Genetic Architecture and QTL Selection Response For Kernza Perennial Grain Domestication Traits

Jared Crain
Kansas State University
Steve Larson
USDA-ARS
Kevin Dorn
USDA-ARS
Lee DeHaan
The Land Institute
Jesse Poland (jesse.poland@kaust.edu.sa)
King Abdullah University of Science and Technology https://orcid.org/0000-0002-7856-1399

Research Article

Keywords: Domestication, genome-wide association study, intermediate wheatgrass, Kernza, perennial grain, Thinopyrum intermedium

Posted Date: October 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-933049/v1

Genetic architecture and QTL selection response for Kernza perennial grain domestication traits

Jared Crain[1], Steve Larson[2], Kevin Dorn[3], Lee DeHaan[4], Jesse Poland[1,5]

[1] Department of Plant Pathology, Kansas State University, 4024 Throckmorton Plant Sciences Center, Manhattan KS, USA 66506
[2] USDA-ARS, Forage and Range Research, Utah State University, Logan, UT USA 84322
[3] USDA-ARS, Soil Management and Sugarbeet Research, Fort Collins, CO 80526, USA
[4] The Land Institute, 2440 E. Water Well Rd, Salina KS, USA 67401
[5] Center for Desert Agriculture, King Abdullah University of Science and Technology, Thuwal Saudi Arabia

Author Correspondence:

Jesse Poland
email: jesse.poland@kaust.edu.sa; jpoland@ksu.edu

Author ORCID list:
Jared Crain (0000-0001-9484-8325)
Steve Larson (0000-0003-2742-2134)
Kevin Dorn (0000-0003-2722-6221)
Lee DeHaan (0000-0002-6368-5241)
Jesse Poland (0000-0002-7856-1399)

KEY MESSAGE

Analysis revealed that the genetic architecture of an intermediate wheatgrass population was highly polygenic for both domestication and agronomic traits supporting the use of genomic selection for new crop domestication.

ABSTRACT

Perennial grains have the potential to provide food for humans as well as decrease the negative impacts of annual agriculture. Intermediate wheatgrass (IWG, *Thinopyrum intermedium*, Kernza®) is a promising perennial grain candidate that The Land Institute has been breeding since 2001. We evaluated four consecutive breeding cycles of IWG from 2016-2020 with each cycle containing approximately 1100 unique genets. Using genotyping-by-sequencing markers, quantitative trait loci (QTL) were mapped for 34 different traits using genome-wide association analysis Combining data across cycles and years, we found 93 marker-trait associations (MTA) for 16 different traits, with each association explaining 0.8-5.2% of the observed phenotypic variance. Across the four cycles, only three QTL showed an F_{ST} differentiation > 0.15 with two corresponding to a decrease in floret shattering. Additionally, one marker associated with brittle rachis was 216 bp from an ortholog of the *btr2* gene. Power analysis and quantitative genetic theory was used to estimate the effective number of QTL, which ranged from a minimum of 33 up to 558 QTL for individual traits. This study suggests that key agronomic and domestication traits are under polygenic control, and that molecular methods like genomic selection are needed to accelerate domestication and improvement of this new crop.

KEY WORDS

Domestication, genome-wide association study, intermediate wheatgrass, Kernza, perennial grain, *Thinopyrum intermedium*

ABBREVIATIONS

BLUP, best linear unbiased predictor; FDR, false discovery rate; $F_{ST,}$ fixation index; FSU, floret site utilization; GBS, genotyping-by-sequencing; GRM, genomic relationship matrix; GS, genomic selection; GWAS, genome wide association study; IWG, intermediate wheatgrass; LD, linkage disequilibrium; LOD, logarithm of the odds; MTA, marker-trait association; PVE, percent variance explained; QTL, quantitative trait loci; SNP, single nucleotide polymorphism; TLI, The Land Institute

