Ghat: An R package for identifying adaptive polygenic traits

Medhat Mahmoud[1,3¶*], Ngoc-Thuy Ha[2&], Henner Simianer[2,3&], and Timothy Beissinger[1,3¶*]

[1] Department of Crop Science, University of Goettingen, Goettingen, Germany

[2] Department of Animal Sciences, University of Goettingen, Goettingen, Germany

[3] Center for Integrated Breeding Research, University of Goettingen, Goettingen, Germany

* Corresponding authors
Medhat Mahmoud (mahmoud1@gwdg.de)
Timothy Beissinger (beissinger@gwdg.de)

¶These authors contributed equally to this work.
&These authors also contributed equally to this work.

Abstract

Identifying selection on polygenic complex traits in crops and livestock is key to understanding evolution and helps prioritize important characteristics for breeding. However, the QTL that contribute to polygenic trait variation exhibit small or infinitesimal effects. This hinders the ability to detect QTL controlling polygenic traits because enormously high statistical power is needed for their detection. Recently, we circumvented this challenge by introducing a method to identify selection on complex traits by evaluating the relationship between genome-wide changes in allele frequency and estimates of effect-size. The method involves calculating a composite-statistic across all markers that captures this relationship, followed by implementing a linkage disequilibrium-aware permutation test to evaluate if the observed pattern differs from that expected due to drift during evolution and population stratification. In this manuscript, we describe "Ghat", an R package developed to implement the method to test for selection on polygenic traits. We demonstrate the package by applying it to test for polygenic selection on 15 published European winter wheat traits including yield, biomass, quality, morphological characteristics, and disease resistance traits. The results highlight the power of Ghat to identify selection on complex traits. The Ghat package is accessible on CRAN, The Comprehensive R Archival Network, and on GitHub.

Introduction

Many important traits in plants, animals, and humans are of polygenic nature and depend on the cumulative effects of many loci, each contributing a small proportion of the total genetic variation [1]. Studying the evolution and selection history of such traits can be extremely challenging due to the small effects of the quantitative trait loci (QTL) that contribute to genetic variation. However, regardless of effect size, the theory of natural selection tells us that alleles with positive effects on fitness will tend to increase in frequency over successive generations [2]. A number of statistical methods have been developed to identify loci carrying such beneficial alleles. These methods were reviewed in [3], and include approaches relying on polymorphism variability, haplotype frequency, linkage disequilibrium, allele frequency change over time, and techniques that combine some or all of these signals. However, in the case of highly-polygenic traits, it is difficult or impossible to identify regions carrying individually-selected alleles/haplotypes due to their small effects, regardless of which sophisticated methodologies are leveraged. For instance, it has been shown that selected loci are difficult to identify when per-locus selection intensity is low [3], as is the case with low heritability traits and those controlled by many alleles of small effect [4]. This imposes serious limitations to the previous methods when identifying selection on polygenic traits (e.g. [5]).

Recently, large sample sizes utilized in human genetic studies have allowed more powerful understanding of the genetic architecture of polygenic traits using genome-wide association studies (GWAS) [6–8]. After conceptual work solidified the importance of polygenic adaptation [9,10], researchers began placing more of an emphasis on identifying this phenomenon using an infinitesimal approach [11–14]. For example, Zeng et al recently developed a Bayesian mixed linear model (BayesS) that detected natural selection on 23 complex traits in humans, based on the relationship between SNP effect sizes and minor allele

frequency [15]. However, it turns out that results from these methods, particularly those relating to selection on human height, are highly sensitive to the confounding effects of population stratification [16,17]. Additional methods recently developed to detect polygenic adaptation include SBayesR [18], which is an extension of the Bayesian multiple regression model (BayesR) [19] to utilize the summary statistics from GWAS instead of using individual-level genotype data, and an approach by Josephs et al [20] which performs principal component analysis on the relatedness matrix to detect polygenic adaptation.

Another test, denoted $\hat{G}$ (or Ghat), was recently developed by Beissinger et al [21] to identify selection on polygenic traits. The idea behind this test is to combine information from all loci simultaneously to evaluate whether there is a trend for loci contributing to a phenotype to collectively show evidence of selection. Thereby, signals that are individually insignificant for selection (at the locus level) may become highly significant when evaluated at the genome-wide scale (at the level of finding selection on a trait). The test depends on the relationship between changes in allele frequency over time and estimated additive effects for every genotyped locus.

We have developed and released an R [22,23] package called "Ghat" to implement the $\hat{G}$ test. The package allows users to test for selection and evolution in both qualitative and quantitative traits, and as shown by [21], it is particularly powerful when testing for selection on highly polygenic traits. For the purpose of demonstrating the package, in this manuscript we applied the method and corresponding Ghat package to 15 different European winter wheat traits from a collection of 191 wheat cultivars, registered in Europe between 1966 and 2013, which were published by Voss-Fels et al. [24], demonstrating the historical adaptation of one of the most important crops worldwide.

Implementation of the Ghat Package

Overview

Ghat identifies selection using all genome markers/sequences information [21]. This makes it possible to identifying selection in complex traits that are controlled by many genes with small effect. First, all allelic effects are estimated using a genomic prediction approach. Next, allele frequencies are estimated in two or more different generations (e.g. generation 0 and 10). Third, the rate of linkage disequilibrium decay is estimated to approximate the number of independent genome segments in the study population. Finally, an LD-aware permutation test is implemented to evaluate if there is evidence of significant selection between generations, and in what direction selection operated.

Installation

The Ghat package is installed within the R terminal (Fig1, step-1) by typing: `>install.packages("Ghat")` and then loading to package into R by typing: `>library(Ghat)`. In addition, the package has the following optional R-package dependencies:

- Parallel-Package {parallel}: Support for parallel computation, to speed the computation time in large data analysis. [25]
- rrBLUP-Package {rrBLUP}: Ridge regression for estimating marker effects (RR-BLUP). [26]

Optional dependencies that are not automatically installed with the package, include rrBLUP [26] and parallel [25]. These maybe installed and loaded before running Ghat by typing: `>install.packages(c("parallel", "rrBLUP"))` and then typing `>library(parallel)` and `>library(rrBLUP)` (Fig1, step-1).

Input Data format

As inputs, Ghat requires two vectors. The first of these is a vector of allele substitution effects, and the second is a vector of the corresponding change in the frequency of the substituted allele between two different generations (Table 1). Required parameters are, "method": to specify how the Ghat permutation test should account for LD between markers ("scale" is typically the most appropriate method to implement), "perms": for the number of permutations to perform, "plot": to specify whether or not simple Ghat plots should be output, "blocksize": for setting the size of blocks that should be used for LD-trimming (only required when method = "trim"), and "num_eff": for setting the effective number of independent segments of the genome (only required when method = "scale"). All parameters details are in Table 1.

