bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Title: A simple test identifies selection on complex traits in breeding and experimentally-evolved populations

Authors: Tim Beissinger[1,2,3,**], Jochen Kruppa[4,5], David Cavero[7], Ngoc-Thuy Ha[4], Malena Erbe[6], Henner Simianer[4,**]

1 USDA Agricultural Research Service, Columbia Missouri, USA
2 University of Missouri Division of Biological Sciences, Columbia, Missouri, USA
3 University of Missouri Division of Plant Sciences, Columbia, Missouri, USA
4 University of Göttingen, Center for Integrated Breeding Research, Göttingen, Germany
5 Institute for Animal Breeding and Genet., Univ. of Veterinary Medicine, Hannover, Germany
6 Institute for Animal Breeding, Bavarian State Research Centre for Agriculture, Grub, Germany
7 H&N International GmbH, Cuxhaven, Germany

**** To whom correspondence may be addressed:**

Tim Beissinger
USDA-ARS Plant Genetics Research Unit
University of Missouri
Columbia, MO 65211
Tim.beissinger@ars.usda.gov
+1-573-882-7606

Henner Simianer
University of Goettingen
Department of Animal Sciences
Albrecht-Thaer-Weg 3, 37075 Goettingen
hsimian@gwdg.de
+49-551-395604

Abstract: Important traits in agricultural, natural, and human populations are increasingly being shown to be under the control of many genes that individually contribute only a small proportion of genetic variation. However, the majority of modern tools in quantitative and population genetics, including genome wide association studies and selection mapping protocols, are designed to target the identification of individual genes with large effects. We have developed an approach to identify traits that have been under selection and are controlled by large numbers of loci. In contrast to existing methods, our technique utilizes additive effects estimates from all available markers, and relates these estimates to allele frequency change over time. Using this information, we generate a composite statistic, denoted $\hat{G}$, which can be used to test for significant evidence of selection on a trait. Our test requires genotypic data from multiple time points but only a single time point with phenotypic information. Simulations demonstrate that $\hat{G}$ is powerful for identifying selection, particularly in situations where the trait being tested is controlled by many genes, which is precisely the scenario where classical approaches for selection mapping are least powerful. We apply this test to breeding populations of maize and chickens, where we demonstrate the successful identification of selection on traits that are documented to have been under selection.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Introduction

Quantitative traits encompass an inexhaustible number of phenotypes that vary in populations, from characters such as height (Yang *et al.* 2010) to weight (Barsh *et al.* 2000) to disease resistance (Poland *et al.* 2009). These types of traits are so essential for agriculture and human health that the entire field of quantitative genetics revolves around their study (Plomin *et al.* 2009; Wallace *et al.* 2014). However, the nature of quantitative traits makes it difficult to study their genetic basis; for nearly a century, scientists have modeled quantitative traits by assuming that their underlying control involves many loci each contributing a very small proportion to genetic variance (Fisher 1918), the so-called 'infinitesimal model'. Therefore, conducting studies with enough power to identify a substantial proportion of the loci that contribute to a quantitative trait requires a massive sample size, imposing financial and logistical barriers. However, this model of quantitative trait variation does an excellent job when predicting important characteristics such as response to selection (Visscher *et al.* 2008). For instance, the recent development of genomic prediction methodologies (Meuwissen *et al.* 2001) allow the breeding value and/or phenotype of individuals to be predicted with remarkable precision from genomic information alone.

The models of quantitative genetics have had a less dramatic impact on studies of evolutionary adaptation, where genomes are often scanned to identify adaptive loci with large effects (Akey 2009). Positive selection on such loci leaves behind pronounced signatures, deemed "selective sweeps". There is an abundance of evidence for such sweeps in humans (Sabeti *et al.* 2007), natural populations (Schweizer *et al.* 2016), livestock (Qanbari and Simianer 2014), and crops (Hufford *et al.* 2012; Qanbari and Simianer 2014; Schweizer *et al.* 2016). However, alternative forms of selection, including purifying selection against new mutations (Lawrie *et al.* 2013), selection on standing variation (Garud *et al.* 2015), or selection on many loci of small effect (Turchin *et al.* 2012), rarely leave these discernible signatures at individual loci. Evidence of these forms of selection can be difficult to identify. When they can be found, it is often through the pooling of weak evidence at individual loci into a stronger signal across a class of loci. For example, Beissinger et al (2016) demonstrated the importance of purifying selection during maize evolution by combining evidence from all maize genes. An approach implemented by Berg and Coop (2014) tests for evidence of selection on a quantitative trait by evaluating allele frequencies at all loci that have previously been implicated by genome-wide association studies (GWAS) as putatively associated with that trait. This approach has since been used to test for selection on multiple human traits, including height (Mathieson *et al.* 2015) and telomere length (Hansen *et al.* 2016).

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

In studies of model organisms or agricultural species, large collections of previously identified "GWAS hits" are not as abundant as in humans, on which the Berg and Coop (2014) method depends. This is due in part to the more modest sample sizes that tend to be used in experimental settings compared to clinical studies, often combined in large-scale meta-analyses (Evangelou and Ioannidis 2013). In contrast to humans, however, genotypic data across multi-generational time points are often available for model and agricultural species. We have developed a test for selection on complex traits that leverages such genotype-over-time data. Our test depends on the relationship between the change in allele frequency across generations and the estimated additive effect of the same allele, computed for every genotyped locus. We use these values to compute an estimate of the direction of genetic gain, which can be shown to be additive across all loci considered and lends itself to a straight-forward positional dissection of the accumulated genetic gain. Because phenotypic data are incorporated, a permutation-based test for significance can be utilized so that significance testing is not limited by many of the demographic history and population structure related caveats that complicate determining significance when testing for selection (de Villemereuil *et al.* 2014). Our method utilizes additive effects estimates for each locus calculated simultaneously, using shrinkage-based methods that have been honed over the past 15 years for the purpose of genomic selection and prediction (Campos *et al.* 2013). Therefore, this text can be considered analogous to reverse genomic selection; rather than using predictions of breeding value to drive selection and hence future changes in allele frequency, we use the same data coupled with knowledge of past changes in allele frequency to make inferences regarding which traits were effectively under selection in the past. Interestingly, we find by simulation that this approach is most powerful for identifying selection on traits controlled by many loci of small effect, which is exactly the situation where other tests for selection and/or association are least powerful.

Herein, we first describe our test for selection on complex traits. Then, we perform simulations demonstrating the validity of the method and explore the situations where it is most and least powerful. Finally, we apply the method to breeding populations of maize and chicken. In both of these experimental situations, we successfully identify the traits that are known to have been selected. Collectively, our results demonstrate that this approach may be leveraged to identify novel traits or component-traits that may be used to inform future breeding decisions and/or for enhanced historical, ecological, and basic scientific understanding.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Results

Theoretical Motivation:

Assume that a trait is fully controlled by additive di-allelic loci $j = 1, \ldots m$. Then the genotypic value, α_j, of an allele at locus j, is equal to its gene substitution effect, α_j. Based on this equivalency, the mean phenotypic effect, M_j, attributable to the locus is given by $M_j = \alpha_j(2p_j-1)$, where p_j is the frequency of the reference allele at this locus. It follows that the change in the population mean resulting from selection on this locus, what we may consider the locus-specific response to selection, is given by

$R_j = M_{j1} - M_{j0} = \alpha_j(2p_{j1}-1) - \alpha_j(2p_{j0}-1) = 2\alpha_j(p_{j1}-p_{j0})$,

where p_{j0} is the allele frequency before selection and p_{j1} is the allele frequency after selection. Define $\Delta_j = (p_{j1}-p_{j0})$, leading to $R_j = 2\,\Delta_j\alpha_j$. Based on our earlier assumption of complete additivity, summing over all m loci provides a genome-wide estimate of the response to selection (Falconer and Mackay 1996):

$$\hat{R} = 2\sum_{j=1}^{m} \Delta_j \alpha_j \qquad (1)$$

This estimate of selection response also naturally arises from the logic of random regression BLUP (RRBLUP) (Meuwissen *et al.* 2001). Here, a model is used

$$y = Xb + Zs + e \ , \qquad (2)$$

where $\boldsymbol{y}$ is a vector of length n containing phenotypes for a specific trait, b are fixed effects, $s \sim N(0, I\sigma_s^2)$ is the vector of length m containing additive SNP effects at m loci; $e \sim N(0, I\sigma_e^2)$ is the vector of random residual terms, and σ_s^2 and σ_e^2 are the corresponding variance components. X and Z are incidence matrices linking observations in y to the respective levels of fixed effects in b and random SNP effects in s. In more detail, Z is an $n \times m$ matrix where element z_{ij} contains the genotype of individual i at SNP locus j. Since such models are invariant with respect to linear transformations of the allele coding (Strandén *et al.* 2011), we may use the notation $z_{ij} = 0, \frac{1}{2}, or\ 1$, standing for zero, one, or two copies of the reference allele. Note that with this coding, s_j is equivalent to $2\alpha_j$ in the coding above, since it reflects the contrast between the two homozygous genotypes at locus j. Due to the equivalence of genomic BLUP (GBLUP; VanRaden 2008) and RRBLUP (Endelman 2011), it is possible to calculate genomic breeding values of the genotyped individual as $\hat{u} = Z\hat{s}$, where $\hat{s}$ are the solutions for the SNP effects obtained using RRBLUP with model (2).

