.

766 Mazaheri M, M Heckwolf, B Vaillancourt, JL Gage, B Burdo, S Heckwolf, K Barry, A Lipzen, C Bastos
767 Ribeiro, TJY Kono, HF. Kaeppler, EP. Spalding, CN Hirsch, CR Buell, N de Leon, SM. Kaeppler. Genome-
768 wide association analysis of stalk biomass and anatomical traits in maize. BMC Plant Biology. 2019;
769 19(45): 1–19.

770 Moose SP, JW Dudley, TR Rocheford. Maize selection passes the century mark: a unique resource for
771 21st century genomics. Trends in plant science. 2014; 9: 358–364.

772 Michalak P, L Kang, MF Schou, HR Garner, V Loeschcke. Genomic signatures of experimental adaptive
773 radiation in Drosophila. Molecular Ecology. 2019; 28: 600–614.

774 Nieh SC, WS Lin, YH Hsu, GJ Shieh, BJ Kuo. The effect of flowering time and distance between pollen
775 source and recipient on maize. GM Crops & Food. 2014; 5(4): 287–295.

776 Orozco-terwengel P, M Kapun, V Nolte, R Kofler, T Flatt, C Schlötterer. Adaptation of Drosophila to a
777 novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles.
778 Molecular Ecology. 2012; 21: 4931-4941.
779
780 Parts L, FA Cubillos, J Warringer, K Jain, F Salinas, SJ Bumpstead, M Molin, A Zia, JT Simpson, MA Quail, A
781 Moses, EJ Louis, R Durbin, G Liti. Revealing the genetic structure of a trait by sequencing a population
782 under selection. Genome Research. 2011; 21(7): 1131–1138.
783
784 Payen C, AB Sunshine, GT Ong, JL Pogachar, W Zhao, MJ Dunham. High-Throughput Identification of
785 Adaptive Mutations in Experimentally Evolved Yeast Populations. PLOS Genetics. 2016; 12(10), p.
786 e1006339.

787 Peiffer JA, MC Romay, MA Gore, SA Flint-Garcia, Z Zhang, MJ Millard, CAC Gardner, MD McMullen, JB
788 Holland, PJ Bradbury, ES Buckler. The genetic architecture of maize height. Genetics. 2014; 196(4): 1337–
789 1356.

790 Phillips MA, AD Long, ZS Greenspan, LF Greer, MK Burke, B Villeponteau, KC Matsagas, CL Rizza, LD
791 Mueller, MR Rose. Genome-wide analysis of long-term evolutionary domestication in Drosophila
792 melanogaster. Scientific Reports. 2016; 6: 39281.

793 Pook, T, M Schlather, G de los Campos, M Mayer, CC Schoen, H Simianer. HaploBlocker: Creation of
794 Subgroup-Specific Haplotype Blocks and Libraries. Genetics. 2019; 212(4): 1045–1061.

795 R Core Team. R: A Language and Environment for Statistical. 2021. https://CRAN.R-
796 project.org/package=doParallel.

797 Ramos Báez R, RY Buckley, H Yu, Z Chen, A Gallavotti, JL Nemhauser, BL Moss. A Synthetic Approach
798 Allows Rapid Characterization of the Maize Nuclear Auxin Response Circuit. Plant Physiology. 2020;
799 182(4): 1713–1722.

800 Remington DL, JM Thornsberry, Y Matsuoka, LM Wilson, SR Whitt, J Doebley, S Kresovich, MM Goodman,
801 ES Buckler. Structure of linkage disequilibrium and phenotypic associations in the maize genome. PNAS.
802 2001; 98(20): 11479–11484.

803 Remolina SC, PL Chang, J Leips, SV. Nuzhdin, KA Hughes. GENOMIC BASIS OF AGING AND LIFE-HISTORY
804 EVOLUTION IN DROSOPHILA MELANOGASTER. Evolution, 2012; 66: 3390-3403.

805 Romay MC, MJ Millard, JC Glaubitz, JA Peiffer, KL Swarts, TM Casstevens, RJ Elshire, CB Acharya, SE
806 Mitchell, SA Flint-Garcia, MD McMullen, JB Holland, ES Buckler, CA Gardner. Comprehensive genotyping
807 of the USA national maize inbred seed bank. Genome Biology. 2013; 14(R55).