Establishing the number of effective markers

Treating each marker as an independent observation can lead to overestimated test statistics. This is because allele frequencies and effect estimates from markers that are in LD are not independent [21]. Any level of LD between markers produces less independent observations than the total number of markers. In Ghat, we provide a function called "ld_decay" that can estimate the LD decay across all genome and use this information to approximate the total number of independent genome segments, or "genotype blocks". The following command runs the ld_decay function `>Ld <- ld_decay(gen, map, max_win_snp, max.chr, cores=1, max_r2)` (Fig 1, Step-3). The output information is used by Ghat to scale the variance of the permutation test statistic according to the actual number of independent markers (Fig 1, Step-4).

Running the test

Before running Ghat, one needs to either estimate the allele substitution effects and allele frequencies in at least to different generations, or get that information from some already published analysis i.e. publicly available summary association statistics [27]. To estimate allele substitution effects and allele frequencies, Ghat provides some simple examples to compute both from individual row genotype data. Next, to perform the Ghat test with this information, only one line of code is required: `>Ghat_res <- Ghat(effects, change, method, perms, plot, num_eff)` (Fig 1, Step-4).

Testing for selection using The Wisconsin Quality Synthetic (WQS) data object.

To illustrate the use of the Ghat package, we used a previously published maize population dataset (Fig 1) as a case of polygenic adapted species [5]. A very brief summary of the selection protocol is that to improve silage quality, the researchers performed reciprocal recurrent selection using a selection index calculated from multiple traits related to maize quality, including a trait called Acid Detergent Fiber (ADF) (http://www.cornbreeding.wisc.edu). The use of Ghat to analyse selection on ADF in the WQS Maize population can be summarized in four steps, which we depict in Fig 1.

Fig 1. Four steps to go from raw individual marker data to results from the Ghat test for selection. An illustrated example using the "Maize_wqs" data available in the Ghat package. **Step-1**: install the package, load the dependencies and "Maize_wqs" data and estimate allele substitution effects (e.g., by using rrBLUP [26]). **Step-2**: calculate the differences in allele frequency between two different generations (cycle 1 and cycle 3). **Step-3**: estimate the decay of linkage disequilibrium (LD) to calculate the effective number of independent markers. **Step-4**: calculate the Ghat value and p-value from the permutation test.

All functions, options and data set available in Ghat are provided in Table 1

Table 1. Objects, data, functions and options found in Ghat package.

Function		**Description**
Ghat (Function)	effects	Input vector of allele substitutional effects.
	change	Input vector of changes in allele frequency between two different generations (could be positive, negative or zero).
	method	Input option, "vanilla" (assumes complete linkage equilibrium between markers), "trim" (excludes markers to approximate linkage equilibrium some of the extreme values) or "scale" (scales results to reflect underlying levels of linkage LD).
	perms	Input number to set the number of permutations to test for significance.
	plot	Input option to select which plot you need to be returned: "Ghat", "Cor", or "Both".
	blockSize	Input number to set the size of the window on each chromosome for trimming, only required if method = "trim".
	num_eff	Input number to set the effective number of independent variants, only required if method = " scale". Can be calculated using "ld_decay" function above.
ld_decay (Function)	gen	Input matrix of genotype data. Individuals in rows, genotypes (0, 1, 2) in columns.
	Map	Input data frame including the name for each marker with a corresponding
	max_win_snp	Input number to set the maximum number of markers allowed per window within a chromosome before estimating the LD. The default value is 2000.
	max.chr	Input number to select to which chromosome you want to run the analysis, chromosomes above this number will be excluded from the analysis.
	cores	Input number to set the optimal number of cores for using parallelized calculation, since calculation of LD require too much memory. The default value is 1 for windows machine.
	max_r2	Input number to set the threshold of correlation between the neighbourhood variant (r^2) to calculate the effective number of independent variants.
Maize_wqs (Data object)	Maize_wqs[[1]]	Data frame, including all SNP names (SNP Names), effects on Acid detergent fiber (adf) (effects.adf), and changes in allele frequencies (change) between cycle 1 and cycle 3 of selection.
	Maize_wqs[[2]]	Data frame, including plant ID's (taxa), estimated breeding values for Acid detergent fiber (adf), and generation when the phenotyped strains were grown (year).
	Maize_wqs[[3]]	Data frame, map file, each line of the Map file describes a single marker and contain three columns. 1: Name (SNP id); 2: Chromosome (Chromosome number); 3: SNP position (in base-pairs).
	Maize_wqs[[4]]	Data frame, Genotype (Illumina MaizeSNP50 BeadChip); an Infinium HD assay (Illumina, Inc. San Diego, CA). 10,017 SNP markers (0, 1 and 2) after filtration, distributed across the maize genome [28].

Methods

Testing for selection during 50 years of wheat breeding

To demonstrate the power and flexibility of the Ghat package, we used it to test for selection during the past 50 years of wheat breeding in Europe. To achieve this, we leveraged a large dataset representing 50 years of commercial European winter wheat varieties that was previously published by Voss-Fels and colleagues [24]. Our analysis methods and results are described below.

Genotypes and phenotypes. A data set of 191 cultivars, registered in Europe between 1966 and 2013, was generated and previously published by [24]. In total, 19 traits were considered and used to evaluate and quantify the phenotypic progress of wheat breeding during the last five decades. However, we used only 15 traits after filtering based on data quality. Spikes per sqm, Heading, Sedimentation, Falling number, and GCD were removed from the analysis because of missing some phenotypes information (>70% of total phenotypes were missing). All traits were tested in three different environments by [24]. The environments were: 1: A high-intensity nitrogen supply along with best-practices fungicide, insecticide, and growth regulator applications (HiN/HiF treatment), 2: A high level of nitrogen fertilization with a fungicide-free treatment (HiN/NoF), and 3: A low level of nitrogen fertilization with a fungicide-free treatment (LoN/NoF). The evaluation was based on five main groups of traits: **1- Yield parameters traits**, including: Grain yield, Harvest index, Radiation use efficiency and Radiation interception efficiency; **2- Biomass parameter traits**, including: Above-ground Biomass, Kernels per sqm, and Thousand-kernel weight. **3- Quality parameter traits**, including: Crude protein, Protein yield and Nitrogen use efficiency. **4- Morphological traits**, including: Plant height, Green canopy duration and Kernel per spike. **5- Disease resistances traits**, including: Resistance to powdery mildew and Resistance to stripe rust. All cultivars were genotyped with a 15K SNP Illumina Infinium iSelect genotyping array [29] for

195 by [24], and after filtering 8,710 SNPs passed the genotype quality control criteria [24] and
196 were included in their subsequent analyses. In our analysis, we re-analyzed the exact set of
197 SNPs that was published by [24], as described above.