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Assume now further that individuals in the vector y can be assigned to g discrete generations and that the individuals of the oldest generation come first and the individuals of the last generation come latest. We then can define a $g \times n$ matrix

$$L = \begin{bmatrix} l_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_g \end{bmatrix},$$

where l_i is a row vector of length n_i, which is the number of individuals in generation i, of which all elements are $1/n_i$. With this, a vector $\bar{u}$ of length g reflecting average breeding values per generation can be calculated as $\bar{u} = L\hat{u}$, and estimated selection response results as $\hat{R} = \bar{u}_g - \bar{u}_1$.

Now, $\bar{u} = L\hat{u} = LZ\hat{s}$, where LZ is a $g \times m$ matrix in which element i, j reflects the average allele frequency of the reference allele at SNP j in generation i. The allele frequency change between generation 1 and generation g can be obtained as a linear contrast between the first and the last row of this matrix as $\Delta = k'LZ$ where k is a vector of length g with $k_1 = -1$, $k_g = 1$, and all other elements are zero. Finally, the selection response can be written as $\hat{R} = \Delta\hat{s}$, which is identical to equation (1), given that s is equivalent to 2α.

Furthermore, theory suggests that under the assumption that selection intensity is equal for all genes across the genome, the change of allele frequency Δ_i should be approximately proportional to the allele effect α_i, such that for a trait under selection a non-zero correlation between allele frequency change and the additive effect of alleles on that trait is expected (Wright 1937). Alternatively stated, (1) emphasizes the temporal component of the Breeder's Equation, $R = h^2S$, where h^2 is the narrow-sense heritability of a trait and S is the selection differential. Given a population of individuals with more than one time-point of genotypic data, it is simple to compute Δ_i for every genotyped locus. Furthermore, the shrinkage methods of genomic prediction (Campos *et al.* 2013), including Ridge Regression (Endelman 2011) and GBLUP (VanRaden 2008) allow additive effects, α_i, to be approximated for every genotyped position. For this, a set of individuals genotyped and phenotyped in at least one generation is needed.

A notable benefit of the estimator in (1) is that by leveraging temporal data from genotypes rather than from phenotypes, it only requires one generation of phenotyping. Additionally, this suggests that if we consider R a random variable, then given the distribution of R in a scenario without selection, a test of whether or not $\hat{R}$ is different from zero may be performed. Since $\hat{R}$ is the genomic response to

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

selection, this is equivalent to testing whether or not a trait has been under selection during the timeframe under study.

Significance Testing:

We implemented a permutation-based strategy to test whether or not $\hat{R}$ is significantly different from zero. Genetic drift and selection jointly determine changes in allele frequency, Δ_i, but without selection these changes in frequency should not be related to effect size. The reverse is also true; effect sizes, α_i, are estimated based on a genomic prediction model applied to phenotypes measured in a single panel of individuals, and therefore they are not related to changes in allele frequency. This suggests that a null distribution for $\hat{R}$ in a no-selection scenario may be generated via a permutation approach. Assuming no linkage disequilibrium (LD) between markers, a simple shuffling of Δ_i and α_i can be implemented to generate the desired null distribution. However, LD between markers compromises the applicability of this simplified approach for most populations—such an approach overestimates the sample size of the permutation test by treating each marker as an independent observation, while in reality any level of LD between markers leads to fewer independent observations than markers. Therefore, we have employed a semi-parametric approach that scales the variance of the permutation test statistic according to the realized extent of LD to alleviate this discrepancy.

Let $\hat{G} = \sum_{j=1}^{m} \Delta_j \alpha_j$, which is proportional to $\hat{R}$ as defined in (1). Define **p** to be a vector of length m that is a permutation of the vector **J** = [1,..,m]. A permuted value of $\hat{G}$ may be obtained via $\hat{G}_{perm} = \sum_{j}^{m} \Delta_j \alpha_{p_j}$. Because Δ_j and α_{p_j} are no longer indexed to the same locus, $\hat{G}_{perm}$ does not reflect selection, but instead captures genetic drift over time (Δ_j terms) as well as the genetic architecture of the underlying trait (α_j terms). Generating repeated values of $\hat{G}_{perm}$ through repeated permutations of **J** therefore generates a null distribution for $\hat{G}$ which assumes no selection and complete linkage equilibrium.

The Central Limit Theorem dictates that realizations of $\hat{G}_{perm}$ are normally distributed with approximate mean $\overline{\hat{G}_{perm}}$ and standard deviation $SE(\hat{G}_{perm})$. Therefore, σ, the underlying standard error of a single-locus estimate for $\hat{G}_{perm}$ is given by $\sigma = SE(\hat{G}_{perm})\sqrt{m}$, where $SE(\hat{G}_{perm})$ is the observed standard error of $\hat{G}_{perm}$. Consider the quantity m_{ind}, representing the effective number of independent loci. If the standard deviation of $\hat{G}_{perm}$ were calculated using m_{ind} independent markers, its expectation would be $SE_{ind}(\hat{G}_{perm}) = \sigma/\sqrt{m_{ind}}$. Plugging in the estimate for σ obtained above,

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

$SE_{ind}(\hat{G}_{perm})$ becomes $SE_{ind}(\hat{G}_{perm}) = SE(\hat{G}_{perm})\sqrt{m/m_{ind}}$. In summary, to test for selection, $\hat{G}$ may be calculated from data, and then a permuted null distribution for $\hat{G}$ that assumes linkage equilibrium may be generated. This permutation distribution is then approximated with a normal distribution and the variance is scaled according to the effective number of independent markers. Ultimately, significance is evaluated by comparing $\hat{G}$ to a normal distribution with mean $\overline{\hat{G}_{perm}}$ and standard deviation $SE(\hat{G}_{perm})\sqrt{m/m_{ind}}$.

Simulations:

We conducted a series of simulations to evaluate the power of the $\hat{G}$ statistic for identifying selection on complex traits. Genotypic data were simulated with the software program QMSim (Sargolzaei and Schenkel 2009). We simulated selection in a generic species with ten chromosomes, each 100 cM in length, with a total of 100,000 equally-spaced markers (10,000 per chromosome). In the first step of each simulation, the total population was established based on 10,000 individuals randomly mating for 5,000 generations. Then, 500 males and 500 females were randomly chosen to establish a base population that would undergo selection. Each generation, 1,000 individuals (500 males and 500 females) were permitted to mate out of a population of 5,000, providing a selection proportion of 20%. For each simulation, heritability was set to 0.5. This general scheme encapsulates characteristics of most plant and animal breeding populations, including the large number of progeny typical of plants and the truncation selection protocol often associated with animal breeding and/or selection in the wild. Additional details regarding the simulated population are included in Supplemental Table 1.

We simulated variable numbers of additive QTL controlling traits, from 10, representing a simple trait controlled by large-effect QTL, to 1,000, representing a highly quantitative trait controlled nearly infinitesimally. QTLs were evenly spaced along each chromosome and QTLs themselves were not included in the marker set for analysis. One hundred simulations were performed for each level of trait complexity. First, we used these simulations to establish the appropriate number of independent markers, m_{ind} as defined above, for this test. We calculated how distant two markers must be to have an expected LD level of $R^2 \leq 0.03$. Then we counted the total number of blocks of this size genome-wide. The 0.03 level was established by performing a grid-search of potential values and tuning the false positive rate (Supplemental Figure 1). An LD cutoff that is too high leads to a high false-positive rate,

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

while one that is too low weakens the power of the test. For populations similar to those we simulated, we expect that requiring $R^2 \leq 0.03$ will be appropriate.

When we tested for selection in our simulated data, we observed a direct relationship between the number of QTL controlling a trait and the power of $\hat{G}$ to identify selection on that trait. $\hat{G}$ powerfully identifies selection on highly polygenic traits, but is not powerful for identifying selection on traits controlled by a small number of QTLs. Analyses of the same simulations using F_{ST}-based selection mapping showed that traits controlled by a small number of QTLs can be mapped using traditional selection mapping approaches. However, as traits become increasingly polygenic, our simulations demonstrate that the ability to map individual selected genes diminishes (Figure 1). These findings demonstrate how $\hat{G}$ and traditional mapping are complementary depending on the underlying genetic architecture of a trait. Table 1 depicts detection and false positive rates for $\hat{G}$ and F_{ST}-based mapping under different genetic architectures.