INTRODUCTION

Perennial grain crops have the potential to revolutionize agriculture. In contrast to their annual counterparts that require regular tillage and anthropogenic disturbances (Crews et al. 2018), perennials could provide a host of ecosystem services (Glover et al. 2010; Crews et al. 2018). Documented ecosystem services by perennial crops include reduced nitrate leaching (Culman et al. 2013; Jungers et al. 2019), more complex soil communities (Culman et al. 2010), greater ability to store and retain carbon (Sprunger et al. 2018), and increased nutrient cycling (Pugliese et al. 2019). Although there are currently no large scale perennial grain crops, the development and utilization of such crops could transform both the sustainability and economic foundations of agriculture (Crews et al. 2018).

Intermediate wheatgrass (IWG, *Thinopyrum intermedium* (Host) Barkworth & D.R. Dewey, trade name Kernza) is the closest perennial relative of wheat and has a similar allohexaploid genome ($2n=6x=42$). Based on comparisons of nearly 100 species of perennial grasses, IWG was first identified for domestication in the 1980's by work at the Rodale Institute (Kutztown, Pennsylvania, USA) because of its relatively large seed size, promising yield, and palatability (Wagoner 1990a, b). In addition to more favorable agronomic traits, the grain has a soft endosperm comparable to soft wheat (*Triticum aestivum*) (Bajgain et al. 2020b), with quality evaluations showing IWG has higher levels of amino acids, protein, and bran percentage than wheat (Becker et al. 1991). Even though IWG has higher grain yield than many perennials, it is estimated to only be 10-20% of the yield of annual wheat (DeHaan et al. 2014; DeHaan and Ismail 2017), necessitating sustained breeding efforts to increase the yield of this potential grain crop. Additionally, several other agronomic and domestication traits such as reduced shattering, increased seed size, and improved threshability are needed to make IWG a mainstream crop.

Uninterrupted breeding efforts to improve IWG have been conducted at The Land Institute (TLI), Salina, Kansas, USA, since 2001 (DeHaan et al. 2018), with new breeding programs being initiated in Minnesota, USA (2011), Manitoba, Canada (2011), Utah, USA (2019), and Uppsala, Sweden (2019) (Cattani 2016; Zhang et al. 2016; Bajgain et al. 2020b). While the initial cycles of selection relied on recurrent phenotypic selection (Zhang et al. 2016; DeHaan et al. 2018), advances in low cost, high-throughput DNA sequencing has permitted IWG breeding to harness the power of genomic selection (GS) (Zhang et al. 2016; Bajgain et al. 2020a; Crain et al. 2020a,

2021b, a). Within TLI's breeding program, GS has reduced the breeding cycle from three years to one year per cycle (DeHaan et al. 2018) and simultaneously maintained an estimated 8% year^{-1} increase in spike yield (Crain et al. 2021a). Furthermore, decreased sequencing cost has resulted in a wealth of genomic information for crop improvement including genetic maps (Kantarski et al. 2016), a draft genome sequence (https://phytozome-next.jgi.doe.gov/info/Tintermedium_v2_1). These genomics resources have enabled genome-wide association studies (GWAS) for agronomic traits including seed size (Zhang et al. 2017; Larson et al. 2019), flowering time (Altendorf et al. 2021b), and grain yield components (Bajgain et al. 2019; Larson et al. 2019) that can be used to better understand and guide IWG breeding.

Since initiating GS in 2017, TLI has completed four cycles of selection for reducing seed shattering, enhancing threshability to produce naked seed (free-threshing trait), increasing seed mass, and developing higher yield per spike. Even though selections have been primarily based on GS models for these few primary traits, up to 34 traits have been measured which allow for a holistic assessment of the breeding program. The estimated genetic gains have generally been favorable and at a more rapid rate than phenotypic selection alone, yet there has been some evidence of unanticipated results. Within the breeding program, increasing spike yield has been associated with increased seeds per spike, number of florets per spike, and florets per spikelet, yet the floret site utilization (FSU, referred to as percent seed set in Crain *et al.*, 2021a) decreased, suggesting less efficient use of resources. While FSU has not been a direct target of selection in the TLI program, Altendorf et al. (2021a) have found that FSU was the primary driver of yield for spaced plants grown on 1m centers (e.g. one-meter spacing between plants).