198 **Phenotypic progress.** As technical validation, we re-estimated the phonotypic progress of the
199 15 wheat traits over 50 years of selection and adaptation using simple *Pearson* correlation
200 [30].

201 **Ghat.** Allelic effects at every marker locus for the 15 traits were estimated using rrBLUP
202 [26], and changes in allele frequencies were calculated with R [22,23]. Ghat is calculated as
203 the summation of the estimated effect of every SNP scored multiplied by its effect size,
204 according to

205 $$\hat{G} = \sum_{j=1}^{m} \Delta_j \alpha_j \ , \qquad (1)$$

206 where Δ_j is the frequency change in locus *j* between two time-points and α_j is the additive
207 effect of locus *j.* We implemented two different analyses with Ghat to test for selection, which
208 enabled us to measure the efficiency of the test in different situations. In the first analysis,
209 which we call "**All Phenotypes"**, allelic effects were estimated using phenotypic information
210 from all available cultivars in the Voss-Fels study, which represent the last 50 years of wheat
211 breeding. This represents the best-case Ghat analysis, when all phenotypic information is
212 available. In the second analysis, which we call **"Modern Phenotypes",** allelic effects were
213 estimated using only modern phenotypic information (from 2010 to 2013), while changes in
214 allele frequency were still calculated based on genotypes from opposite ends of the 50-year
215 dataset. This represents a realistic use-case for Ghat, when the experimenter has all genotypic
216 information, but only phenotypic information from the present. This represents a situation
217 where reliably and comparably measured historical phenotypes no longer exist, but historical
218 tissue or seed are available and can be used for DNA collection.

Results

Testing for selection during 50 years of wheat breeding

The main theory behind this test for polygenic adaptation/selection is that alleles at every locus spread across the genome may contribute a small proportion of the total variance for each trait (infinitesimal model), regardless whether or not a mapping study has enough power to significantly associate markers with traits. To demonstrate the Ghat package, we used it to test for selection on the 15 wheat traits and replicated for two different analysis mentioned above, enabling us to establish how much power is lost when phenotypic data are only available from modern lines.

Phenotypic progress. After estimating the Phenotypic trend (correlation between phenotypic value and the registration year) (Fig 2). We can clearly classify the 15 traits into three groups. The first group is comprised of traits that increased in value during breeding. These were above-ground plant biomass, grain yield, green canopy duration, harvest index, kernels per spike, kernels per square meter, nitrogen use efficiency, protein yield, radiation interception efficiency, radiation use efficiency, resistance to powdery mildew, and resistance to stripe rust. Next, we identified a group of traits with a decreasing phenotypic trend, which was comprised of plant height and crude protein. Finally, thousand-kernel weight showed no significant trend in either direction.

Ghat test for selection. Fig 2 shows the performance of Ghat when testing for selection using genotypes and phenotypic information from lines produced during all 50 years of wheat breeding and comparing it with the actual phenotypic progress of these lines for the past 50 years (phenotypic trend). First, we investigated the precision and accuracy of the Ghat in estimating the size and direction of selection (using Ghat-correlation) when all phenotypic information were available (50 years). Overall, the selection size and direction on the 15 traits

using all phenotypes were in the highest agreement with the phenotypic trend. Next, we evaluated Ghat results when the phenotypic information were only available for the past 3 years (from 2010 to 2013). Notably, this was also in a strong agreement with the phenotypic trend. Ghat successfully detected the selection history of each trait, even when only modern phenotypes were utilized. For example, Ghat found that Plant height and Crude protein traits were under intensive negative selection, which agrees with their phenotypic trends during the last 50 years. Ghat showed very weak selection for thousand-kernel weight, which again aligned with this trait's phenotypic trend. The rest of the traits were under high to moderate positive selection. To validate the results of Ghat on wheat data, a *spearman* correlation between the Ghat results from 3 years and 50 years of phenotypes were estimated. A very high significant ($P<0.001$) correlation of 0.57 was found between Ghat values from 3 and 50 years of phenotypes.

Fig 2. The power of Ghat test of selection in European winter wheat under three agrochemical treatments. 1: A high-intensity nitrogen supply along with best-practices fungicide, insecticide, and growth regulator applications (HiN/HiF), 2: A high level of nitrogen fertilization with a fungicide-free treatment (HiN/NoF), and 3: A low level of nitrogen fertilization with a fungicide-free treatment (LoN/NoF). **Left**: the phenotypic trend over the last 50 years. **Right**: the Ghat test of selection based on two analyses. "All Phenotypes" refers to estimating allelic effects using all available phenotypic information over the 50 years of wheat breeding; "Modern Phenotypes" refers to estimating allelic effects using modern phonotypic information only (2010 to 2013). Blue, positive selection; red, negative selection. Large light ellipses correspond to weaker selection intensity; thin dark ellipse correspond to more selection intensity. P-values printed on each cell on the left side of the figure (phenotypic trend section) correspond to the significance of the phenotypic trend. P-values printed on each cell on the right side of the figure (Ghat test section) correspond to implementing the Ghat permutation test.

Adaptation history in the European winter wheat.

To more precisely dissect the history of selection in wheat, we divided the 50 years of breeding into 10 adaptation periods and applied Ghat successively between each period. We excluded the green canopy duration trait from this analysis since it was only measured under one environmental condition. For the remaining 14 traits, we calculated the average phenotypic value across the three different environmental conditions. For example, during the period from 1966 to 1975, significantly selected traits according to Ghat included grain yield, harvest index (Fig 3-A), above-ground plant biomass, thousand-kernel weight (Fig 3-B), crude protein, nitrogen use efficiency (Fig 3-C), kernel per spike and resistance to strip rust (Fig 3-D). However, during the same period a selection operated significantly in the negative direction against radiation use efficiency (Fig 3-A), thousand-kernel weight (Fig 3-B), plant height and resistance to powdery mildew (Fig 3-D). Moreover, by following to the time series of the selection over the past 50 years (Fig 3) we can get a complete picture of how breeders adapted the 191 wheat cultivars through time, by focusing on certain traits in some time periods and other traits in other time periods.

Fig 3. Ghat correlation estimates of the selection size and direction for 50 years of wheat breeding. Results are for 14 productivity traits in European winter wheat measured during the past 50 years. Points above zero level (red dashed lines) indicate positive selection, while points under zero indicate negative selection; bold lines represent 99% confidence interval. A- Adaptation history for the **Yield parameters traits**. B- Adaptation history for the **Biomass parameters traits**. C- Adaptation history for **Quality parameters traits**. D- Adaptation history for **Morphological and disease resistance traits**.

