G3: Genes|Genomes|Genetics Early Online, published on February 2, 2015 as doi:10.1534/g3.114.015263

Title: Selection for silage yield and composition did not affect genomic diversity within the Wisconsin Quality Synthetic maize population

Aaron J. Lorenz[1], Timothy M. Beissinger[2], Renato Rodrigues Silva[3], Natalia de Leon[3]

1. Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583
2. Department of Plant Sciences, University of California, Davis, CA 95616
3. Department of Agronomy, University of Wisconsin, Madison, WI 53706

Short title: Selection mapping for maize silage traits

Key words: Genomic diversity, plant breeding, silage composition, association mapping

Corresponding author: Aaron J. Lorenz

Phone: 402-472-5361

Email: alorenz2@unl.edu

Abstract: Maize silage is a forage of high quality and yield, and represents the second most important use of maize in the U.S. The Wisconsin Quality Synthetic (WQS) maize population has undergone five cycles of recurrent selection for silage yield and composition resulting in a genetically improved population. The application of high-density molecular markers allows breeders and geneticists to identify important loci through association analysis and selection mapping, as well as monitor changes in the distribution of genetic diversity across the genome. The objectives of this study were: 1) Identify loci controlling variation for maize silage traits through association analysis and the assessment of selection signatures. 2) Describe changes in the genomic distribution of gene diversity through selection and genetic drift in the WQS recurrent selection program. We failed to find any significant marker-trait associations using the historical phenotypic data from WQS breeding trials combined with 17,719 high-quality, informative single nucleotide polymorphisms. Likewise, no strong genomic signatures were left by selection on silage yield and quality in the WQS despite genetic gain for these traits. These results could be due to the genetic complexity underlying these traits, or the role of selection on standing genetic variation. Variation in loss of diversity through drift was observed across the genome. Some large regions experienced much

greater loss in diversity than what is expected, suggesting limited recombination combined with small populations in recurrent selection programs could easily lead to fixation of large swaths of the genome.

Introduction

Silage production is the second most important use of maize in the U.S. following grain production (USDA-National Agricultural Statistics Service, 2014). With the expected increase in consumption of animal, especially dairy, products worldwide, as well as regulations currently in place in the U.S. and other countries in the world related to the need to increase the presence of biomass-derived biofuel production, improving forage yield and composition has become an area of substantial research and development.

Maize silage is a forage of high quality and yield (Coors and Lauer, 2001). A major difference between maize silage and other types of forage relates to the contribution of the grain, which represents approximately 50% of the total biomass in average temperate maize hybrids in the U.S. (Lorenz et al., 2010). Cell wall-bound carbohydrates provide another important energy source for ruminant animals. Breeding for silage production in maize, therefore, involves the simultaneous improvement of forage yield and cell wall composition. A substantial amount of work has been dedicated to establishing associations between the relative contribution of compositional properties -- digestibility, carbohydrate concentration and protein -- to animal productivity (Schwab et al., 2003). Summative equations, such as MILK2006 (Shaver et al., 2006), combine forage composition and yield to calculate expected milk per hectare, which can be directly used as a selection criterion.

The Wisconsin Quality Synthetic (WQS) maize population was developed by the University of Wisconsin maize breeding program nearly three decades ago and is currently in its fifth cycle of recurrent selection for high-quality stover and high forage

93 yield (Frey et al., 2004; Gustafson et al., 2010). Gustafson et al. (2010) evaluated forage
94 yield and composition for each cycle of WQS *per se* as well as topcrosses to two
95 commercial testers. Linear improvements were observed in whole plant yield, stover
96 yield, and whole-plant composition both in the population *per se* as well as topcross
97 evaluations. Although stover quality *per se* did not improve through selection, milk yield
98 on a Mg ha^{-1} basis has increased 24%. Changes in silage yield have been greater than
99 changes in silage composition, suggesting that the current selection protocol tends to
100 emphasize improvements in forage yield compared to composition (Gustafson et al.,
101 2010). Eight inbred lines have been released from the different cycles of this population
102 (W601S, W602S, W603S, W604S, W611S, W612S, W613S and W614S) and made
103 available to the public.

104

105 Recurrent selection in plant breeding is a cyclical process of evaluation, selection, and
106 recombination practiced within a closed population with the goal of improving the mean
107 population performance while maintaining genetic variation (Bernardo, 2010).
108 Maintaining genetic variation in a population undergoing recurrent selection is critical for
109 continued response to selection, but achieving an intensity of selection sufficient for
110 making genetic gain can be antagonistic to this goal. Studies reporting changes in average
111 diversity at the molecular marker level within maize recurrent selection programs have
112 been frequently performed and have found decreases in diversity in proportion to that
113 expected through genetic drift alone (Butruille et al., 2004; Romay et al., 2012; Labate et
114 al., 1999). Low marker densities prevented these studies from examining the distribution
115 of the genetic diversity across the genome. This is important to examine because diversity

116 in some genomic regions may have been maintained by chance, while diversity in other
117 regions may have been completely lost by chance through the fixation of large swaths of
118 the genome due to infrequent recombination. Only two to three crossovers per
119 chromosome are expected in maize (Anderson et al., 2003)

120 Dense genotyping of populations undergoing recurrent selection can also be used
121 for identifying signatures of selection, as has been performed with model organisms
122 (Parts, 2011; Turner, 2011), and agricultural species such as maize (Wright et al., 2005;
123 Hufford et al., 2012; Wisser et al., 2008; Coque et al., 2006; Falke et al., 2007; Hirsch et
124 al., 2014; Beissinger et al., 2014). When combined with genomic information, an array of
125 statistical methods, both widely recognized and recently proposed, hold great promise for
126 identifying genes underlying phenotypic response to selection and impacts of selection on
127 genomic structure (Lewontin and Krakauer, 1973; Barrett and Hoekstra, 2011). A
128 disadvantage of selection mapping stems from the fact that selection is often not
129 performed for a single trait, making it impossible to estimate effects of individual loci on
130 specific traits.

131 Association mapping is another option for identifying loci underlying variation
132 for traits of interest within breeding populations. A major setback of this approach,
133 however, is low power to detect rare alleles for populations of moderate size (Myles et
134 al., 2009). Another issue highly relevant to the application of association mapping to
135 populations undergoing recurrent selection is the fact that alleles conferring favorable
136 values for traits are expected to change in frequency through selection and thus contribute
137 to structure between the different cycles of selection. When population structure is
138 corrected for using a mixed linear model (Yu et al., 2006), power to detect these alleles

contributing to genetic differences between cycles is reduced (Rincent et al., 2014). Wisser et al. (2011) proposed combining selection mapping and association mapping to overcome deficiencies of both methods for dissecting the genetic architecture underlying response to selection.

Assessing the impact of recurrent selection on the distribution of diversity across the genome would further the understanding of how drift and selection shape genomic architecture. Moreover, identifying genomic regions influencing forage composition and yield would be beneficial to silage breeding. With this in mind, the objectives of this study were: 1) Identify loci controlling variation for maize silage traits through association analysis and the assessment of selection signatures. 2) Describe changes in the genomic distribution of gene diversity through selection and genetic drift in the WQS recurrent selection program. To accomplish this, individuals from multiple cycles of the WQS recurrent selection program were genotyped using a high-density SNP array. Phenotypic data was collated from historical records of the long-term WQS recurrent selection program.