808 Ross AJ, AR Hallauer, M Lee. Genetic analysis of traits correlated with maize ear length. Maydica. 2006;
809 51: 301–313.

810 Tange, O. GNU Parallel - The Command-Line Power Tool. The USENIX Magazine. 2011; 42-47.

811 Teng F, L Zhai, R Liu, W Bai, L Wang, D Huo, Y Tao, Y Zheng, Z Zhang. ZmGA3ox2, a candidate gene for a
812 major QTL, qPH3.1, for plant height in maize. The Plant Journal. 2013; 73(3): 405–416.

813 Teotónio H, IM Chelo, M Bradić, MR Rose, AD Long. Experimental evolution reveals natural selection on
814 standing genetic variation. Nature Genetics. 2009; 41(2): 251–257.

815 Teotónio H, MR Rose. Variation in the reversibility of evolution. Nature. 2000; 408(6811): 463–466.

816 Tobler R, SU Franssen, R Kofler, P Orozco-terWengel, V Nolte, Joachim Hermisson, C Schlötterer. Massive
817 Habitat-Specific Genomic Response in D. melanogaster Populations during Experimental Evolution in Hot
818 and Cold Environments. Molecular Biology and Evolution. 2014; 31(2): 364–375.

819 Turner TL, AD Stewart, AT Fields, WR Rice, AM Tarone. Population-Based Resequencing of Experimentally
820 Evolved Populations Reveals the Genetic Basis of Body Size Variation in Drosophila melanogaster. PLoS
821 Genetics. 2011; 7(3): e1001336.

822 Turner TL, Miller PM. Investigating Natural Variation in Drosophila Courtship Song by the Evolve and
823 Resequence Approach. Genetics. 2012; 191(2): 633–642.

824 Scheet P, M Stephens. A fast and flexible statistical model for large-scale population genotype data:
825 applications to inferring missing genotypes and haplotypic phase. American Journal of Human Genetics.
826 2006; 78(4): 629-644.

827 Silverman BW. Some Aspects of the Spline Smoothing Approach to Non-Parametric Regression Curve
828 Fitting. Journal of the Royal Statistical Society. 1985; 47(1): 1-52

829 Shin J, S Blay, B McNeney, J Graham. "LDheatmap: An R Function for Graphical Display of Pairwise
830 Linkage Disequilibria Between Single Nucleotide Polymorphisms." Journal of Statistical Software. 2006;
831 16(3).

832 U.S. National Plant Germplasm System. Details for: PI 269743, Zea mays L. subsp. mays, Shoepeg. 2021.
833 https://www.ars-grin.gov/.

834 Wallace JG, X Zhang, Y Beyene, K Semagn, M Olsen, BM Prasanna, ES Buckler. Genome-wide Association
835 for Plant Height and Flowering Time across 15 Tropical Maize Populations under Managed Drought
836 Stress and Well-Watered Conditions in Sub-Saharan Africa. Crop Science. 2016; 56(5): 2365–2378.

837 Wang K, P Wu, Q Yang, D Chen, J Zhou, A Jiang, J Ma, Q Tang, W Xiao, Y Jiang, L Zhu, X Li and G Tang.
838 Detection of Selection Signatures in Chinese Landrace and Yorkshire Pigs Based on Genotyping-by-
839 Sequencing Data. Frontiers in Genetics. 2018; 9:119.

840 Weir BS and CC Cockerham. Estimating F-Statistics for the Analysis of Population Structure. Evolution.
841 1984; 38(6): 1358– 1370.

842 Wichman HA, MR Badgett, LA Scott, CM Boulianne, JJ Bull. Different Trajectories of Parallel Evolution
843 During Viral Adaptation. Science. 1999; 285(5426): 422–424.

844 Wickland DP, G Battu, KA Hudson, BW Diers, ME Hudson. A comparison of genotyping-by-sequencing
845 analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. BMC
846 Bioinformatics. 2017; 18:586.
847

848