Relative selection pressure

Relative selection pressure (RSP) can defined as: The average of absolute changes in allele frequency of the alleles that controlling traits. RSP is genetic property of biological traits. Like heritability, it has a specific value for each trait and varies from zero to one. RSP is a measurement of how allele frequencies of the effective variants changes per year or more under natural/artificial selection [31,32]. We estimated the RSP in this study by calculating the average of all absolute values of Ghat correlation every 5 years. In the case of the European winter wheat data, we estimated the RSP for the past 50 years and we plotted the accumulative selection pressure on these 14 traits in Fig 4. In addition, the actual RSP of the 14 traits listed in Table 2. RSP quantifies how elastic each trait is elastic to the breeding decisions regardless of the selection direction (positive or negative). One notable example of the response to selection in the evolutionary process, is the "Above the ground biomass" trait, with RSP of 0.085 this trait starting from close to zero selection on 1970 to near to 0.75 of accumulative selection. On the other hand, "Thousand-kernel weight" trait did not show high response to selection, since the increase of accumulative selection size did not increased compared to the other traits, and it has the lowest RSP in the list (Table 2)

Table 2. Relative selection pressure (RSP) for 14 traits of European winter wheat. RSP were estimated by calculating the average of the absolute Ghat correlation every 5 years for 50 years.

Traits	Relative selection pressure (RSP)
Above-ground plant biomass	0.086
Crude protein	0.085
Grain yield	0.076
Harvest index	0.061
Kernel per spike	0.090
Kernels per sqm	0.100
Nitrogen use efficiency	0.079
Plant height	0.064
Protein yield	0.079
Radiation interception efficiency	0.097
Radiation use efficiency	0.067
Resistance to powdery mildew	0.069
Resistance to stripe rust	0.084
Thousand-kernel weight	0.050

Overall, the average value of RSP in wheat traits is 0.078 (Table 1), the highest was 0.10 in Kernels per sqm, reveals that this traits is most affected traits of in wheat breeding may be because of a direct selection on the traits or a high genetic correlation with other important traits.

Fig 4. Cumulative selection in European winter wheat. Results are for the 14 traits of wheat during the past 50 years ago.

Testing and system requirements

We have released three versions of Ghat package in CRAN (https://cran.r-project.org/package=Ghat) for the three major operating systems: MS Windows, Linux/Unix, and Mac OS. We have also released the source code on GitHub (https://github.com/timbeissinger/ComplexSelection) so that users can compile for specific platforms or modify the code as needed.

For users who have already computed allele effect and allele frequencies changes and are applying the Ghat test to these data, the computational cost of Ghat is unnoticeable. Practically, estimating Ghat for maize data [5] with 1000 permutations takes one second per one phenotype on a 2019 Dell XPS 23 with 2.4 Intel Core i7 processor.

Availability and future directions

We have presented a new package for identifying selection and adaptation from GWAS or Genomic prediction summary statistics, which we applied to a dataset of Wheat plant after calculating their rrBLUP summary statistics of 15 traits and 119 cultivars phenotype measurements. We reported several new findings that would have been difficult or impossible to obtain with the traditional methods of scanning for signatures of selection at

342 particular loci. These findings including historical adaptation and selection direction on the
343 most important 15 traits of Wheat plant.

344 A cause of false positive findings in any genome wide analysis is hidden population
345 structure [33]. In additional to the importance of considering population and environmental
346 effects in the estimation model, an effective way to solve this problem is to run principal
347 component analyse and using the most important PC-axis as additional fixed effects in
348 estimation model before running Ghat [34].

349 Unlike methods for detecting selection, Ghat can quantify selection magnitude and
350 direction even if we do not have any phenotypes from before selection (generation zero). This
351 can be achieve by estimating allele effect by the current population phenotype and estimation
352 allele frequencies from the ancient population by genotyping some ancient tissues that are
353 available in i.e. gene-banks or from other resources. Moreover, we can still approximate allele
354 frequencies from pre-selection generations even without tissue by using current allele
355 frequencies coupled with known pedigree information [35].

356 For breeders and breeding organizations, Ghat represents a practical solution for
357 tracing selection and adaptation history based on their recorded traits. Furthermore, RSP
358 values can be used to understand how each trait responds to selection and gives breeders an
359 indication of how traits may response to any future selection pressure.

360 The Ghat test and implementation via the package described herein does not require
361 immense sample size, individual genotypes and/or genome-wide significant SNPs, as
362 demonstrated by our implementation with a relatively modest sample size of 191 wheat
363 individuals genotyped for 8,710 SNPs. Notably, this approach is more powerful when traits
364 are controlled by large number of QTLs (Beissinger et at., 2018). These advantages allow the
365 Ghat package to be used for identifying selection for various complex traits in other species of
366 plants, animals and in humans. The package is available through CRAN (https://cran.r-

project.org/package=Ghat) and on GitHub (https://github.com/timbeissinger/ComplexSelection).

Acknowledgements

This project was supported with funds from the University of Göttingen, Faculty of Agriculture, Center for Integrated Breeding Research, and Division of Plant Breeding Methodology. We are indebted to the authors of [24] for making their data publicly available.

377 # References

378 1. Fisher RA. XV.—The Correlation between Relatives on the Supposition of Mendelian
379 Inheritance. Earth Environ Sci Trans R Soc Edinb. 1919;52: 399–433.
380 doi:10.1017/S0080456800012163

381 2. Wright S. Evolution in Mendelian Populations. Genetics. 1931;16: 97–159.

382 3. Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H. Properties of different
383 selection signature statistics and a new strategy for combining them. Heredity.
384 2015;115: 426–436. doi:10.1038/hdy.2015.42

385 4. Beissinger T, Kruppa J, Cavero D, Ha N-T, Erbe M, Simianer H. A Simple Test Identifies
386 Selection on Complex Traits. Genetics. 2018;209: 321–333.
387 doi:10.1534/genetics.118.300857

388 5. Lorenz AJ, Beissinger TM, Silva RR, de Leon N. Selection for Silage Yield and
389 Composition Did Not Affect Genomic Diversity Within the Wisconsin Quality Synthetic
390 Maize Population. G3 GenesGenomesGenetics. 2015;5: 541–549.
391 doi:10.1534/g3.114.015263

392 6. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of
393 common variation in the genomic and biological architecture of adult human height.
394 Nat Genet. 2014;46: 1173–1186. doi:10.1038/ng.3097

395 7. Visscher PM, Hemani G, Vinkhuyzen AAE, Chen G-B, Lee SH, Wray NR, et al. Statistical
396 Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated
397 Samples. PLOS Genet. 2014;10: e1004269. doi:10.1371/journal.pgen.1004269

398 8. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of
399 GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101: 5–22.
400 doi:10.1016/j.ajhg.2017.06.005

401 9. Pritchard JK, Di Rienzo A. Adaptation – not by sweeps alone. Nat Rev Genet. 2010;11:
402 665–667. doi:10.1038/nrg2880