Materials and Methods

Germplasm

Details on the formation of WQS can be found in Frey et al. (2004) and Gustafson et al. (2010). The breeding protocol utilized to advance WQS is depicted in Figure 1. Briefly, for cycles zero through three, between 400 and 500 S_1 families of WQS were initially screened for general agronomic suitability in a high-plant density replicated trial in South Central Wisconsin. The same S_1 families were simultaneously self-pollinated in

the breeding nursery. Approximately 50 to 67 percent of the S_1 families were discarded based on the stress trial. During the following season, $S_{1:2}$ families descended from random plants within selected S_1 families are crossed to testers belonging to the Stiff Stalk heterotic group. Resulting topcross hybrids are evaluated at two locations the following summer. Evaluations used standard field plot techniques for silage hybrids to estimate forage yield and composition (Frey et al., 2004; Gustafson et al., 2010). Details of the forage composition analysis are provided below. Advancing WQS from cycle four to cycle five involved the same procedures except 200 S_1 families were initially screened instead of 400-500.

After the fall harvest, the top 20 $S_{1:2}$ families presenting the highest milk-production index based on the MILK2006 prediction were selected. These $S_{1:2}$'s were recombined using the bulk-entry method whereby each selected progeny is crossed with each other selected progeny and each cross contributes equally to the next cycle of WQS. In this set up, population improvement and inbred development occured simultaneously as superior finished (S_6) lines were identified through the process of selfing, topcrossing and evaluating.

Phenotypic data

Starting in 1997, WQS silage yield trials were conducted to select the best 20 S_2 families for advancing the WQS to the next cycle. WQS C0, C1, C2, C3, and C4 were trialed in 1997, 2000, 2003, 2006, and 2010-11, respectively. All trials were planted at either West Madison Agricultural Research Station (Madison, WI; WMARS) or

184 Arlington Agricultural Research Station (Arlington, WI; AARS) or both. In 1997, a trial
185 of WQS C0 S_2 topcross families was planted at one location, WMARS, with four
186 replications. In 2000, a trial of WQS C1 S_2 topcross families was planted at two
187 locations, WMARS and AARS, with two replications per location. The WQS C2 and C3
188 S_2 topcross families were evaluated at WMARS and AARS using three replications at
189 each location in 2003 and 2006, respectively. However, the AARS location in 2006 was
190 abandoned because of a severe windstorm that caused extensive lodging. In 2010, a trial
191 of WQS C4 S_2 topcross families was planted at WMARS and AARS with two
192 replications, but data quality from AARS in 2010 was very poor and it was therefore
193 discarded. The WQS C4 trial was replanted at WMARS in 2011 using two replications in
194 order to provide an additional environment for evaluation. All trials consisted of two-row
195 plots, 6.08 m long and 0.76 m spacing between rows, arranged in a randomized complete
196 block design. Planting densities were common for silage production in the region.
197 Different testers were used across cycles. Testers LH119 was used in WQS C0; LH198 in
198 WQS C1; HC33 in WQS C2; and LH244 in WQS C3 and C4. All testers used are highly
199 related to B73.

200 Most recently, nutritional quality is evaluated using MILK2006, a summative
201 equation for calculating milk yield based on factors that affect whole-plant maize silage
202 feed quality including yield, dry matter (DM) content, neutral detergent fiber (NDF)
203 content, NDF digestibility (NDFD), protein, and starch (Schwab et al, 2003). Previous
204 versions of this summative equation were utilized in earlier cycles of the WQS selection
205 program. In MILK2006, as well as previous versions, each component (NDF, protein,
206 and starch) is weighted to take into account its respective digestibility. Starch and protein

digestibility are traditionally treated as constant, while the digestibility of the cell wall, or NDFD, is measured separately for each resulting hybrid. In vitro true digestibility (IVTD), acid detergent fiber (ADF), NDF, crude protein, and starch are predicted using a global near-infrared reflectance calibration developed in-house at the University of Wisconsin (https://cornbreeding.wisc.edu/nirs). Wet chemistry procedures to develop the calibration set are described elsewhere (Frey et al., 2004; Gustafson et al., 2010). The summative equation is then used to develop predictions of milk yield described as kg milk yield Mg^{-1} DM and kg milk yield ha^{-1}.

Genotyping

Remnant seed of available S_2 families from WQS C0 to WQS C4 and S_1 families from WQS C5 was germinated. Immature leaf tissue was collected from ten individual plants and pooled to represent each of the S_2 (for WQS C0 to WQS C4) and S_1 (in the case of WQS C5) families, respectively, selected at each cycle. Genomic DNA was extracted from each sample using a modified CTAB method (Saghai-Maroof *et al.* 1984). Samples were then genotyped using the Illumina MaizeSNP50 BeadChip, an Infinium HD assay with 56,110 SNP markers distributed across the maize genome (Ganal *et al.* 2011; Illumina, Inc. San Diego, CA). Alleles for each sample were called using the Genotyping Module within the Illumina Genome-Studio software. The built in GenCall data analysis software, which relies on the GenTrain clustering algorithm, was used for automatic clustering and calling of genotypes (Oliphant *et al.* 2002; Fan *et al.* 2003). In order to maintain only the highest quality SNPs, a GenCall threshold of 0.6 was used.

This filtering resulted in a dataset of 17,719 high-quality SNPs to be used for further analysis. The mean frequency of missing data was 0.07 with a range of 0 to 0.20. Of these markers, 15,646 were polymorphic, with polymorphic markers being defined as those with minor allele frequencies greater than 0.025. Missing marker scores were imputed using Beagle (Browning and Browning, 2009) implemented in the R package Synbreed (Wimmer et al., 2012). Imputation accuracy was defined as the mean posterior probability of the most likely genotypes and calculated using the *gprobsmetrics* utility in the Beagle package. The average imputation accuracy in this SNP dataset was greater than 99% for all chromosomes.

Analysis of phenotypic data

Data from different cycles was kept separate and the initial phenotypic data analysis was performed for each cycle separately. The following mixed linear model was fit to the phenotypic data

$$y_{ijk} = \mu + g_i + l_j + gl_{ij} + b_{k(j)} + \varepsilon_{ijk}$$

where y_{ijk} is the observation of the i^{th} family evaluated in the j^{th} environment in the k^{th} replication; μ is the intercept; g_i is the effect of the i^{th} family, l_j the effect of the j^{th} environment, gl_{ij} is the interaction between the i^{th} family and j^{th} environment, $b_{k(j)}$ is the effect of the k^{th} replicate nested within the j^{th} environment, and ε_{ijk} is the residual. Environment and replicate effects were modeled as fixed effects. Family and family-by-

environment interaction effects were modeled as random effects assumed to be independent and identically distributed. Variance components were estimated using restricted maximum likelihood and best linear unbiased predictions (BLUPs) for each trait were calculated for families. Each cycle was analyzed separately. All calculations were performed using the statistical analysis software ASReml-R (Butler et al., 2009).