Genetic Architecture	**10 QTL**	**50 QTL**	**100 QTL**	**1,000 QTL**
Ĝ				
True positive rate	0.04	0.54	0.94	1.0
False positive rate	0.03	0.03	0.02	0.03
***F_{ST}*-based Selection Mapping**				
Mean # true positives (rate)	5.6 (56%)	22 (44%)	39 (39%)	187 (18.7%
Mean # false positives	52	280	715	1,745

***Table 1*: Detection and false positive rates for Ĝ and Selection mapping. One Ĝ test is conducted per simulation, so true and false positive rates are shown. For selection mapping, one test is conducted per marker in each simulation, so the mean number of markers that were declared true and false positives is shown. A marker was declared a false positive in selection mapping if it exceeded a 5% simulation-based experiment-wide significance threshold but was not within a .1 cM region around a simulated QTL.**

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

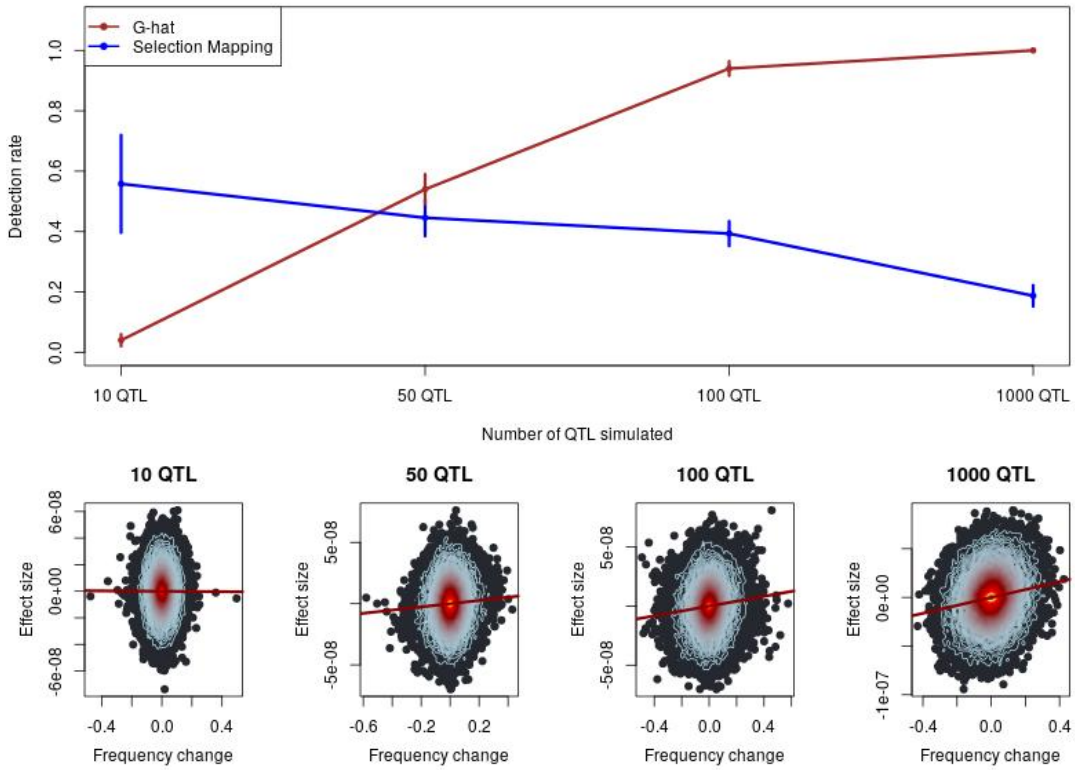


***Figure 1*: The power of Ĝ to identify selection. Top: The detection rate of Ĝ compared to Fst-based selection mapping. Vertical lines indicate one standard deviation. Standard deviations for selection mapping were estimated empirically. Standard deviations for Ĝ were estimated based on the binomial distribution. Bottom: Exemplary heat plots depicting individual-SNP allelic effect estimates linearly regressed on allele frequency change over time. Each point represents a SNP, while the contour lines indicate the density of SNPs. From the regression line, observe that a stronger relationship between frequency change and effect size corresponds to increasing polygenicity.**

Selection on maize silage traits:

We re-analyzed data from a previous study that tested for selection in a decades-long breeding program for maize silage quality (Lorenz *et al.* 2015). Very briefly, a selection index comprised of experimentally-measured traits related to silage quality was used to perform reciprocal recurrent selection for breeding improved maize. Traits comprising the index included acid detergent fiber (ADF), protein content, starch content, *in-vitro* digestibility, and yield (www.cornbreeding.wisc.edu). In total, 648 individuals from various stages of selection were genotyped. Between 240 and 300 of these individuals were also phenotyped, depending on the trait. Selection mapping (Wisser *et al.* 2008) was previously performed utilizing simulations of drift to scan for selection, but the analysis did not identify any loci that showed

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

significant evidence of selection. This is in spite of quantifiable improvement of the population and demonstrated heritability of the index-composing traits (Lorenz *et al.* 2015). We re-analyzed the same data to evaluate evidence for polygenic selection on the measured traits, which included NDF, *in-vitro* digestibility, crude protein content, starch content, yield, and dry matter. After filtering, these data consisted of 10,023 polymorphic markers. Due to the relatively small population size and recurrent selection breeding scheme, we expect slow LD decay and therefore for most of the genome to be represented with this marker set. Further analysis of LD to determine the value of M_{eff} to utilize in our test for selection confirms this (Supplemental Figure 2).

Figure 2 depicts the maize patterns of selection that were observed in our analysis. In these plots, the histogram shows the null distribution of $\hat{G}$ that was observed from a permutation test, while the vertical line depicts the observed value of $\hat{G}$ when applied to the experimental data. We observed that with the exception of protein, for the traits where we had an *a priori* expectation of selection, we not only identified that selection did occur, but we correctly estimated the direction of selection (positive or negative) from the data. One of the traits measured was silage dry matter (DM), which was not a part of the selection index. We did not identify evidence of selection on DM, as was expected.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

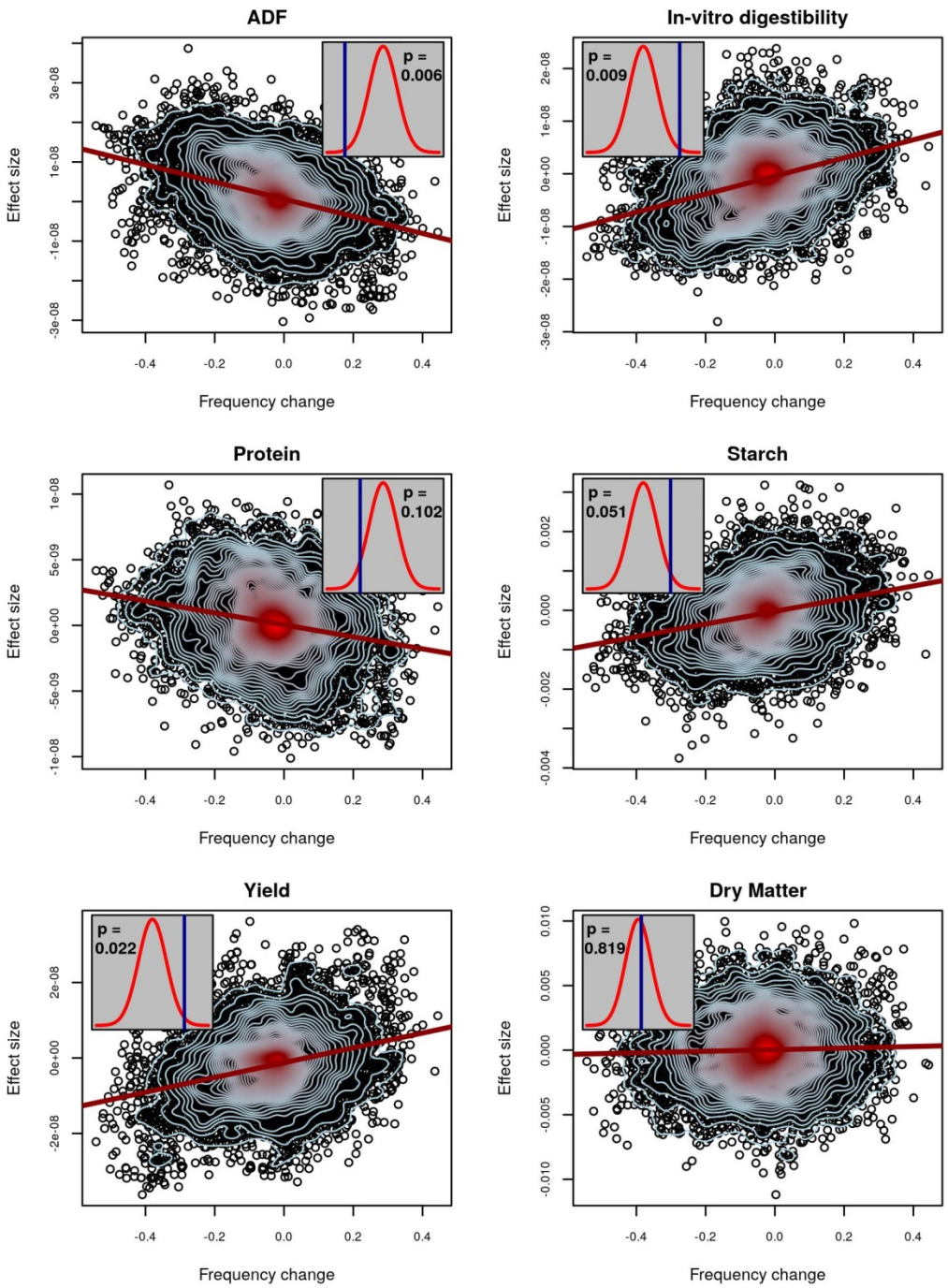


Figure 2: Evidence of selection for maize silage traits. For six traits, the relationship between estimated allelic effects at individual SNPs and the change in allele frequency over generations is plotted. The red line is a regression of effect size on allele frequency change. Contour lines indicate the density of points, with blue contours indicating fewer points than red. Inset plots depict observed values of $\hat{G}$ (blue lines) and their statistical significance based on a comparison to permuted null

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

distributions (red densities) for no-selection scenarios. An exact two-sided p-value is given within each inset. Significant values of $\hat{G}$ above the permuted mean indicate selection operated in the positive direction, while significant values below the permutation mean indicated selection operated in the negative direction.

Selection on chicken traits:

We tested for evidence of selection in two panels of commercial lines of laying hens: one white layer (WL) and one brown layer (BL). Both closed lines have been selected over decades with a similar composite breeding goal, comprised of laying rate, body weight and feed efficiency, egg weight, and egg quality, among other objectives. The respective weights applied to the different traits varied between lines and over time. Traits analyzed included laying rate, egg weight, and breaking strength of eggs. Genotypes were available only for the post-selection population, so initial allele frequencies were inferred based on pedigree data (Gengler *et al.* 2007). M_{ind} was determined based on separate evaluations of LD in the WL (Supplemental Figure 3) and BL (Supplemental Figure 4) populations.