403 10. Pritchard JK, Pickrell JK, Coop G. The Genetics of Human Adaptation: Hard Sweeps, Soft
404 Sweeps, and Polygenic Adaptation. Curr Biol. 2010;20: R208–R215.
405 doi:10.1016/j.cub.2009.11.055

406 11. Hancock AM, Alkorta-Aranburu G, Witonsky DB, Di Rienzo A. Adaptations to new
407 environments in humans: the role of subtle allele frequency shifts. Philos Trans R Soc B
408 Biol Sci. 2010;365: 2459–2468. doi:10.1098/rstb.2010.0032

409 12. Berg JJ, Coop G. A Population Genetic Signal of Polygenic Adaptation. W. Feldman M,
410 editor. PLoS Genet. 2014;10: e1004412. doi:10.1371/journal.pgen.1004412

411 13. Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, Golan D, et al. Detection of human
412 adaptation during the past 2000 years. Science. 2016;354: 760–764.
413 doi:10.1126/science.aag0776

414 14. Csilléry K, Rodríguez-Verdugo A, Rellstab C, Guillaume F. Detecting the genomic signal
415 of polygenic adaptation and the role of epistasis in evolution. Mol Ecol. 2018;27: 606–
416 612. doi:10.1111/mec.14499

417 15. Zeng J, Vlaming R de, Wu Y, Robinson MR, Lloyd-Jones LR, Yengo L, et al. Signatures of
418 negative selection in the genetic architecture of human complex traits. Nat Genet.
419 2018;50: 746–753. doi:10.1038/s41588-018-0101-4

420 16. Sohail M, Maier RM, Ganna A, Bloemendal A, Martin AR, Turchin MC, et al. Polygenic
421 adaptation on height is overestimated due to uncorrected stratification in genome-
422 wide association studies. Nordborg M, McCarthy MI, Nordborg M, Barton NH,
423 Hermisson J, editors. eLife. 2019;8: e39702. doi:10.7554/eLife.39702

424 17. Berg JJ, Harpak A, Sinnott-Armstrong N, Joergensen AM, Mostafavi H, Field Y, et al.
425 Reduced signal for polygenic adaptation of height in UK Biobank. Nordborg M,
426 McCarthy MI, Nordborg M, Barton NH, Hermisson J, editors. eLife. 2019;8: e39725.
427 doi:10.7554/eLife.39725

428 18. Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, et al. Improved
429 polygenic prediction by Bayesian multiple regression on summary statistics. Nat
430 Commun. 2019;10: 1–11. doi:10.1038/s41467-019-12653-0

431 19. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, et al. Improving
432 accuracy of genomic predictions within and between dairy cattle breeds with imputed
433 high-density single nucleotide polymorphism panels. J Dairy Sci. 2012;95: 4114–4129.
434 doi:10.3168/jds.2011-5019

435 20. Josephs EB, Berg JJ, Ross-Ibarra J, Coop G. Detecting Adaptive Differentiation in
436 Structured Populations with Genomic Data and Common Gardens. Genetics. 2019;211:
437 989–1004. doi:10.1534/genetics.118.301786

438 21. Beissinger T, Kruppa J, Cavero D, Ha N-T, Erbe M, Simianer H. A Simple Test Identifies
439 Selection on Complex Traits. Genetics. 2018;209: 321–333.
440 doi:10.1534/genetics.118.300857

441 22. R: The R Project for Statistical Computing. [cited 2 Jan 2020]. Available: https://www.r-
442 project.org/

443 23. Beissinger T. timbeissinger/ComplexSelection. 2019. Available:
444 https://github.com/timbeissinger/ComplexSelection

445 24. Voss-Fels KP, Stahl A, Wittkop B, Lichthardt C, Nagler S, Rose T, et al. Breeding improves
446 wheat productivity under contrasting agrochemical input levels. Nat Plants. 2019;5:
447 706–714. doi:10.1038/s41477-019-0445-5

25. Ooi H, Corporation M, Weston S, Tenenbaum D. doParallel: Foreach Parallel Adaptor for the “parallel” Package. 2019. Available: https://CRAN.R-project.org/package=doParallel

26. Endelman JB. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome. 2011;4: 250–255. doi:10.3835/plantgenome2011.08.0024

27. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. Nat Rev Genet. 2017;18: 117–127. doi:10.1038/nrg.2016.142

28. Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, et al. 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. 2011;6: e28334. doi:10.1371/journal.pone.0028334

29. Muqaddasi QH, Brassac J, Börner A, Pillen K, Röder MS. Genetic Architecture of Anther Extrusion in Spring and Winter Wheat. Front Plant Sci. 2017;8. doi:10.3389/fpls.2017.00754

30. Benesty J, Chen J, Huang Y, Cohen I. Pearson Correlation Coefficient. In: Cohen I, Huang Y, Chen J, Benesty J, editors. Noise Reduction in Speech Processing. Berlin, Heidelberg: Springer; 2009. pp. 1–4. doi:10.1007/978-3-642-00296-0_5

31. Avgar T, Lele SR, Keim JL, Boyce MS. Relative Selection Strength: Quantifying effect size in habitat- and step-selection inference. Ecol Evol. 2017;7: 5322–5330. doi:10.1002/ece3.3122

32. Coleman ML, Chisholm SW. Ecosystem-specific selection pressures revealed through comparative population genomics. Proc Natl Acad Sci. 2010;107: 18634–18639. doi:10.1073/pnas.1009480107

33. Burton PR, Clayton DG, Cardon LR, Craddock N, Deloukas P, Duncanson A, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447: 661–678. doi:10.1038/nature05911

34. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38: 904–909. doi:10.1038/ng1847

35. Gengler N, Mayeres P, Szydlowski M. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. animal. 2007;1: 21–28. doi:10.1017/S1751731107392628