Variance components were used to calculate broad-sense heritability (*H*) on a family-mean basis as $H = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{GE}^2/e + \sigma_\varepsilon^2/re}$, where σ_G^2 is the variance among families, σ_{GE}^2 is the variance due to family-by-environment interaction effects, σ_ε^2 is the residual variance, *e* is the number of environments, and *r* is the number of replications in each environment.

Genomic heritability

The proportion of variation among S_2 family BLUPs across cycles explained by the genomic relationship matrix was calculated. The genomic relationship matrix among all families was calculated as

$$\mathbf{G} = \frac{\mathbf{W_C W_C}'}{2\sum_l p_l q_l}$$

where $\mathbf{W_C}$ is the centered genotype matrix, and p_l and q_l are allele frequencies at the l^{th} locus (Endelman and Jannink, 2012). The following G-BLUP model was fit to the data

$$\hat{\mathbf{g}} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where $\hat{\mathbf{g}}$ is the vector of family BLUPs; **b** is a vector of fixed year effects (corresponding to selection cycle); **u** is a vector of random additive genetic values where $\mathbf{u} \sim MVN(0, \mathbf{G}\sigma_u^2)$; **e** is a vector of residuals; and **X** and **Z** are incidence matrices relating **b** and **u** to $\hat{\mathbf{g}}$, respectively. All calculations were made using ASReml-R (Butler et al., 2009) and the variance components σ_u^2 and σ_e^2 were estimated. Genomic heritability was calculated as $h_G^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$ (de los Campos et al., 2013).

Association mapping

A genome-wide association analysis for each trait was performed using the model

$$\hat{\mathbf{g}} = \mathbf{Xb} + \mathbf{Wm} + \mathbf{Zu} + \mathbf{e}$$

where $\hat{\mathbf{g}}$, **X**, **b**, **Z**, **u**, and **e** are as above; **m** is a vector of marker effects; and **W** is a matrix comprised of marker scores. The association analysis was implemented using EMMA (Kang et al., 2008). A statistical threshold of $P=10^{-4}$ was used to declare significant marker-trait associations. Because a preliminary analysis indicated no markers surpassed this threshold, no further effort was made to better define the statistical threshold to correct for multiple testing.

Selection mapping and gene diversity

Allele frequencies in WQS C2 and WQS C5 were calculated based on their maximum likelihood estimate, i.e. the observed number of copies of the minor allele divided by twice the number of individuals with an observed genotype. WQS C2 was utilized rather than WQS C0 or C1 because samples from the earlier cycles of selection did not include enough individuals for reliable estimates of allele frequencies. SNP-specific F_{ST} values based on a comparison of C2 and C5 were computed according to , where s^2 is the sample variance of allele frequency between populations, is the mean allele frequency across populations, and r is the number of populations (Weir and Cockerham, 1984).

Significance thresholds were determined via drift simulations of the demographic history of the WQS population, assuming linkage equilibrium between markers. Simulations were conducted within R (R Core Team, 2014). For each SNP in C2, a simulated allele frequency in C5 was created according to the WQS selection protocol, incorporating generations of selfing, crossing, evaluating, and recombining based on the precise number of individuals utilized at each step in the WQS program. Allele frequencies in C2 and C5 were used to calculate simulated F_{ST} values for each SNP. The maximum F_{ST} value observed across SNPs was recorded. This process was repeated 1,000 times. The 95% quantile of maximum F_{ST} values observed from simulations was taken as a simulated 95% significance threshold that accounts for multiple testing across all 17,590 SNPs. The R script used for simulations is provided (File S1).

The above simulations assumed linkage equilibrium. This is a conservative approach because it allows for more independent tests than may truly be appropriate: since SNPs are inherited in linked segments, the true number of independent loci may be lower than

the number of SNPs. To explore this possibility, the effective number of markers (M_{eff}), were computed with the *simpleM* software (Gao *et al.*, 2008). The above simulation strategy was again employed, but with the results of *simpleM* incorporated. To achieve this, the C2 starting population was simulated by sampling M_{eff} SNPs, where M_{eff} was obtained utilizing the parameters PCA_cutoff = 0.99 and 0.95. Ultimately, utilizing the M_{eff} SNPs instead of the total number of SNPs did not result in a substantive difference in the estimated significance threshold. Therefore, thresholds obtained via linkage equilibrium simulations were utilized throughout this experiment.

We also performed an enrichment analysis to assess if there is an excess of loci displaying a large change in allele frequency. This was achieved by using the previously described simulations to identify the expected 95% and 99% quantiles of F_{ST} over the course of the experiment. Then, the observed proportion of loci exceeding these quantiles was computed. Theoretically, 5% and 1% of loci will exceed these quantiles, assuming no selection.

Gene diversity (D; i.e., expected heterozygosity) was estimated for each SNP and for each selection cycle from WQS C2 through C5 using $\hat{D}_{lc} = 1 - (\hat{p}_{lc}^2 + \hat{q}_{lc}^2)$, where $\hat{D}_{lc}$ is the gene diversity estimate for the l^{th} SNP in the c^{th} selection cycle, $\hat{p}_{lc}$ is the allele frequency of that SNP, and $\hat{q}_{lc} = 1 - \hat{p}_{lc}$ (Weir, 1996).

Results

Association mapping

A total of 648 individuals from the WQS population were genotyped for this study. Most individuals came from WQS C2 to C5, while only 16 individuals were genotyped from WQS C0 and C1 (Table 1) because of germination problems most likely a result of seed source age. Both genotype and phenotype data were available on between 240 and 300 families from WQS C1 to C4, depending on the trait.

Trait means and ranges are displayed in Table 2 in order to provide an overview of the phenotypic data analyzed for the association analysis. The range in silage yield within a given cycle was, on average, 45% of the mean. On the other extreme, the range in IVTD within a given cycle was, on average, 8.5% of the mean. Broad-sense heritabilities on a family-mean basis for each cycle were mostly moderate to high (Table 3). An exception was the *H* for yield in WQS C2. Broad-sense heritability was generally higher for dry matter and starch, and lower for ADF and NDF. The h_G^2 within cycles was mostly low except for dry matter, CP in WQS C2, and starch WQS C4. The h_G^2 across cycles, was also low. This indicates that while the genotype accounts for a fair proportion of the phenotypic variation within a given cycle or year of evaluation, little of the variation is captured by an additive relationship matrix.

Using phenotypic data from historical field trials, an association analysis was performed to identify important genomic regions controlling variation for these yield and compositional traits. Unfortunately no statistically significant associations ($P < 10^{-4}$) were made for any trait (Figure S1).

Selection mapping

A wide range of F_{ST} values was observed between SNPs. Since the selection protocols and number of selected individuals at each generation were recorded throughout this experiment, this enabled a simulation-based approach for identifying significance thresholds for the boundaries of F_{ST} expected to result from drift alone. The 20 S_2 families selected within each cycle led to a strong bottleneck that the population repeatedly experienced, resulting in a high significance threshold. To obtain 95% and 99% probabilities of no false positives, respectively, significance thresholds were set at $F_{ST} = 0.743$ and 0.707.

There were no SNPs that exceeded these significance thresholds (Figure 2). Often, much more lenient outlier thresholds are utilized for selection-mapping experiments. Outlier thresholds involve comparing the observed data to its own empirical distribution, thereby guarantee that a specified proportion of the data are identified as candidates for selection. Utilizing a 99% outlier threshold in this study would have meant setting the significance threshold at $F_{ST} = 0.340$ and identifying 175 "significant" SNPs. Simulations demonstrated that this significance value is substantially lower than the effects of drift may allow.