Among the traits evaluated, we observed significant evidence of selection for increased laying rate in both WLs ($p = 0.027$) and BLs ($p = 0.0193$). Tests were also suggestive of selection for increased eggshell breaking strength in WLs ($p < 0.1$; one-sided $p < 0.05$), while there was no evidence of directed selection for egg weight (Figure 3). To verify that these results were not driven by a small number of SNPs with high estimated effect sizes, we repeated the analysis with the 10 largest effect-size SNPs removed and saw virtually identical results (Supplemental Figure 5). The result for egg weight can be seen as a 'negative control' since for this trait an optimum value is already achieved and maintained by balancing selection. The fact that we were not able to detect significant evidence of selection in a trait such as eggshell breaking strength in both lines (although a tendency can be observed) may be due to the fact that improving those traits is part of a complex multi-objective breeding program, or simply that our test was underpowered for these traits. The unavailability of experimentally-estimated initial frequencies and our alternative use of pedigree-inferred initial allele frequencies likely weakened the power of the test as compared to the more complete data available for maize and in the simulations.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

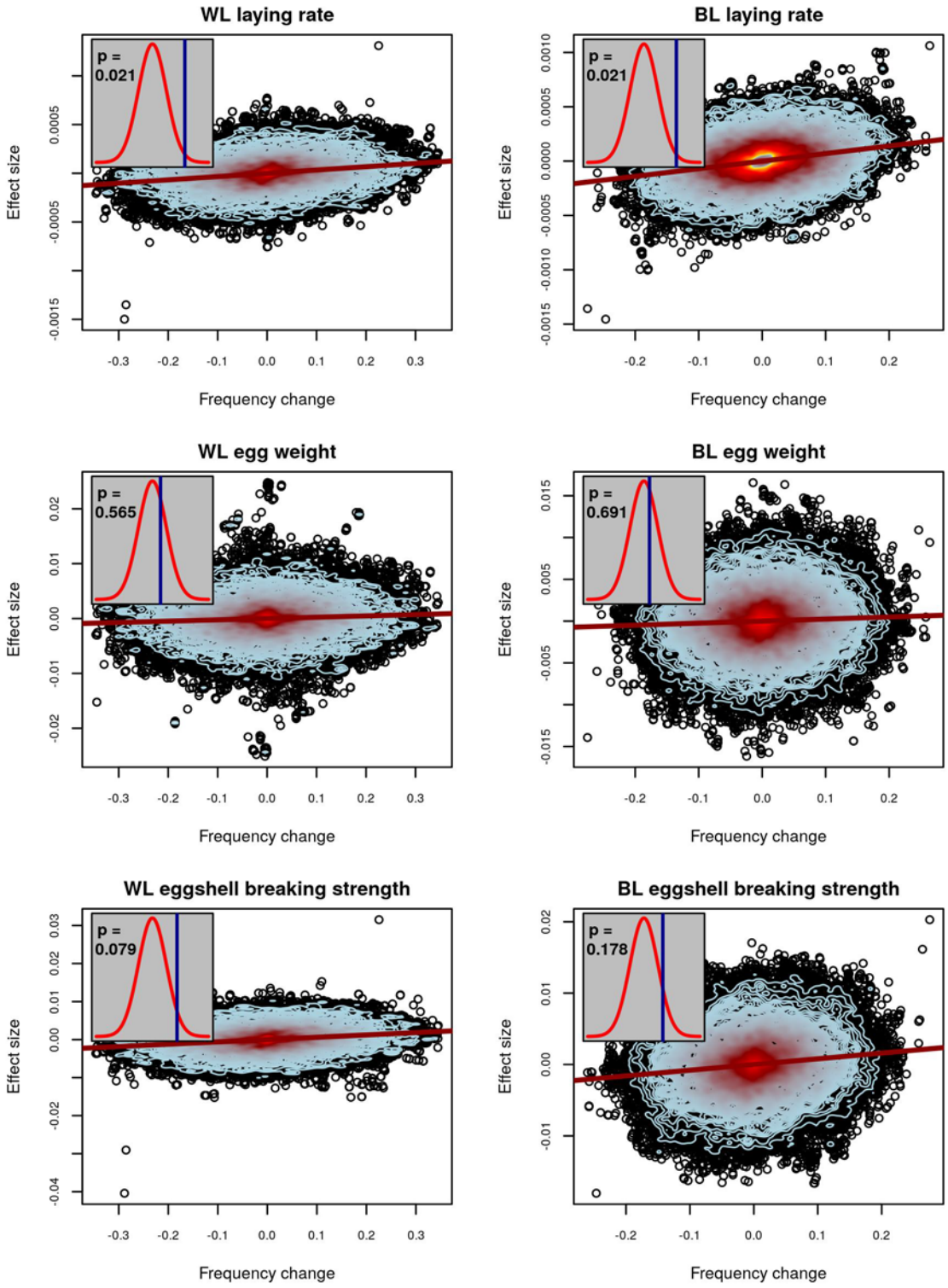


Figure 3: Evidence of selection for chicken traits. For three traits in white (left column) and brown (right column) laying hens, the relationship between estimated allelic effects at individual SNPs and the change in allele frequency over generations is plotted. The red line is a regression of effect size on

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

allele frequency change. Contour lines indicate the density of points, with blue contours indicating fewer points than red. Inset plots depict observed values of $\hat{G}$ (blue lines) and their statistical significance based on a comparison to permuted null distributions (red densities) for no-selection scenarios. An exact two-sided p-value is given within each inset. Significant values of $\hat{G}$ above the permuted mean indicate selection operated in the positive direction, while significant values below the permutation mean indicated selection operated in the negative direction.

Discussion

We have defined a test statistic, $\hat{G}$, that combines phenotypic and genotypic information to test for selection on traits controlled by many loci of small effect. The approach utilizes estimated effect sizes for individual loci and allele frequency changes over time resulting from selection on those loci. Therefore, $\hat{G}$ is most applicable in experimental or breeding populations, where both pieces of information are readily available via genotyping individuals from multiple generations. However, phenotypic information for estimating allelic effects is only required from a single time-point, so this approach can be applied post-hoc using DNA samples from previous generations even if phenotyping is no longer possible. As the practice of sequencing ancient DNA from archeological sites, museum samples, or other sources becomes progressively commonplace (Orlando *et al.* 2015), we envision that this approach may become increasingly applicable for ecological questions, evolutionary studies, and for human research.

Powerful for highly quantitative traits

Methods for mapping genes associated with important traits or for identifying loci that are under selection are most powerful for large-effect genes. A simple explanation for the disappointing number of associations that have been uncovered to date through GWAS is that complex traits are often controlled by many genes of small effect (Yang *et al.* 2011). If this is the case, enormous sample sizes are required to map loci regardless of the methodological enhancements that can be applied. Human geneticists have had success studying complex traits by utilizing extremely large sample sizes (Rietveld *et al.* 2013; Wood *et al.* 2014). But, sample sizes of this magnitude are not yet achievable within resource limitations for most species, and, arguably, will never be. Conversely, population genetic studies aiming to scan for selection have been most successful at identifying hard sweeps, where a new mutation of large effect rapidly rises to fixation as a result of selection (Pritchard *et al.* 2010). Only few

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

methodologies with limited power exist for mapping soft sweeps, when the beneficial allele is already at intermediate frequency at the start of selection (Garud *et al.* 2015); Ma *et al.* 2015). A likely explanation for the presence of soft sweeps is that they often result from loci of small effect increasing in frequency slowly in a population and therefore existing on multiple distinct haplotypes or mutating multiple times before fixation. In an agricultural context, many soft sweeps may be due to newly defined breeding goals which put selection pressure on genes that previously were segregating in the populations, but were selectively neutral. The $\hat{G}$ statistic does not attempt to map specific genes—instead it pools information from all SNPs to test for selection on specific traits. This approach completely avoids the question of which loci are associated with a trait. Instead of testing each SNP, we perform one test based on information from all SNPs. Therefore, a strong statistical signal arises when a large proportion of SNPs behave similarly but not when a few SNPs portray strong signals on their own.

It was recently argued that most complex disease traits in humans are controlled by small-effect genes dispersed throughout the genome – the omnigenic hypothesis (Boyle *et al.* 2017). Likewise, many important traits in agricultural animal and plant species tend to be quantitative in nature and are presumably controlled by small-effect genes (Goddard and Hayes 2009; Wallace *et al.* 2014). For these agricultural organisms, geneticists and breeders have long recognized the benefits that can be achieved by predicting breeding values and/or phenotypes based on models that use all SNPs simultaneously (Meuwissen *et al.* 2001; Heffner *et al.* 2009; Goddard and Hayes 2009). In fact, the development of these models has led to dramatic re-designs of modern breeding protocols (Schaeffer 2006; Cabrera-Bosquet *et al.* 2012). The $\hat{G}$ statistic represents one avenue to leverage information from all measured SNPs to gain an understanding of the evolutionary history of a population. This approach is analogous to genomic selection/prediction as utilized by animal and plant breeders, with an important distinction: instead of predicting breeding values to determine which individuals should be selected for the future, it utilizes genotypic frequencies over time coupled with phenotypic information to unravel the history of selection in the past.