Raw input **(Step-1)**

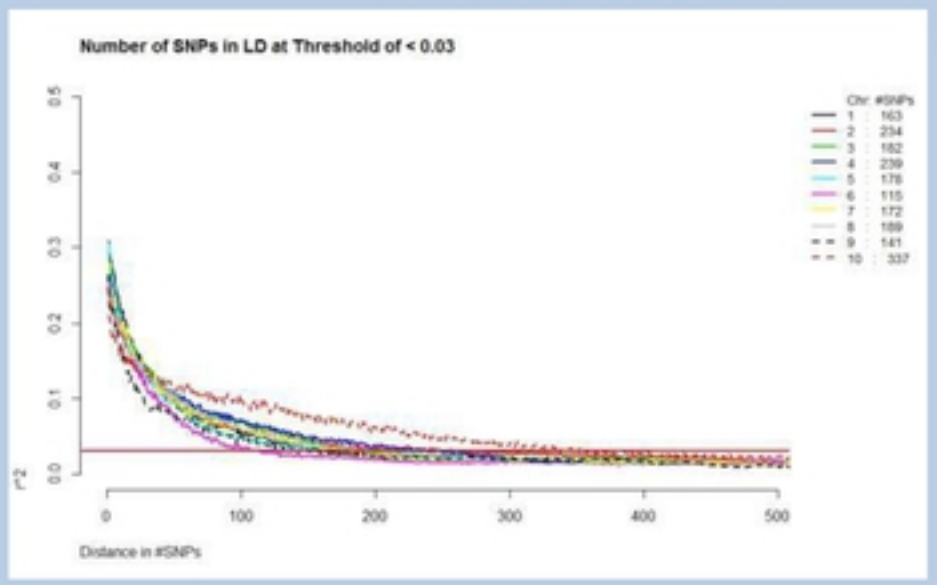
```
#Install Ghat package
install.packages("Ghat")

# Loading Ghat and the dependent packages
library(Ghat)
library(parallel)
library(rrBLUP)

# Loading phenotypes, map, genotype data
# that are available in Ghat package
phe       <- Maize_wqs[[2]]
map       <- Maize_wqs[[3]]
gen       <- Maize_wqs[[4]]

# Runing rrBLUP to estimate all markers effects,
# using full maximum-likelihood method ("ML")
result    <- mixed.solve(phe[,2],
          Z= as.matrix(gen[,2:ncol(gen)]),
          X= model.matrix(phe[,2]~phe[,3]),
          K= NULL, SE= FALSE,
          return.Hinv=FALSE, method="ML")
```

Calculating allele frequency at Cycle (generation) 1 and 3 **(Step-2)**

```
# indicating each chicle
Cy_ind <- as.numeric(unlist(strsplit(gen$X,
          split="_C")) [seq(2,2*nrow(gen),2)])
Cycle1              <- gen[which(Cy_ind == 1),]
Cycle3              <- gen[which(Cy_ind == 3),]
CycleList           <- list(Cycle1,Cycle3)


# Calculating the frequencies
Frequencies         <- matrix(nrow=ncol(gen)-
1,ncol=2)
for(i in 1:2){
  frequencies[,i]   <- colMeans(CycleList[[i]][,-
1],na.rm=TRUE)/2
}
Frequencies         <- as.data.frame(frequencies)
names(frequencies)  <- c("Cycle1","Cycle3")

# Calculating changes in allele frequencies from
cycle 1 to cycle 3
Change<-frequencies$Cycle3-frequencies$Cycle1
```

Calculating Linkage disequilibrium (LD) Decay **(Step- 3)**

```
# Using ld_decay function that implemented in
# Ghat, with a threshold of correlation between
# neighbouring of 0.03

Ld                  <- ld_decay(gen=gen,map=map,
                    max_win_snp=2000, max.chr=10,
                    cores=1, max_r2=0.03)
```



Calculating Ghat **(Step-4)**

```
# Using effective number of independent marker
# of 54.74819 (see step 3)
Ghat.adf     <- Ghat(effects=result$u,
                change=change, method = "scale",
                perms=1000,plot="Ghat",
               num_eff = 54.74819)
```