Additionally, by evaluating drift simulations without accounting for multiple testing, we determined that it is expected that 5% and 1% of SNPs will exceed F_{ST} values of 0.214 and 0.328, respectively, due to drift alone. We used these values to assess whether or not there is enrichment for high-F_{ST} SNPs in the data. We observed that 6.139% and 1.137% of SNPs exceed these uncorrected thresholds, respectively, indicating there is little evidence of enrichment for SNPs displaying high F_{ST}.

375

376 *Reduction of gene diversity*

377 Despite no strong signatures of selection and marker-trait associations, an
378 examination of *D* for each locus shows that reductions were not uniform across the
379 genome (Figure 3). A large reduction in *D* was observed in regions on chromosomes 2
380 (~132 million bp), 3 (~55 million bp), and 4 (~78 million bp). These regions of relatively
381 greater loss in diversity were defined visually by examining the *D* plots in Figure 3.
382 Average *D* across all loci was reduced from 0.352 in C2 to 0.285 in C5. While average
383 genome-wide *D* was only reduced by 19 percent from C2 to C5, average *D* in these
384 regions on chromosomes 2, 3, and 4 was reduced by 62, 79, and 67 percent, respectively.
385 The large region on chromosome 2, for example, had an average *D* of 0.355 in C2, which
386 is very close to the average genome-wide *D* in C2. By C5, however, the average *D* was
387 only 0.135, which is well below one standard deviation of *D* (genome-wide SD = 0.131).

388

389 **Discussion**

390 The first objective of this study was to identify loci controlling variation for traits
391 important to silage breeding using a combination of association and selection mapping.
392 Despite moderate to high entry-mean heritabilities within cycles (Table 2) and
393 documented genetic gain for silage yield and composition in WQS (Frey et al., 2004;
394 Gustafson et al., 2010), no significant results were obtained.

395 The genetic complexity underlying silage quality and yield is antagonistic to
396 identifying loci contributing to variation and therefore selection response. While the

397 genetics underlying mechanisms involved in cell wall digestibility could be complex, the
398 dependence of silage quality on grain content, which is related to grain yield, surely
399 makes silage quality increasingly complex. Grain is highly digestible and accounts for
400 approximately 50% of total dry matter of silage (Coors and Lauer, 2001). Also, variation
401 in the effectiveness of the ear as a sink can influence stover composition through its
402 effect on dry matter partitioning and transport of sugars to the ear (Coors et al., 1997).
403 Stage of plant development at which plants are harvested contributes to variation in silage
404 quality (Jung and Casler, 2006). If genetic variation for time to maturity exists within a
405 population, this variation will be confounded with variation for stover quality. Finally,
406 plant components vary for digestibility and fiber concentrations, and genetic variation
407 exists for digestibility of specific plant components (Hansey et al., 2010). Therefore, the
408 genetic complexity of silage quality on a whole-plant basis could easily equal that of
409 grain yield given its dependence on grain yield and plant morphology.
410 Because starch content in silage and sink-source dynamics are important
411 contributors to quality, and genotype-by-environment (GxE) interactions are an important
412 source of variation for grain yield, it's not surprising that silage compositional traits are
413 highly influenced by GxE interactions, which has been observed in previous studies
414 (Argillier et al., 2000; Mechin et al., 2001). This source of variation reduces the
415 contribution of the genetic signal to the total variation, decreasing power to detect
416 marker-trait associations and selection signatures across years. On top of possible strong
417 GxE effects, epistatic interactions could reduce the contribution of main allelic effects,
418 and thus result in a loss of power for making associations. Although comparing variance
419 components and thus heritabilities is fraught with issues because of high standard errors,

examination of Table 3 shows that the proportion of variation accounted for by the additive genomic relationship matrix is low relative to the entry-mean broad-sense heritability in most cases. This suggests the importance of interactions underlying the variability for these traits, both epistatic interactions within cycles as well as GxE interactions across cycles. Another, confounded source of variation is allele-by-tester interactions. As noted in the *Methods*, different testers were used in the different cycles, opening up the possibility for tester interaction to dilute the main allelic effects. The testers used were all highly related, being B73 types, and therefore the importance of this source of variation is likely less than if unrelated testers were used.

We demonstrate that although significant genetic gain has been realized for important silage traits within WQS, no strong selection signature was left on the genome. There are at least two reasons for this finding. First, the genetic signal underlying variation for silage yield and composition is highly complex, likely comprised of many small main and interaction effects distributed across the entire genome. This hypothesis is supported by the lack of marker-trait associations found in this study. Second, it is possible that selection acted upon standing genetic variation caused by old mutations, meaning that a casual polymorphism is not necessarily associated with any particular haplotype. Such soft selective sweeps (Hermisson and Pennings, 2005) do not leave a strong selection signature and are hard to detect using molecular markers.

The lack of a strong selection signature found by this study is in good company among other similar findings on complex traits in agricultural species. Kemper et al. (2014) found little to no signature on the genome of cattle left by selection for milk yield, despite enormous genetic gain for this trait, and large differences between cattle breeds.

Likewise, selection for grain yield in maize has left only very subtle, if any, selection signatures (Gerke et al., 2014; van Heerwaarden et al., 2012). Once again, this is despite substantial genetic gain for grain yield accomplished within both a recurrent selection program (Gerke et al., 2014) and commercial breeding (van Heerwaarden et al., 2012).

Given that genetic gain has occurred (Frey et al., 2004; Gustafson et al., 2010), these observations indicate that the gain realized has been accomplished through subtle allele frequency shifts at many loci. On the one hand it's encouraging to know that breeders are able to simultaneously increase the frequency of many small-effect alleles, therefore achieving genetic gain on highly complex traits. On the other hand, great difficulty in getting to the bottom of causal mechanisms underlying genetic gain for complex traits limits our understanding of the genetics underlying selection response. It is clear that new and more powerful methods are required to identify signatures left by selection on highly polygenic traits. Researchers in population genomics have realized this and begun developing such methods (e.g., Berg and Coop, 2014).