Genotypes from base population provide high power:

Compared to other methods that test for selection on quantitative traits (Berg and Coop 2014; Zeng *et al.* 2017), $\hat{G}$ is unique in that it leverages genotypic information from multiple time points and that it incorporates information from all SNPs instead of restricting to a previously identified set of SNPs from one or multiple independent GWAS's. With the exception of a few traits in heavily studied species,

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

such as human height (Wood *et al.* 2014), few species, if any, provide the enormous sample sizes required to implicate a large number of loci for any quantitative traits. This includes situations where scientists are reasonably certain that a genetic architecture consisting of small-effect loci persists. Importantly, $\hat{G}$ is powerful because of the independence of the estimation of allele frequency changes across generations and effect sizes, respectively. Even when allelic effects and/or allele frequency changes are small, they cumulatively generate a powerful test since they can be compared across all genotyped loci. However, our analysis of the chicken data suggested that the power of the test can be reduced through noisy estimation of allele frequency change. Our reliance on pedigree data to derive initial allele frequencies was not as precise as the direct measurement of initial allele frequencies that was conducted for maize. Although we were still able to find evidence of selection on traits including laying rate, which was almost certainly under the strongest selection, there were selected traits we did not detect potentially due to this noise.

Future directions and conclusions:

We have developed an approach to test for selected traits that avoids the preliminary identification of candidate genes or regions. The approach is particularly applicable in experimental, agricultural, and natural populations for which available resources dictate limited sample sizes for conducting massive mapping studies for such preliminary identification. In contrast to purely population-genetic analyses, which rely solely on genotypic information, our methodology requires that phenotypic data be collected on at least one time-point of genotyped individuals. Additionally, multiple time-points of genotypic information are needed, either directly or through pedigree-based imputation.

While the $\hat{G}$ statistic is most directly applicable for the discovery of traits that have been previously under selection during recent evolution, it may have additional applications. Recent studies have demonstrated that distinct physical regions of the genome, such as individual chromosomes, often contribute a disproportionate amount to trait variance (Bernardo and Thompson 2016). Rather than applying the $\hat{G}$ statistic genome-wide, future research should be conducted regarding whether it can be applied across any collections of loci such as individual chromosomes, pathways, gene-families, functional classes, or other categories to test if these show evidence of selection on a quantitative trait. This would represent a process allowing researchers to map significant features as opposed to genes.

Although $\hat{G}$ does have potential as a tool for mapping categories of selected loci, the simple identification of traits that have been under selection may prove enormously useful. Whether

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

agricultural, experimental, or natural, it is often difficult to determine all of the traits that are advantageous in a population or respond to natural or anthropogenic selection, including undesired selection responses. The application of the $\hat{G}$ statistic genome-wide allows this determination, which may help scientists select the right traits for maximum agricultural production, determine inadvertently selected lab traits impacting experimental outcomes, and establish ecologically important traits for survival in the wild.

Materials and methods

Simulations:

Each simulation started with a random mating historical population. After 5 thousand generations, selection began and simulations proceeded with more control over each generation. Truncation selection was performed based on high phenotype. Drift simulations were identical to selection simulations in terms of genome layout and genetic basis of the trait, but individuals were selected randomly. Simulations were performed with QMSim (Sargolzaei and Schenkel 2009). Parameters for simulations are provided in full in Supplemental Table 1.

Selection mapping in simulations:

Pre- and post-selection simulated allele frequencies were output from QMSim. These were used to calculate marker-specific F_{ST} values, as was performed by (Lorenz *et al.* 2015). F_{ST} was computed according to $F_{ST} = \frac{s^2}{\bar{p}(1-\bar{p})+s^2/2}$, where s^2 is the sample variance of allele frequency between pre- and post-selection populations, and $\bar{p}$ is the mean allele frequency (Weir and Cockerham 1984). Experiment-wide 5% significance threshold were identified based on the 95% F_{ST} quantile observed from drift simulations. These thresholds were applied to F_{ST} values obtained from selection simulations to determine detection and false positive rates. Simulated QTL were declared detected if a significant marker was identified within a .1 cM window surrounding the QTL. False positives were defined as markers that were not within a .1 cM window surrounding any simulated QTL.

Maize data:

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

All maize data were previously published and described by Lorenz et al. (2015). In brief, a selection index comprised silage-quality traits was used to perform reciprocal recurrent selection. Traits comprising the index were yield, dry matter content, neutral detergent fiber (NDF), protein content, starch content, and *in-vitro* digestibility (www.cornbreeding.wisc.edu). Phenotypic data included five cycles of selection, encompassing approximately 20 generations in total. Tens to hundreds of individuals were sampled from each cycle of selection to be genotyped. Genotyping was performed with the MaizeSNP50 BeadChip, which includes 56,110 markers in total (Ganal *et al.* 2011). After removing monomorphic SNPs, redundant SNPs, quality filtering, and imputing, as described in Lorenz (2015), 10,023 informative SNPs remained.

Allele frequencies were computed for each cycle of selection. Because only 5 and 11 individuals from cycles 0 and 1 were genotyped, respectively, allele frequency change from cycle 2 (n = 163) to cycle 5 (n = 211) was computed for each SNP. Since all SNPs were di-allelic, the frequency of only one allele was tracked, and the frequency change for that allele perfectly mirrored the change for the other allele. For the tracked allele only, allelic effects were estimated using the R package RR-BLUP (Endelman 2011). Phenotypic information was available from individuals representing selection cycles 1 through 4. Therefore, a fixed effect for cycle was included in our model. Our exact analysis scripts are available at github.com/timbeissinger/ComplexSelection.

Chicken data:

Data were available for one white layer (WL) and one brown layer (BL) line from a commercial breeding program. Both closed lines have been selected over decades with a similar composite breeding goal, comprising, among others, laying rate, body weight and feed efficiency of the hens, as well as egg weight and egg quality, where the respective weights of the different traits varied between lines and over time. In total, 673 (743) WL (BL) individuals were genotyped, of which > 80% were from the last generation and the remaining animals were parents, grand-parents, and great-grandparents of the actual birds. For all genotyped individuals, complete pedigree data were available comprising 2109 (1879) individuals and going 13 (9) generations back in WL (BL). The oldest generation was defined as the base population and comprised 111 (64) ungenotyped individuals being separated from the majority of genotyped individuals by 12 (8) generations.

Current individuals were genotyped with the Affymetrix Axiom® Chicken Genotyping Array which initially carries 580K SNPs. This data were pruned by discarding sex chromosomes, unmapped

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

linkage groups, and SNPs with minor allele frequency (MAF) lower than 0.5% or genotyping call rate smaller than 97%. Individuals with call rates smaller than 95% were also discarded. Subsequently, missing genotypes at the remaining loci were imputed with Beagle version 3.3.2 (Browning and Browning 2009),resulting in sets of 277‘522 (334‘143) SNPs for the WL (BL) individuals.

To calculate the allele frequency change in the chicken populations, the allele frequency in the base population individuals had to be reconstructed by statistical means. This was done with the approach of Gengler *et al.* (2007), which, in short, considers the allele frequency in an individual as a quantitative and heritable trait and uses a mixed model approach to obtain a best linear unbiased prediction (BLUP) for the allele frequency of all un-genotyped individuals. This is done by linking the genotyped offspring to the un-genotyped ancestors via the pedigree information (for details, see Gengler *et al.* 2007). This required solving 277‘522 (334‘143) linear equation systems of dimension 2109 (1879) for the WL (BL) data set. Next, Δ_i for locus i was calculated as the difference of the observed allele frequency of the genotyped individuals in the current and the 3 ancestral generations and the average estimated allele frequency of the 111 (64) base population individuals 12 (8) generations back.

For each genotyped individual, conventional (non-genomic) BLUP breeding values and the respective reliabilities for a wide set of traits were available. SNP effects were estimated in a two-step procedure: first, for each trait in each line genomic breeding values were estimated via genomic BLUP (GBLUP), followed by a back-solution of estimated SNP effects. In the GBLUP step, the model $\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Zg} + \boldsymbol{e}$, was solved, where $\boldsymbol{y}$ is the vector of de-regressed proofs DRPs of genotyped individuals for a specific trait; μ is the overall mean; $\boldsymbol{g}$ is the vector of additive genetic values (i.e. genomic breeding values) for all genotyped chickens; $\boldsymbol{e}$ is the vector of residual terms; $\mathbf{1}$ is a vector of 1s and $\boldsymbol{Z}$ is a squared design matrix assigning DRPs to additive genetic values with dimension number of all genotyped individuals. Residual terms were assumed to be distributed $\boldsymbol{e} \sim N(0, \boldsymbol{R}\sigma_e^2)$, where $\boldsymbol{R}$ is a diagonal matrix with diagonal elements $R_{ii} = \frac{[c+(1-r_{DRPi}^2)/r_{DRPi}^2]h^2}{1-h^2}$ (Garrick *et al.* 2009) for an individual *i* in the training set, where r_{DRPi}^2 is the reliability of DRP for individual *i*, σ_e^2 is the residual variance, using *c* set to 0.1. The distribution of additive genetic values is assumed to be $\boldsymbol{g} \sim N(0, \boldsymbol{G}\sigma_g^2)$, where σ_g^2 is the additive genetic variance and $\boldsymbol{G}$ is a realized genomic relationship matrix which was constructed according to (VanRaden 2008). Estimation of variance components and genomic breeding values was done with ASReml 3.0 (Gilmour et al., 2009).