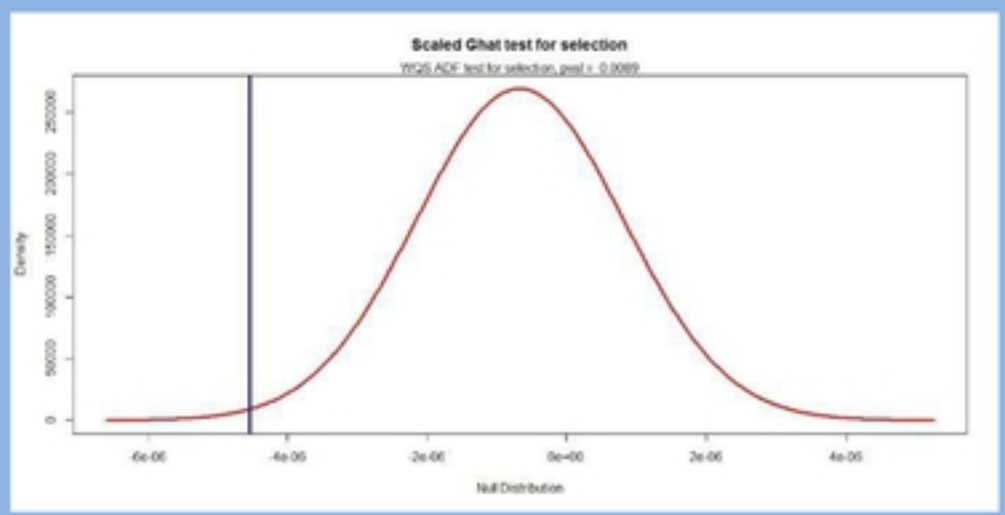


Fig1

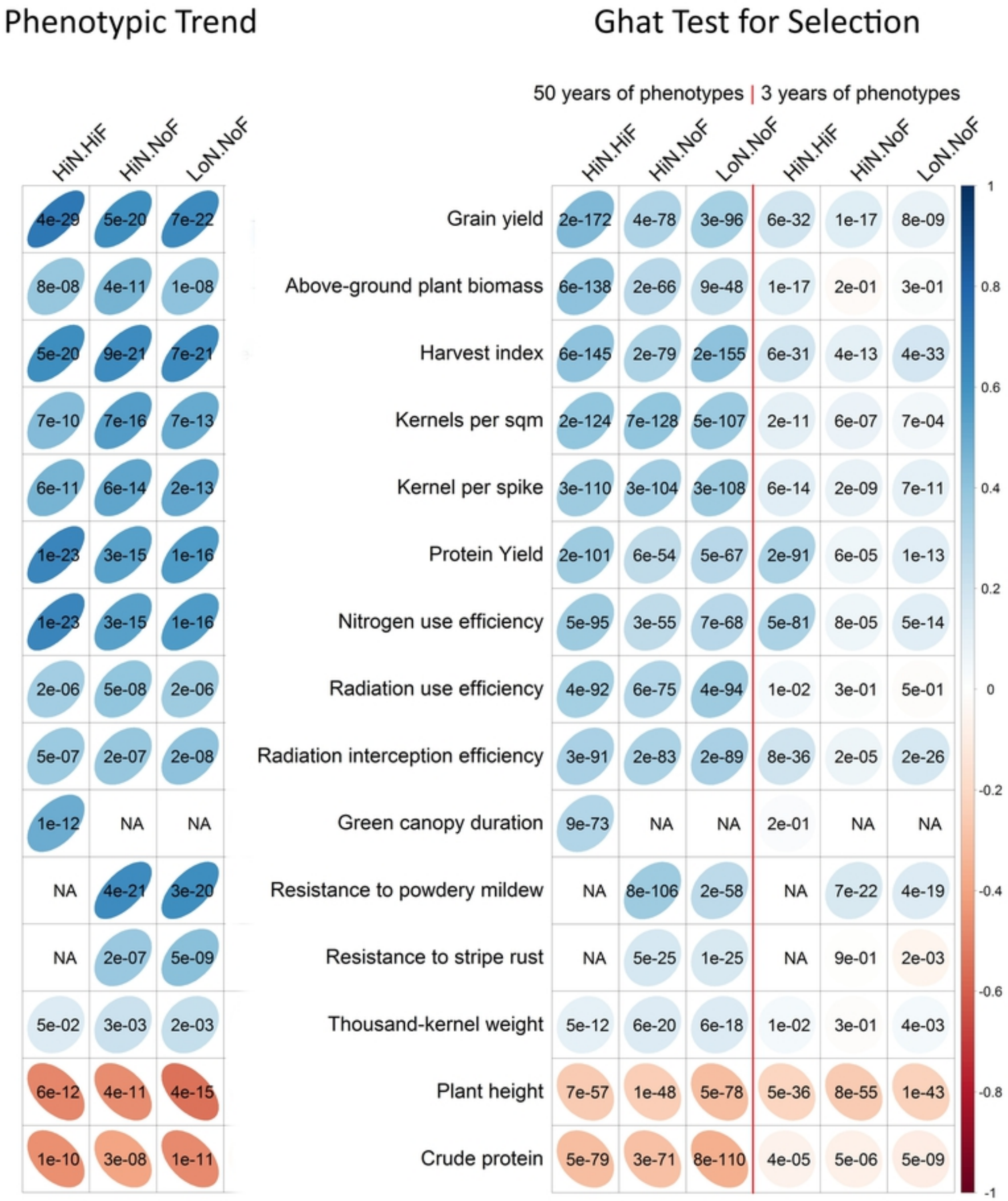

Phenotypic Trend
Ghat Test for Selection
50 years of phenotypes | 3 years of phenotypes
HiN.HiF
HiN.NoF
LoN.NoF
HiN.HiF
HiN.NoF
LoN.NoF
Grain yield
Above-ground plant biomass
Harvest index
Kernels per sqm
Kernel per spike
Protein Yield
Nitrogen use efficiency
Radiation use efficiency
Radiation interception efficiency
Green canopy duration
Resistance to powdery mildew
Resistance to stripe rust
Thousand-kernel weight
Plant height
Crude protein
4e-29 5e-20 7e-22 2e-172 4e-78 3e-96 6e-32 1e-17 8e-09
8e-08 4e-11 1e-08 6e-138 2e-66 9e-48 1e-17 2e-01 3e-01
5e-20 9e-21 7e-21 6e-145 2e-79 2e-155 6e-31 4e-13 4e-33
7e-10 7e-16 7e-13 2e-124 7e-128 5e-107 2e-11 6e-07 7e-04
6e-11 6e-14 2e-13 3e-110 3e-104 3e-108 6e-14 2e-09 7e-11
1e-23 3e-15 1e-16 2e-101 6e-54 5e-67 2e-91 6e-05 1e-13
1e-23 3e-15 1e-16 5e-95 3e-55 7e-68 5e-81 8e-05 5e-14
2e-06 5e-08 2e-06 4e-92 6e-75 4e-94 1e-02 3e-01 5e-01
5e-07 2e-07 2e-08 3e-91 2e-83 2e-89 8e-36 2e-05 2e-26
1e-12 NA NA 9e-73 NA NA 2e-01 NA NA
NA 4e-21 3e-20 NA 8e-106 2e-58 NA 7e-22 4e-19
NA 2e-07 5e-09 NA 5e-25 1e-25 NA 9e-01 2e-03
5e-02 3e-03 2e-03 5e-12 6e-20 6e-18 1e-02 3e-01 4e-03
6e-12 4e-11 4e-15 7e-57 1e-48 5e-78 5e-36 8e-55 1e-43
1e-10 3e-08 1e-11 5e-79 3e-71 8e-110 4e-05 5e-06 5e-09
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Fig2