Another implication of this study is that caution should be taken when using historical phenotypic data from recurrent selection programs for association mapping of complex traits. While we recognized our power was limited because of only modest population sizes (Table 1), we believed, based on the moderate to high *H*, the trait data from individual cycles was of high enough quality to detect marker-trait associations. Clearly that was a wrong assumption. Little of the phenotypic variance across cycles (and thus years) was additive genetic variance, with the majority likely being caused by genetic-by-year interactions given the complexity of the silage compositional traits and their interaction with grain yield. Our experience suggests that historical data is of limited

466 value for association genetics on complex traits prone to genotype-by-year interactions.
467 We recommend that all genotypes be re-evaluated across multiple years and locations to
468 maximize power for detecting associations. It is recognized that the dataset size used
469 herein is relatively small compared to some other historical datasets, and historical data
470 could be useful if vast quantities are available (Vaughn et al., 2014).
471 Recurrent selection is a systematic method to increase allele frequency of a base
472 population, and therefore increase the probability a superior inbred line is derived from
473 that base population through selection and inbreeding (Hallauer, 1990). Recurrent
474 selection in plant breeding generally involves the selection of multiple individuals or
475 families (typically 10 – 50) for recombination each cycle. A theoretical advantage of
476 recurrent selection compared to simple inbreeding and selection is that genetic variation
477 is maintained, leading to sustained genetic gain over time (Bernardo, 2010). Indeed,
478 response to recurrent selection has reportedly continued after many cycles (Dudley and
479 Lambert, 2004; Keeratinijakal and Lamkey, 1993; de Leon and Coors, 2002). Using
480 molecular markers, several studies on maize populations undergoing recurrent selection
481 have shown that observed average losses in gene diversity (i.e., expected heterozygosity)
482 are approximately equal to that expected by theory assuming genetic drift and a given
483 effective population size (Lamkey and Lorenz, 2014; Labate et al., 1999; Hinze et al.,
484 2005; Butruille et al., 2004; Romay et al., 2012). None of the aforementioned studies,
485 however, used marker densities great enough to observe variation in diversity loss across
486 the genome. By genotyping individuals from multiple cycles of selection of the WQS
487 with over 15,000 high-quality, informative SNPs, we were able to assess the degree to
488 which gene diversity reductions vary across the genome. Very few studies in maize have

examined the effects of recurrent selection using high density SNPs (Gerke et al., 2014; Beissinger et al., 2014; Hansey et al., 2014). While we observed that most loci followed expectations, a few genomic regions experienced substantial loss of diversity presumably through the combination of chance and the low number of crossovers occurring on each maize chromosome. A similar observation was made by Gerke et al. (2014). Using the same Illumina Infinium array, these authors observed that a number of large genomic regions within the BSSS/BSCB1 recurrent selection populations that became completely fixed for one haplotype after 16 cycles of selection. Based on the selection procedures used, it was difficult to determine if this was caused by drift or selection. It appears that the regions on chromosomes 2, 3, and 4 are headed for the same fate in the WQS recurrent selection program.

Given the erratic nature of drift in recurrent selection programs with relatively small effective population sizes, combined with the limited number of crossovers occurring on any given maize chromosome each generation, it is entirely possible for a population to become fixed for one haplotype across a large swath of genomic space. This means that while genome-wide diversity in a population may be seemingly satisfactory for continued progress, diversity within specific regions could be inadequate. If these regions harbor loci important for traits of interest, genetic gain would be compromised and the population would be prevented from reaching its full potential. A major advantage to the routine use of high-density markers in a breeding program would be the ability to monitor genomic variability in allelic diversity, and ultimately identifying any regions that would benefit from targeted injections of allelic diversity.

Conclusions

This is the first report of an analysis on genetic gain for silage yield and composition at the genomic level. No strong genomic signatures were left by selection on silage yield and quality in the WQS likely due to the complexity underlying these traits. The role of selection on standing genetic variation could also be contributing to the lack of strong signatures. Variation in loss of diversity through drift was observed across the genome. A few large regions experienced much greater loss in diversity than what is expected, indicating limited recombination and population sizes in recurrent selection programs could lead to fixation of large swaths of the genome.

Acknowledgements

This project is supported by the National Research Initiative or Agriculture and Food Research Initiative Competitive Grants Program grant no. # 2012-67013-19460 from the USDA National Institute of Food and Agriculture. T.M.B. was partially supported by a gift to the University of Wisconsin, Madison, Plant Breeding and Plant Genetics program from Monsanto as well as funding from the University of Wisconsin Graduate School. R.R was supported by a fellowship from CNPq – Brazil.

Figure Captions

Figure 1. Schematic of the selection protocol utilized to advance the Wisconsin Quality Synthetic (WQS) population. A second generation (S_2) topcross selection method is utilized to improve this germplasm. Inbreds derived from succeeding cycles of improvement are developed and released. Population improvement and inbred development occur simultaneously. The red oval highlights the approximately the 20 S_2 families that originate each subsequent cycle of selection.

Figure 2. F_{ST} values between WQS Cycle 2 and WQS Cycle 5, computed for each SNP. The dashed green line depicts an empirical 99% outlier threshold. Blue and red lines show simulation-based multiple-testing corrected significance thresholds, which control for the magnitude of drift that could reasonably be expected according to under the selection protocol that was employed.

Figure 3. Gene diversity for each SNP evaluated in the Wisconsin Quality Synthetic selection program from cycle 2 (WQS C2) to C5

Figure S1. Manhattan plots produced from an association analysis on silage compositional traits neutral detergent fiber (NDF), crude protein (CP), *in vitro* true digestibility (IVTD), and starch. Numbers along the *x*-axis represent the 10 chromosomes of maize. Dashed lines are placed at the arbitrary statistical threshold of $P = 10^{-4}$.

References

Anderson, L. K., G. G. Doyle, B. Brigham, J. Carter, K. D. Hooker *et al*, 2003 High-resolution crossover maps for each bivalent of *Zea mays* using recombination nodules. Genetics **165:** 849-865.

Argillier, O., V. Méchin and Y. Barriere, 2000 Inbred line evaluation and breeding for digestibility-related traits in forage maize. Crop Sci. **40:** 1596-1600.

Barrett, R. D., and H. E. Hoekstra, 2011 Molecular spandrels: Tests of adaptation at the genetic level. Nature Reviews Genetics **12:** 767-780.

Beissinger, T. M., C. N. Hirsch, B. Vaillancourt, S. Deshpande, K. Barry *et al*, 2014 A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. Genetics **196:** 829-840.

Berg, J. J., and G. Coop, 2014 A population genetic signal of polygenic adaptation. PLoS Genetics **10:** e1004412.

Bernardo, R., 2010 *Breeding for Quantitative Traits in Plants.* Stemma Press, Woodbury, MN.

Browning, B. L., and S. R. Browning, 2009 A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. The American Journal of Human Genetics **84:** 210-223.

Butler, D. G., B. R. Cullis, A. R. Gilmour and B. J. Gogel, 2009 *ASReml-R Reference Manual*. Queensland Dep. of Primary Industries, Brisbane, QLD, Australia.

Butruille, D. V., H. D. Silva, S. M. Kaeppler and J. G. Coors, 2004 Response to selection and genetic drift in three populations derived from the Golden Glow maize population. Crop Sci. **44:** 1527-1534.

Coors, J. G., and J. G. Lauer, 2001 Silage corn, pp. 347-392 in *Specialty Corns*, edited by A. R. Hallauer. CRC Press, Boca Raton, FL.

Coors, J., K. Albrecht and E. Bures, 1997 Ear-fill effects on yield and quality of silage corn. Crop Sci. **37:** 243-247.

Coque, M., and A. Gallais, 2006 Genomic regions involved in response to grain yield selection at high and low nitrogen fertilization in maize. Theor. Appl. Genet. **112:** 1205-1220.

de Leon, N., and J.G. Coors, 2002 Twenty-four cycles of mass selection for prolificacy in the Golden Glow maize population. Crop Sci. **42:** 325-333.

de los Campos, G., A. I. Vazquez, R. Fernando, Y. C. Klimentidis and D. Sorensen, 2013 Prediction of complex human traits using the genomic best linear unbiased predictor. PLoS Genetics **9:** e1003608.