Next, estimated SNP effects $\hat{s}$ were obtained following Strandén and Garrick (2009) as

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

$$\hat{\boldsymbol{s}} = \frac{1}{2\sum_{1=1}^{m} p_i(1-p_i)} \boldsymbol{M}^T \boldsymbol{Z}^T \hat{\boldsymbol{g}}$$

where $\boldsymbol{M}$ is a matrix of dimension number of genotyped individuals x number of genotyped SNPs with entry $m_{ij} = x_{ij} - 2p_j$ where x_{ij} is the genotype of individual i at locus j (coded as 0, 1, or 2 which are counts of the reference allele) and p_j is the population frequency of the reference allele at SNP j.

Computational Resources:

Computation was performed using the University of Missouri Informatics Core Research Facility BioCluster (https://bioinfo.ircf.missouri.edu/). Computational nodes where simulations were performed had 64 cores and 512 GB of RAM.

Data availability:

Maize data are available in from Lorenz et al. (2015). Chicken data, including allele frequency change and estimated SNP effects, will be made available on Github or Figshare upon publication. All scripts used for simulations and analysis are available at github.com/timbeissinger/ComplexSelection.

Acknowledgements

We thank Natalia de Leon, Aaron Lorenz, and Lohmann GMBH for generating the maize and chicken biological data used in this study. We are grateful for helpful discussions with Emily Josephs and Aaron Lorenz. This research was supported by the USDA Agricultural Research Service, CRIS project number 5070-21000-038-00-D.

References

Akey J. M., 2009 Constructing genomic maps of positive selection in humans: Where do we go from here? Genome Res. 19: 711–722.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Barsh G. S., Farooqi I. S., O'Rahilly S., 2000 Genetics of body-weight regulation. Nature 404: 644–651.

Beissinger T. M., Wang L., Crosby K., Durvasula A., Hufford M. B., *et al.*, 2016 Recent demography drives changes in linked selection across the maize genome. Nat. Plants 2: 16084.

Berg J. J., Coop G., 2014 A Population Genetic Signal of Polygenic Adaptation. PLOS Genet. 10: e1004412.

Bernardo R., Thompson A. M., 2016 Germplasm Architecture Revealed through Chromosomal Effects for Quantitative Traits in Maize. Plant Genome 9.

Boyle E. A., Li Y. I., Pritchard J. K., 2017 An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169: 1177–1186.

Browning B. L., Browning S. R., 2009 A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. Am. J. Hum. Genet. 84: 210–223.

Cabrera-Bosquet L., Crossa J., Zitzewitz J. von, Serret M. D., Luis Araus J., 2012 High-throughput Phenotyping and Genomic Selection: The Frontiers of Crop Breeding ConvergeF. J. Integr. Plant Biol. 54: 312–320.

Campos G. de los, Hickey J. M., Pong-Wong R., Daetwyler H. D., Calus M. P. L., 2013 Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. Genetics 193: 327–345.

Endelman J. B., 2011 Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome 4: 250–255.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Evangelou E., Ioannidis J. P. A., 2013 Meta-analysis methods for genome-wide association studies and beyond. Nat. Rev. Genet. 14: 379–389.

Falconer D. S., Mackay T. F. C., 1996 *Introduction to Quantitative Genetics*. Pearson, Harlow.

Fisher R. A., 1918 The Correlation between Relatives on the Supposition of Mendelian Inheritance. Trans. Rocal Soc. Edinb. 52: 399–433.

Ganal M. W., Durstewitz G., Polley A., Bérard A., Buckler E. S., *et al.*, 2011 A Large Maize (Zea mays L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome. PLOS ONE 6: e28334.

Garrick D. J., Taylor J. F., Fernando R. L., 2009 Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41: 55.

Garud N. R., Messer P. W., Buzbas E. O., Petrov D. A., 2015 Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. PLOS Genet. 11: e1005004.

Gengler N., Mayeres P., Szydlowski M., 2007 A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. Anim. Int. J. Anim. Biosci. 1: 21–28.

Goddard M. E., Hayes B. J., 2009 Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10: 381–391.

Hansen M. E. B., Hunt S. C., Stone R. C., Horvath K., Herbig U., *et al.*, 2016 Shorter Telomere Length in Europeans than in Africans due to Polygenetic Adaptation. Hum. Mol. Genet.: ddw070.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Heffner E. L., Sorrells M. E., Jannink J.-L., 2009 Genomic Selection for Crop Improvement. Crop Sci. 49: 1–12.

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

Lawrie D. S., Messer P. W., Hershberg R., Petrov D. A., 2013 Strong Purifying Selection at Synonymous Sites in D. melanogaster. PLOS Genet. 9: e1003527.

Lorenz A. J., Beissinger T. M., Silva R. R., Leon N. de, 2015 Selection for Silage Yield and Composition Did Not Affect Genomic Diversity Within the Wisconsin Quality Synthetic Maize Population. G3 GenesGenomesGenetics: g3.114.015263.

Ma Y., Ding X., Qanbari S., Weigend S., Zhang Q., *et al.*, 2015 Properties of different selection signature statistics and a new strategy for combining them. Heredity.

Mathieson I., Lazaridis I., Rohland N., Mallick S., Patterson N., *et al.*, 2015 Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528: 499–503.

Meuwissen T. H. E., Hayes B. J., Goddard M. E., 2001 Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157: 1819–1829.

Orlando L., Gilbert M. T. P., Willerslev E., 2015 Reconstructing ancient genomes and epigenomes. Nat. Rev. Genet. 16: 395–408.

Plomin R., Haworth C. M. A., Davis O. S. P., 2009 Common disorders are quantitative traits. Nat. Rev. Genet. 10: 872–878.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Poland J. A., Balint-Kurti P. J., Wisser R. J., Pratt R. C., Nelson R. J., 2009 Shades of gray: the world of quantitative disease resistance. Trends Plant Sci. 14: 21–29.

Pritchard J. K., Pickrell J. K., Coop G., 2010 The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. Curr. Biol. 20: R208–R215.

Qanbari S., Simianer H., 2014 Mapping signatures of positive selection in the genome of livestock. Livest. Sci. 166: 133–143.

Rietveld C. A., Medland S. E., Derringer J., Yang J., Esko T., *et al.*, 2013 GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. Science 340: 1467–1471.

Sabeti P. C., Varilly P., Fry B., Lohmueller J., Hostetter E., *et al.*, 2007 Genome-wide detection and characterization of positive selection in human populations. Nature 449: 913–918.

Sargolzaei M., Schenkel F. S., 2009 QMSim: a large-scale genome simulator for livestock. Bioinformatics 25: 680–681.

Schaeffer L. R., 2006 Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123: 218–223.

Schweizer R. M., vonHoldt B. M., Harrigan R., Knowles J. C., Musiani M., *et al.*, 2016 Genetic subdivision and candidate genes under selection in North American grey wolves. Mol. Ecol. 25: 380–402.

Strandén I., Garrick D. J., 2009 Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. J. Dairy Sci. 92: 2971–2975.

Strandén I., Christensen O. F., others, 2011 Allele coding in genomic evaluation. Genet Sel Evol 43: 25.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Turchin M. C., Chiang C. W., Palmer C. D., Sankararaman S., Reich D., *et al.*, 2012 Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat. Genet. 44: 1015–1019.

VanRaden P. M., 2008 Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91: 4414–4423.

Villemereuil P. de, Frichot É., Bazin É., François O., Gaggiotti O. E., 2014 Genome scan methods against more complex models: when and how much should we trust them? Mol. Ecol. 23: 2006–2019.

Visscher P. M., Hill W. G., Wray N. R., 2008 Heritability in the genomics era — concepts and misconceptions. Nat. Rev. Genet. 9: 255–266.

Wallace J. G., Larsson S. J., Buckler E. S., 2014 Entering the second century of maize quantitative genetics. Heredity 112: 30–38.

Weir B. S., Cockerham C. C., 1984 Estimating F-Statistics for the Analysis of Population Structure. Evolution 38: 1358–1370.

Wisser R. J., Murray S. C., Kolkman J. M., Ceballos H., Nelson R. J., 2008 Selection Mapping of Loci for Quantitative Disease Resistance in a Diverse Maize Population. Genetics 180: 583–599.

Wood A. R., Esko T., Yang J., Vedantam S., Pers T. H., *et al.*, 2014 Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 46: 1173–1186.

Wright S., 1937 The Distribution of Gene Frequencies in Populations. Proc. Natl. Acad. Sci. U. S. A. 23: 307–320.

Yang J., Benyamin B., McEvoy B. P., Gordon S., Henders A. K., *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42: 565–569.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Yang J., Lee S. H., Goddard M. E., Visscher P. M., 2011 GCTA: A Tool for Genome-wide Complex Trait Analysis. Am. J. Hum. Genet. 88: 76–82.

Zeng J., Vlaming R. de, Wu Y., Robinson M., Lloyd-Jones L., *et al.*, 2017 Widespread signatures of negative selection in the genetic architecture of human complex traits. bioRxiv: 145755.

Gilmour AR, Gogel BJ, Cullis BR, Thompson R. ASReml User Guide 3.0. Hemel Hempstead, UK: VSN International Ltd; 2009.