Within annual wheat, increasing the number of seeds per spikelet (Würschum et al. 2018) or spike fertility, percent of grain weight to total spike weight, has been shown to increase yield (Alonso et al. 2018), yet Philipp et al. (2018) reported that there appears to be little evidence that the number of spikelets per spike has been improved in elite varieties from landraces or wild germplasm. Wheat and barley display indeterminant numbers of fertile florets per spikelet or spikelets per spike, respectively, but are otherwise determinate for these traits, illustrating different ways of increasing the number of seed per spike (Zhong et al. 2021). Although IWG is indeterminate for the number of fertile florets per spikelet and spikelets per spike, a key element that should be considered is the difference between annual and perennial life cycle, specifically if

a high yielding perennial grain crop is viable. Research has shown that perennials devote more resources below ground than do their annual counterparts, and that this allocation is a precursor to switching between perennial and annual life cycles (Lindberg et al. 2020). Additionally, selecting for higher seed yield may induce concessions from below ground resources and plant longevity (Vico et al. 2016).

While there are some arguments against perennial grains due to the ecological and physiological limitation of perennial plants (Smaje 2015), current work suggest that favorable gains can be made through artificial selection (DeHaan et al. 2014; Zhang et al. 2016; Crain et al. 2021a). As breeding programs mature, they should assess whether the realized gains in perennial crops are matching the target gains for both agronomic yield and increased ecosystem services. Given the rapid cycling nature of the TLI IWG breeding program and the results from the first few cycles of GS (Crain et al. 2021a), our objectives are to 1) conduct a GWAS for observed traits to identify associated loci for key agronomic traits, 2) determine the genetic architecture of the observed traits, 3) assess allele frequency changes across the four cycles of selection for significant marker-trait associations, and 4) evaluate the potential selection opportunities to drive genetic gains for desirable physiological and agronomic outcomes such as high grain yield and high FSU.

MATERIALS AND METHODS

Plant Material

All plant material used in this study came from the TLI breeding program, Cycles 6 to 9, with TLI-Cycle 6 being extensively described in DeHaan *et al*. (2018) and TLI-Cycles 7 to 9 detailed in Crain *et al*. (2021a, b). Briefly, TLI-cycle 6 formed the initial training population for GS and consisted of 3,658 space planted genets that were evaluated in 2016 and 2017 at Salina, KS (Crain et al. 2021b). As outcrossed IWG plants are all unique and heterozygous (excluding clones or ramets), the term "genet" herein refers to a genetically unique individual which is typically a single plant but possibly cloned ramets, while genotype herein refers to the DNA sequence of a particular genet (Zhang et al. 2016). Phenotypic data alone was used to select TLI-Cycle 6 genets that were randomly intermated to form TLI-Cycle 7. Genomic selection was used to identify 118 TLI-Cycle 7 genets, out of 4,183 genotyped, to intermate to form TLI-Cycle 8 seed. Another 1,216 TLI-Cycle 7 genets were selected for field evaluations to train future GS models and divided randomly between an irrigated and non-irrigated site. Genets were space planted on 0.91m centers in the fall of 2017 with phenotypic evaluations in 2018, 2019, and 2020. TLI-Cycle 8 and 9 were formed in a same manner with around 100 selected genets intermated to form each subsequent cycle out of nearly 3,500 genotyped genets. Planting was similar to TLI-Cycle 7, where individual genets were divided between irrigated and non-irrigated sites and planted on 0.91m centers. TLI-Cycle 8 training population consisted of 1,092 genets planted in the field in the fall of 2018 and evaluated during 2019 and 2020. The TLI-Cycle 9 training population was comprised of 1,004 genets, planted in the fall of 2019 with first year phenotypic observations in 2020. Across all cycles, there was no replication of genets, thus each genet was evaluated as a unique single plant.