Dudley, J., and R. Lambert, 2004 100 generations of selection for oil and protein in corn. Plant Breed. Rev. **24:** 79-110.

Endelman, J. B., and J. Jannink, 2012 Shrinkage estimation of the realized relationship matrix. G3-Genes Genomes Genetics **2:** 1405-1413.

Falke, K., C. Flachenecker, A. Melchinger, H. Piepho, H. Maurer *et al*, 2007 Temporal changes in allele frequencies in two European F2 flint maize populations under modified recurrent full-sib selection. Theor. Appl. Genet. **114:** 765-776.

Fan, J. B., A. Oliphant, R. Shen, B. G. Kermani, F. Garcia *et al*, 2003 Highly parallel SNP genotyping. Cold Spring Harb. Symp. Quant. Biol. **68:** 69-78.

Frey, T., J. Coors, R. Shaver, J. Lauer, D. Eilert *et al*, 2004 Selection for silage quality in the wisconsin quality synthetic and related maize populations. Crop Sci. **44:** 1200-1208.

607 Ganal, M. W., G. Durstewitz, A. Polley, A. Berard, E. S. Buckler *et al*, 2011 A large
608 maize (zea mays L.) SNP genotyping array: Development and germplasm
609 genotyping, and genetic mapping to compare with the B73 reference genome.
610 Plos One **6:** e28334.

611 Gao, X., J. Starmer and E. R. Martin, 2008 A multiple testing correction method for
612 genetic association studies using correlated single nucleotide polymorphisms.
613 Genet. Epidemiol. **32:** 361-369.

614 Gerke, J.P., J.W. Edwards, K.E. Guill, J. Ross-Ibarra, and M.D. McMullen The genomic
615 impacts of drift and selection for hybrid performance in maize. Genetics (in
616 press).

617 Gustafson, T., J. Coors and N. de Leon, 2010 Evaluation of S2-topcross selection for
618 maize (*Zea mays* L.) silage yield and quality in the Wisconsin Quality Synthetic
619 population. Crop Sci. **50:** 1795-1804.

620 Hallauer, A., 1990 Methods used in developing maize inbreds. Maydica **35:** 1-16.

621 Hansey, C. N., A. J. Lorenz and N. de Leon, 2010 Cell wall composition and ruminant
622 digestibility of various maize tissues across development. BioEnergy Research **3:**
623 295-304.

624 Hermisson, J., and P. S. Pennings, 2005 Soft sweeps: Molecular population genetics of
625 adaptation from standing genetic variation. Genetics **169:** 2335-2352.

626 Hinze, L. L., S. Kresovich, J. D. Nason and K. R. Lamkey, 2005 Population genetic
627 diversity in a maize reciprocal recurrent selection program. Crop Sci. **45:** 2435-
628 2442.

629 Hirsch, C. N., S. A. Flint-Garcia, T. M. Beissinger, S. R. Eichten, S. Deshpande *et al*,
630 2014 Insights into the effects of long-term artificial selection on seed size in
631 maize. Genetics **198:** 409-421.

632 Hufford, M. B., X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J. Chia *et al,* 2012
633 Comparative population genomics of maize domestication and improvement. Nat.
634 Genet. **44:** 808-811.

635 Jung, H., and M. Casler, 2006 Maize stem tissues: Impact of development on cell wall
636 degradability. Crop Sci. **46:** 1801-1809.

637 Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman *et al*, 2008 Efficient
638 control of population structure in model organism association mapping. Genetics
639 **178:** 1709-1723.

640 Keeratinijakal, V., and K. R. Lamkey, 1993 Genetic effects associated with reciprocal
641 recurrent selection in BSSS and BSCB1 maize populations. Crop Sci. **33:** 78-82.

642 Kemper, K. E., S. J. Saxton, S. Bolormaa, B. J. Hayes and M. E. Goddard, 2014 Selection
643 for complex traits leaves little or no classic signatures of selection. BMC
644 Genomics **15:** 246.

645 Labate, J. A., K. R. Lamkey, M. Lee and W. L. Woodman, 1999 Population genetics of
646 increased hybrid performance between two maize populations under reciprocal
647 recurrent selection, pp. 127-137 in *The Genetics and Exploitation of Heterosis in*
648 *Crops*, edited by J. G. Coors and S. Pandey. ASA-CSSA-SSSA, Madison, WI.

649 Lamkey, C., and A. Lorenz, 2014 Relative effect of drift and selection in diverging
650 populations within a reciprocal recurrent selection program. Crop Sci. **54:** 576-
651 585.

652 Lewontin R. C., Krakauer J. , 1973 Distribution of gene frequency as a test of the
653 theory of the selective neutrality of polymorphisms. Genetics **74**: 175–195.

654 Lorenz, A., T. Gustafson, J. Coors and N. d. Leon, 2010 Breeding maize for a
655 bioeconomy: A literature survey examining harvest index and stover yield and
656 their relationship to grain yield. Crop Sci. **50:** 1-12.

657 Méchin, V., O. Argillier, Y. Hébert, E. Guingo, L. Moreau *et al*, 2001 Genetic analysis
658 and QTL mapping of cell wall digestibility and lignification in silage maize. Crop
659 Sci. **41:** 690-697.

660 Myles, S., J. Peiffer, P. J. Brown, E. S. Ersoz, Z. Zhang *et al*, 2009 Association mapping:
661 Critical considerations shift from genotyping to experimental design. Plant Cell
662 **21:** 2194-2202.

National Agricultural Statistics Service, 2014 National Statistics for Corn. Available at: http://www.nass.usda.gov/Statistics_by_Subject/result.php?08638011-B478-3942-B44F-FA130ADDE283&sector=CROPS&group=FIELD%20CROPS&comm=CORN

Oliphant, A., D. Barker, J. Stuelpnagel and M. Chee, 2002 BeadArray (TM) technology: Enabling an accurate, cost-effective approach to high throughput genotyping. BioTechniques **32**: 56-58.

Parts L., Cubillos F. A., Warringer J., Jain K., Salinas F., et al. , 2011 Revealing the genetic structure of a trait by sequencing a population under selection. Genome Res. **21**:1131–1138.

R Core Team, *2014 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, Vienna. Available at: http://www.R-project.org/.

Rincent, R., L. Moreau, H. Monod, E. Kuhn, A. E. Melchinger *et al*, 2014 Recovering power in association mapping panels with variable levels of linkage disequilibrium. Genetics **197:** 375-387.

Romay, M. C., A. Butrón, A. Ordás, P. Revilla and B. Ordás, 2012 Effect of recurrent selection on the genetic structure of two broad-based Spanish maize populations. Crop Sci. **52:** 1493-1502.

Saghaimaroof, M., K. Soliman, R. Jorgensen and R. Allard, 1984 Ribosomal DNA spacer-length polymorphisms in barley - Mendelian inheritance, chromosomal location, and population-dynamics. Proc. Natl. Acad. Sci. **81:** 8014-8018.

Schwab, E.C., R.D. Shaver, J.G. Lauer and J.G. Coors, 2003 Estimating silage energy value and milk yield to rank corn hybrids. Anim. Feed Sci. Technol. **109:** 1-18.

Shaver, R., J. Lauer, J. Coors and P. Hoffman, 2006 Corn silage evaluation: MILK2000 challenges and opportunities with MILK2006. Home page address: http://www.uwex.edu .