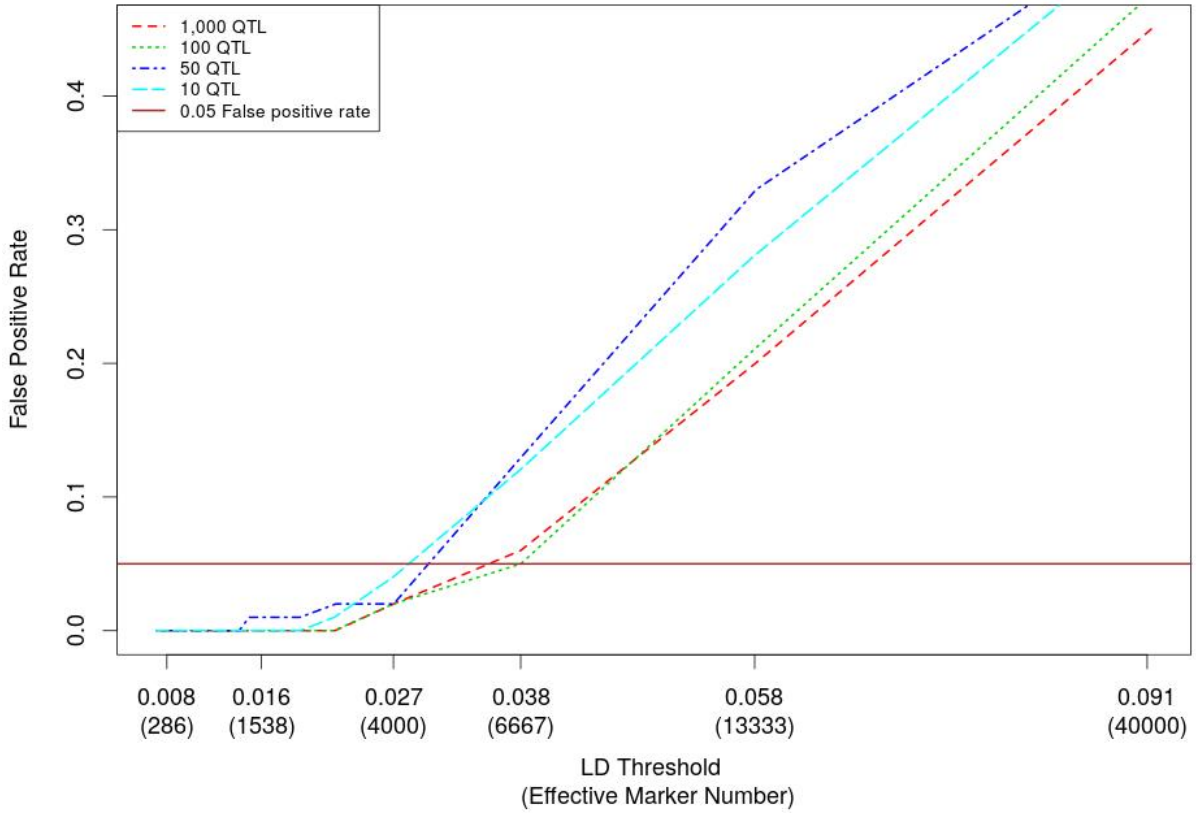
Endelman, J. B., 2011 Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4**:** 250-255.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Supplemental Table 1: Simulation parameters.

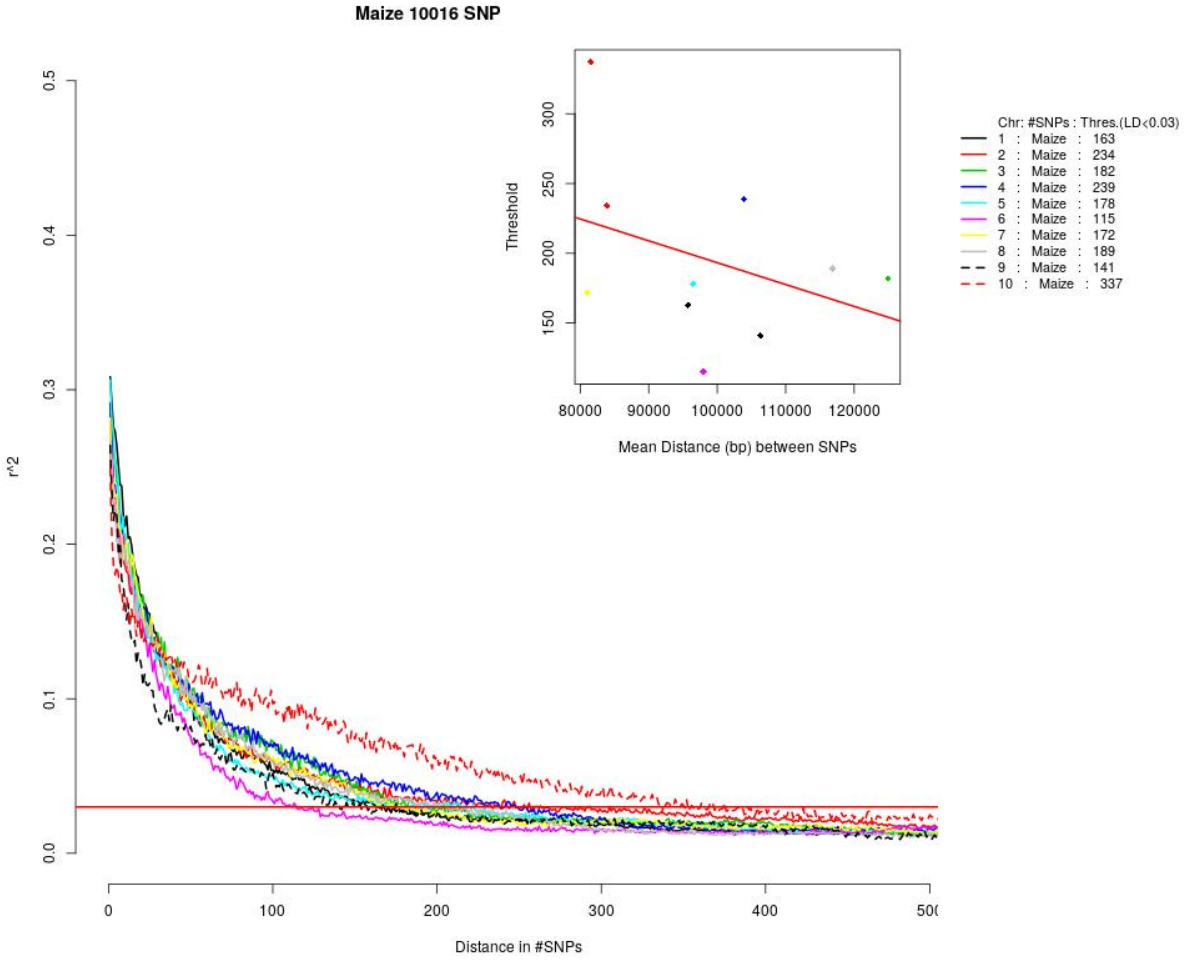
Parameter	Value(s)
Genetic basis of the trait	
Heritability	0.5
QTL Heritability	0.5 (all heritability attributable to QTL)
Phenotypic variance	1
Historical Population	
Population size	10,000
Number of generations	5,000
Marker mutation rate (only historical gens)	2.5e-5
QTL mutation rate (only historical gens)	2.5e-5
Breeding (selected) Population	
Number of selected males/generation	500
Number of selected females/generation	500
Litter size	10 (total pop size = 500*50 = 25,000)
Number of generations	20
Mating design	Random union of gametes, discrete generations
Genome	
Number of chromosomes	10
Chromosome size	100 cM
Markers/chromosome	10,000
Marker spacing	Even
Alleles/marker	2
Marker allele frequencies	Random (uniformly distributed)
Number of QTL	10, 50, 100, 1,000
QTL spacing	Even
Alleles/QTL	2
QTL allele frequencies (in first gen)	Equal (0.5)
QTL allele effects	Random (uniformly distributed)

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



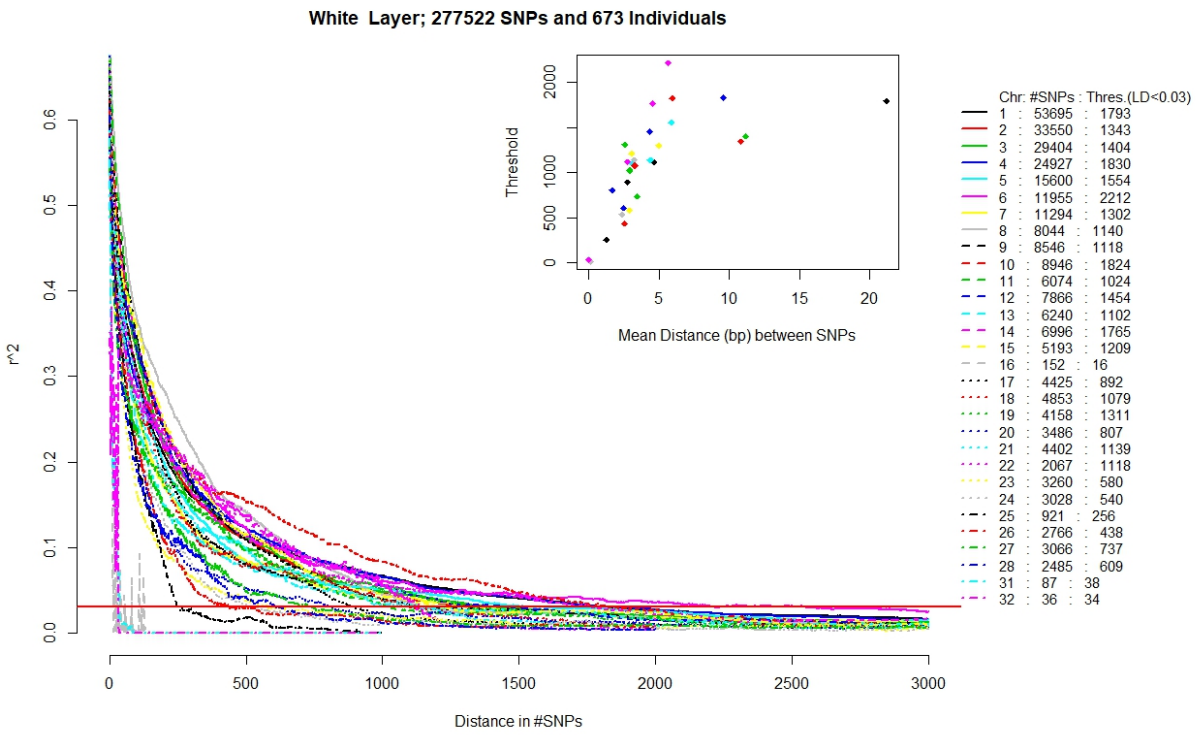
Supplemental Figure 1: False positive rate depends on the *number of effective markers*. The y-axis of this plot shows the false positive rate for simulations of different genetic architectures that was realized with varying effective numbers of markers. The x-axis depicts the mean LD-threshold across simulations that corresponded to a particular effective number of markers. Simulations suggested that defining the effective number of markers as the number of genome-segments such that LD across each segment is expected to be in the interval $R^2 \in [0.027, 0.038]$ appropriately controls false positive rate.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