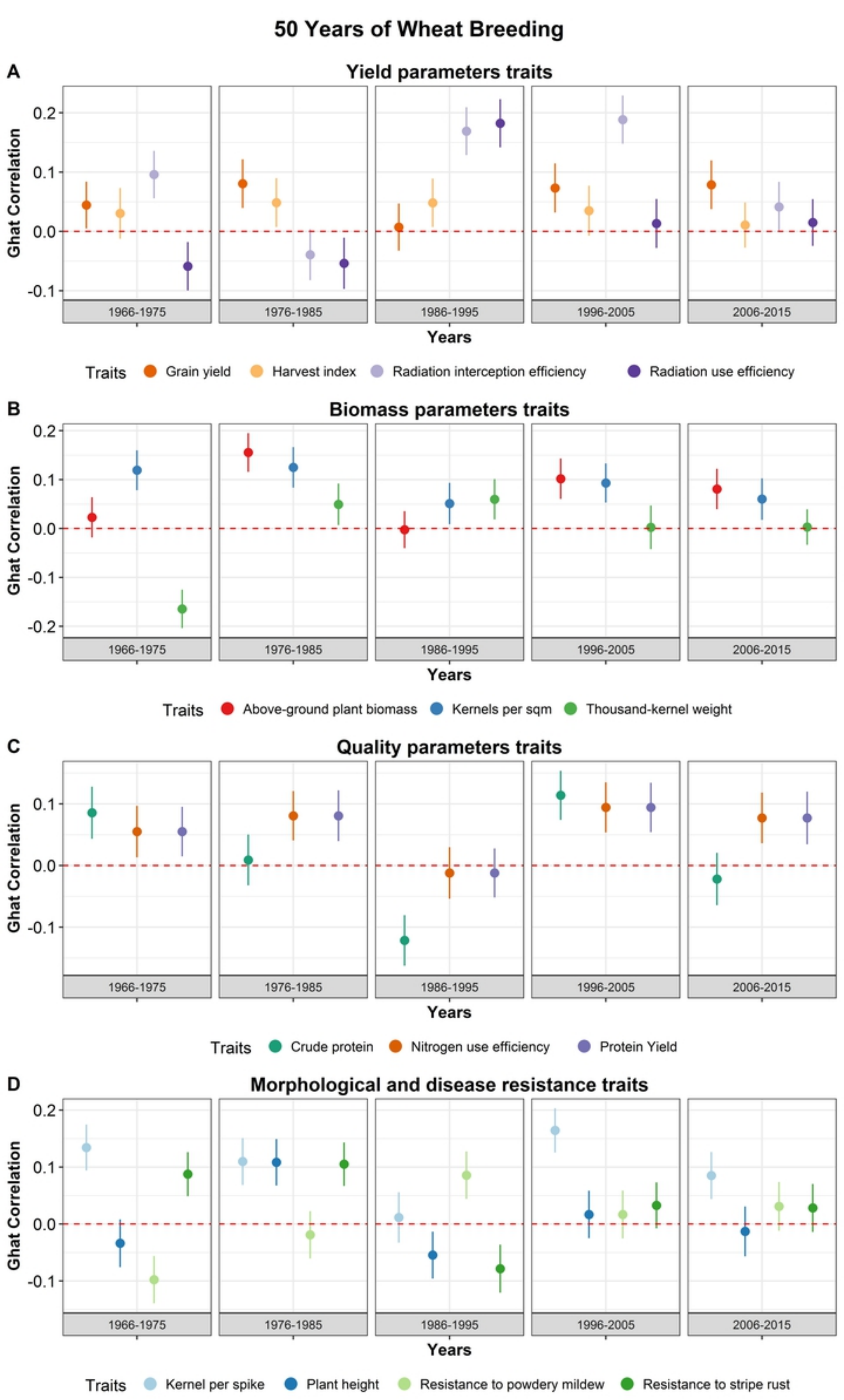

50 Years of Wheat Breeding
A
Yield parameters traits
Ghat Correlation
0.2
0.1
0.0
-0.1
1966-1975
1976-1985
1986-1995
1996-2005
2006-2015
Years
Traits
Grain yield
Harvest index
Radiation interception efficiency
Radiation use efficiency
B
Biomass parameters traits
Ghat Correlation
0.2
0.1
0.0
-0.1
-0.2
1966-1975
1976-1985
1986-1995
1996-2005
2006-2015
Years
Traits
Above-ground plant biomass
Kernels per sqm
Thousand-kernel weight
C
Quality parameters traits
Ghat Correlation
0.1
0.0
-0.1
1966-1975
1976-1985
1986-1995
1996-2005
2006-2015
Years
Traits
Crude protein
Nitrogen use efficiency
Protein Yield
D
Morphological and disease resistance traits
Ghat Correlation
0.2
0.1
0.0
-0.1
1966-1975
1976-1985
1986-1995
1996-2005
2006-2015
Years
Traits
Kernel per spike
Plant height
Resistance to powdery mildew
Resistance to stripe rust

Fig3

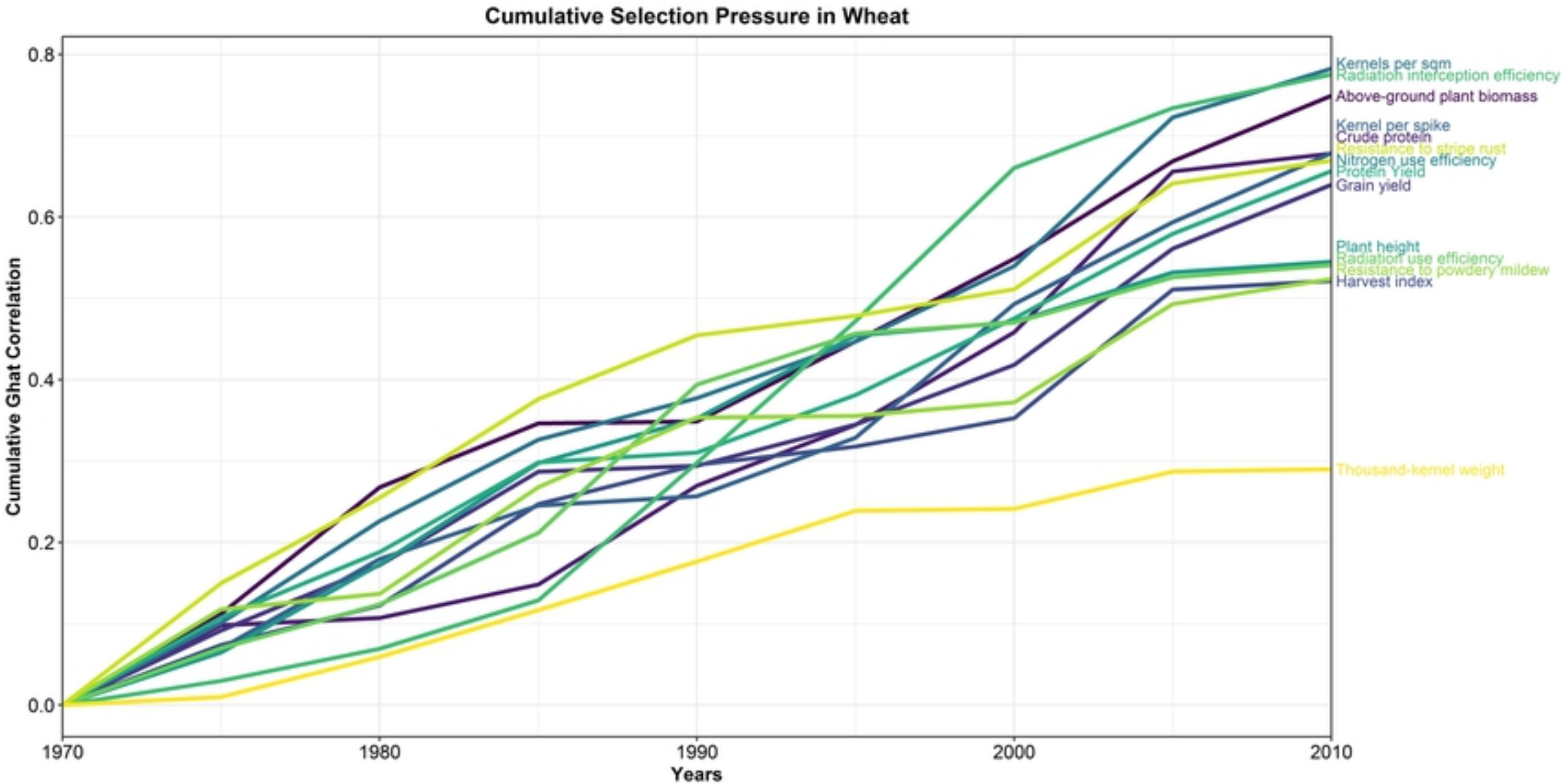

Cumulative Selection Pressure in Wheat
Cumulative Ghat Correlation
0.0
0.2
0.4
0.6
0.8
1970
1980
1990
2000
2010
Years
Kernels per sqm
Radiation interception efficiency
Above-ground plant biomass
Kernel per spike
Crude protein
Resistance to stripe rust
Nitrogen use efficiency
Protein Yield
Grain yield
Plant height
Radiation use efficiency
Resistance to powdery mildew
Harvest index
Thousand-kernel weight

Fig4