Turner T. L., Stewart A. D., Fields A. T., Rice W. R., Tarone A. M., 2011 Population-based resequencing of experimentally evolved populations reveals the genetic

basis of body size variation in Drosophila melanogaster. PLoS Genet. **7**: e1001336.

van Heerwaarden, J., M. B. Hufford and J. Ross-Ibarra, 2012 Historical genomics of North American maize. Proc. Natl. Acad. Sci. **109:** 12420-12425.

Weir, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.

Weir, B. S., and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. Evolution 1358-1370.

Wimmer, V., T. Albrecht, H.J. Auinger and C.C. Schon, 2012 Synbreed: A framework for the analysis of genomic prediction data using R. Bioinformatics **28:** 2086-2087.

Wisser, R. J., P. J. Balint-Kurti and J. B. Holland, 2011 A novel genetic framework for studying response to artificial selection. Plant Genetic Resources **9:** 281-283.

Wisser, R. J., S. C. Murray, J. M. Kolkman, H. Ceballos and R. J. Nelson, 2008 Selection mapping of loci for quantitative disease resistance in a diverse maize population. Genetics **180:** 583-599.

Wright, S. I., I. Vroh Bi, S. G. Schroeder, M. Yamasaki, J. F. Doebley *et al*, 2005 The effects of artificial selection on the maize genome. Science **308:** 1310-1314.

Yu, J. M., G. Pressoir, W. H. Briggs, I. V. Bi, M. Yamasaki *et al*, 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38:** 203-208.

717

718

719

720

721

722

723

724

725

726

727 Table 1. Numbers of individuals with genotypic data and phenotypic data included in the evaluation of the Wisconsin Quality
728 Synthetic population selected for five cycles for silage yield and compositional traits.

Cycle	Both phenotypic and genotypic data							Genotypic data
	Silage yield	Dry matter	NDF†	ADF†	IVTD†	CP†	Starch	
WQS C0	0	0	0	0	0	0	0	5
WQS C1	6	6	6	0	6	6	6	11
WQS C2	0‡	60	60	60	60	60	60	163
WQS C3	80	80	79	79	79	79	79	88
WQS C4	154	154	114	114	114	114	114	170
WQS C5	0	0	0	0	0	0	0	211
Total	240	300	259	253	259	259	259	648

729 † NDF = neutral detergent fiber; ADF = acid detergent fiber; IVTD = *in vitro* true digestibility; CP = crude protein

730 ‡ Data for silage yield from WQS C2 was not included because of zero heritability (see Table 2).

731

732

733

734

735 Table 2. Mean, minimum, and maximum of each trait in each year (cycle) of evaluation.

		WQS C1					WQS C2					WQS C3					WQS C4			
Trait	Units	Mean	Min	Max	S/σ_P†		Mean	Min	Max	S/σ_P		Mean	Min	Max	S/σ_P		Mean	Min	Max	S/σ_P
Silage yield	Mg ha^{-1}	7.2	5.7	8.5	0.74		8.9	8.5	10.5	0.24		9.7	5.3	12	1.17		8.7	6.1	10.5	0.81
Dry matter	%	32.1	26.6	42.0	0.07		39.5	33.5	49.2	-0.34		34.1	29.3	39.8	0.44		36.9	30.1	43.7	0.31
NDF	%	53.3	48.2	58.5	-0.23		50.5	46.7	55.4	-1.11		46.9	42.9	51.1	-0.35		44.8	42.0	49.3	0.23
ADF	%	--‡	--	--	--		26.2	23.6	29.7	-1.04		24.7	22.1	27.7	-0.42		23.0	21.4	25.8	0.24
IVTD	%	70.9	65.9	75.0	0.26		82.4	80.0	86.2	1.07		78.6	76.1	81.5	0.26		81.1	78.2	83.7	-0.33
CP	%	7.2	6.3	8.0	-0.18		8.0	7.2	8.8	0.15		7.0	6.2	8.1	-0.24		6.9	6.2	7.5	0.51
Starch	%	21.9	12.7	30.6	-0.01		27.9	20.8	33.7	0.61		30.5	20	34.9	0.37		31.4	25.2	35.7	0.04

736

737 † A standardized selection differential was calculated for each cycle by dividing the selection differential by the phenotypic standard
738 deviation.

739 ‡ ADF was not measured in WQS C1.

740

741

742

743

744

745 Table 3. Broad-sense heritability on a family-mean basis (*H*) and genomic heritability (h_G^2) for each trait in each selection cycle of the
746 Wisconsin Quality Synthetic selection program, and h_G^2 across cycles.

Cycle	Silage yield		Dry matter		NDF		ADF		IVTD		CP		Starch	
	H	h_G^2	H	h_G^2	H	h_G^2	H	h_G^2	H	h_G^2	H	h_G^2	H	h_G^2
WQS C1	0.43	-- †	0.73	--	0.33	--	-- ‡	--	0.38	--	0.54	--	0.58	--
WQS C2	0	0	0.71	0.53	0.52	0.04	0.49	0.05	0.52	0.01	0.73	0.57	0.71	0.21
WQS C3	0.59	0.16	0.82	0.42	0.34	0	0.35	0	0.48	0	0.60	0.10	0.66	0.31
WQS C4	0.33	0.06	0.64	0.69	0.41	0	0.41	0	0.59	0.27	0.32	0	0.61	0.81
Across cycles	--	0.11	--	0.42	--	0.01	--	0.06	--	0.02	--	0.23	--	0.18