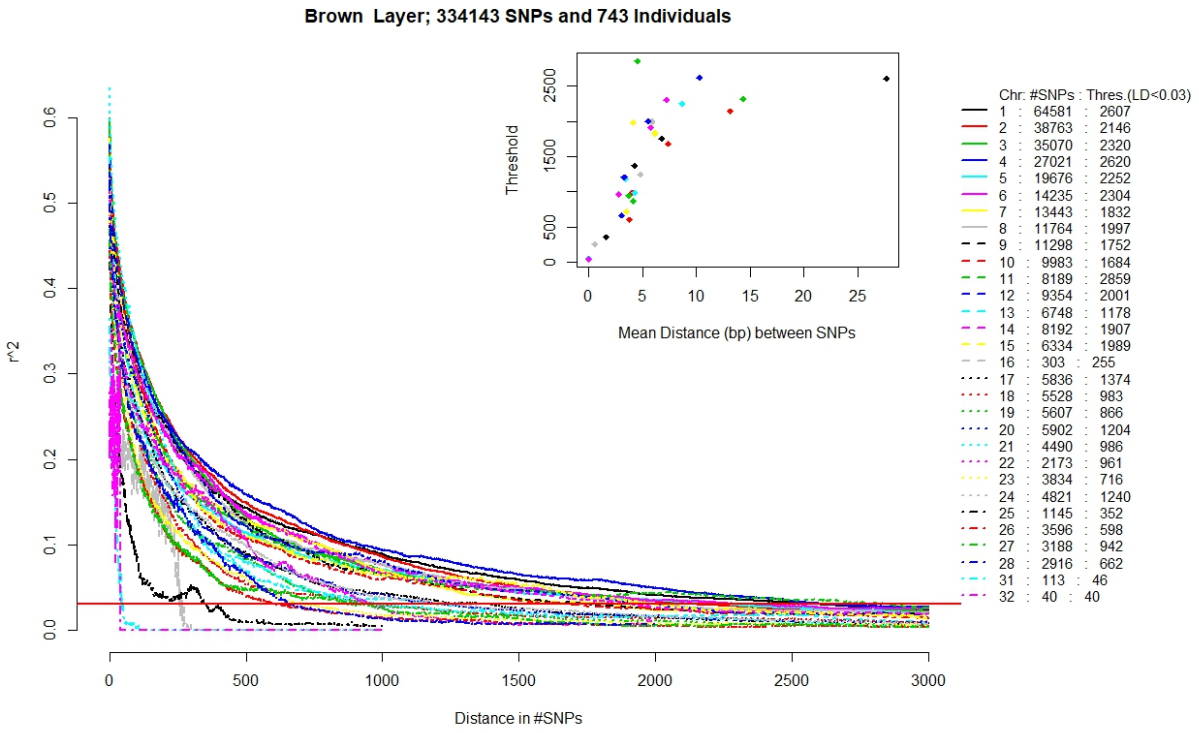
Supplemental Figure 2: LD Decay by chromosome in the WQS maize population. For each chromosome, LD is plotted against the distance between SNPs (in number of markers). The effective number of markers for our test was determined by dividing the total number of markers by the mean distance between markers such that $R^2 \leq 0.03$.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



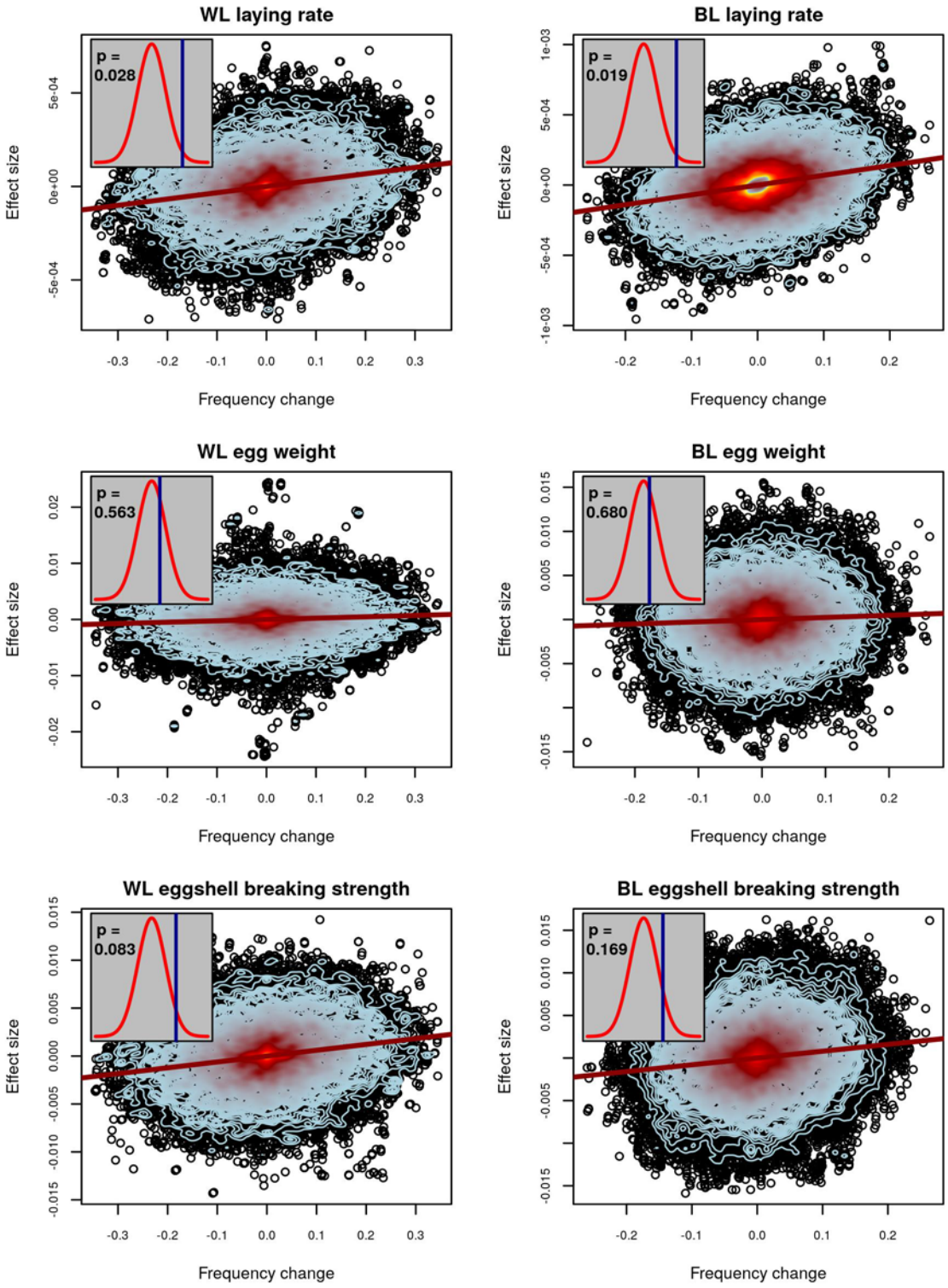
Supplemental Figure 3: LD Decay by chromosome in the White Layer chicken population. For each chromosome, LD is plotted against the distance between SNPs (in number of markers). The effective number of markers for our test was determined by dividing the total number of markers by the mean distance between markers such that $R^2 \leq 0.03$.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



Supplemental Figure 4: LD Decay by chromosome in the Brown Layer chicken population. For each chromosome, LD is plotted against the distance between SNPs (in number of markers). The effective number of markers for our test was determined by dividing the total number of markers by the mean distance between markers such that $R^2 \leq 0.03$.

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



Supplemental Figure 5: Evidence of selection for chicken traits, with potential outliers removed. This plot demonstrates a reanalysis of the chicken data shown in Figure 3 after removing of the 10 SNPs with the largest-magnitude effect size for each trait. For three traits in white (left column) and brown (right column) laying hens, the relationship between estimated allelic effects at individual SNPs and the

bioRxiv preprint doi: https://doi.org/10.1101/238295; this version posted December 21, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

change in allele frequency over generations is plotted. Contour lines indicate the density of points, with blue contours indicating fewer points than red. Inset plots depict observed values of $\hat{G}$ (blue lines) and their statistical significance based on a comparison to permuted null distributions (red densities) for no-selection scenarios. An exact two-sided p-value is given within each inset. Significant values of $\hat{G}$ above the permuted mean indicate selection operated in the positive direction, while significant values below the permutation mean indicated selection operated in the negative direction.