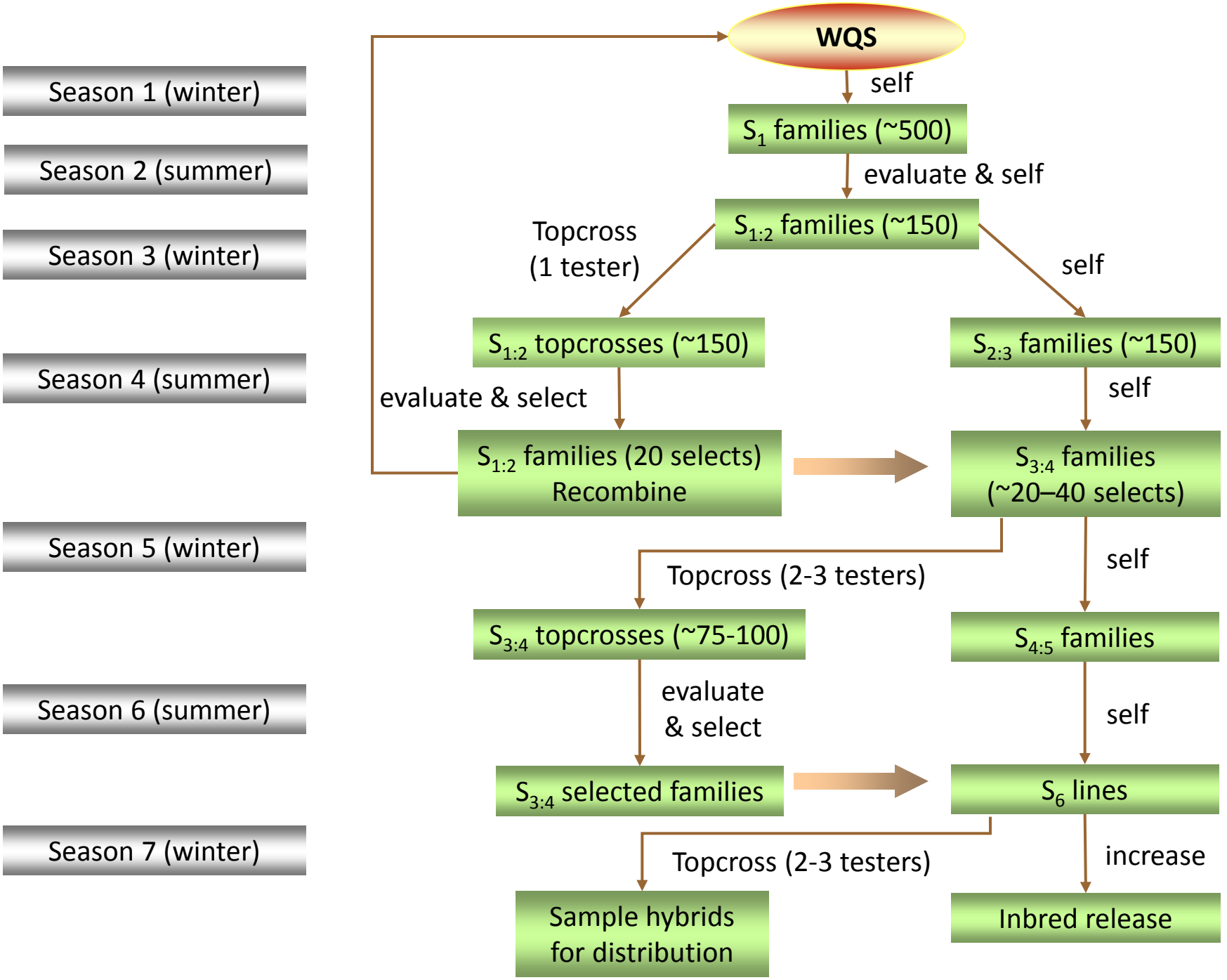
747

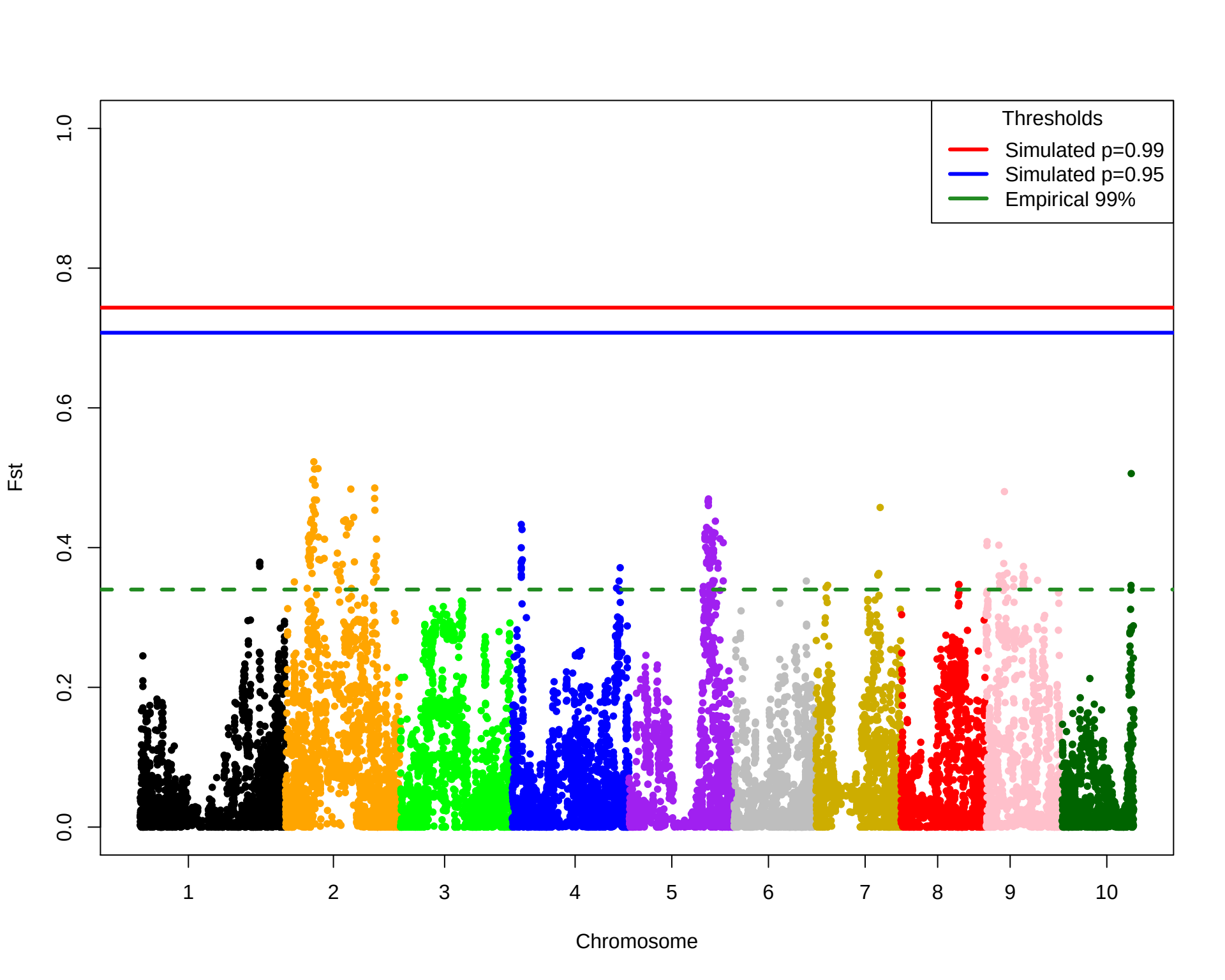
748 † Not enough individuals were genotyped in Cycle 1 to calculate h_G^2.

749 ‡ ADF was not measured in WQS C1.

750

751


Season 1 (winter)
Season 2 (summer)
Season 3 (winter)
Season 4 (summer)
Season 5 (winter)
Season 6 (summer)
Season 7 (winter)
WQS
self
S1 families (~500)
evaluate & self
S1:2 families (~150)
Topcross
(1 tester)
self
S1:2 topcrosses (~150)
S2:3 families (~150)
evaluate & select
self
S1:2 families (20 selects)
Recombine
S3:4 families
(~20–40 selects)
Topcross (2-3 testers)
self
S3:4 topcrosses (~75-100)
S4:5 families
evaluate
& select
self
S3:4 selected families
S6 lines
Topcross (2-3 testers)
increase
Sample hybrids
for distribution
Inbred release


Thresholds
Simulated p=0.99
Simulated p=0.95
Empirical 99%
Fst
0.0
0.2
0.4
0.6
0.8
1.0
1
2
3
4
5
6
7
8
9
10
Chromosome


1.00
0.75
0.50
0.25
0.00
Gene Diversity
C2
C3
C4
C5
chr
1
2
3
4
5
6
7
8
9
10
chromosomes