Phenotypic Assessment

Each year phenotypic traits were measured to evaluate genet performance, with a total 34 unique traits (Crain et al. 2021a). Within the breeding program the most important traits which are key selection targets include shattering, percent free-threshing seed, seed mass, and spike yield. Shattering was rated on a scale of 0 to 5, where 0 indicated no shattering and 5 indicated more than 12 florets shattering per evaluated spike (DeHaan *et al.*, 2018). From 2016 to 2018, shattering was considered a single trait; however, work by Altendorf (2020) indicated that floret

and brittle rachis shattering should be scored separately, so beginning in 2019 brittle rachis was scored as a separate trait in the IWG population. In addition, many other secondary traits including seeds spike^{-1}, spikelets spike^{-1}, florets spike^{-1}, and FSU were evaluated. While most traits were assessed consistently across years and cycles, it should be noted that TLI-Cycle 6 had significant missing data due to flooding, and reduced data collection in 2020 reflected limited labor due to the COVID-19 pandemic. A subset of 1,470 TLI-Cycle 6 genets were selected to make approximately equal representation of genet number between cycles and follows previous work by Crain *et al*. (2021a).

A linear mixed model, Eqn. 1, was used to calculate trait best linear unbiased predictors (BLUPs) for each genet using ASREML version 4.1 (Gilmour et al. 2015).

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e} \qquad \text{Equation 1}$$

In Eqn. 1, $\boldsymbol{y}$ is a vector of phenotypic observations, fixed and random effects are given by vector $\boldsymbol{b}$ and $\boldsymbol{u}$ respectively, and $\boldsymbol{e}$ is a vector of residuals. The incidence matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$ allocate each fixed or random effect to their corresponding observation in **y** (Isik et al. 2017). For each model no fixed effects were added, so $\boldsymbol{Xb}$ reduces to the mean vector. Random effects that were normal, independent and identically distributed ~NIID(0, σ^2_{effect}) were included for site-year combination, multiple measurements for each genet representing observations across years, and a nugget effect for residual error variance. A random term for genet was included that had a mean 0 and a known variance-covariance matrix of the genomic relationship matrix (GRM) ~(0, σ^2_{genet}**GRM)**, which explained the genet effect that accounts for the relationship between genets using the GRM (Isik *et al.*, 2017 pg 124-125). The GRM was calculated as θ**MM'** where M is an matrix of marker scores with dimensions *n* individuals by *m* markers, and θ is a proportionality constant (Endelman and Jannink 2012). The GRM was computed with the *A.mat* function in the *rrBLUP* R package (Endelman 2011). Within the model, residual error was formed of two parts with the nugget being NIID and then a correlated error term for rows and columns (AR1 x AR1, autoregressive first order correlation structure) (Isik *et al.*, 2017 pg 93; 217). A separate AR1 x AR1 structure was fit for each cycle-site-year combination (14 total combinations), with ASREML requiring a complete row column matrix, with any incomplete observations filled in with dummy variables. This model fit one BLUP per genet regardless of if a trait had been measured one or multiple times and will be referred to as combined analysis as all years and

cycles of data observations were combined in one model. For some traits, convergence failed using the AR1xAR1 model, and a reduced model with no row and column error structure was fit. Eqn. 1 was also fit individually for each cycle-year combination by dropping terms for cycle-year combination and repeated measurements across years.

Genomic Profiling

All genets were profiled using genotyping-by-sequencing (GBS) using a two enzyme protocol as in Poland et al. (2012). DNA extraction and pooled 192-plex GBS libraries were prepared at Kansas State University with all sequencing conducted at Hudson Alpha, Huntsville, AL using Illumina HiSeq machines. Single nucleotide polymorphisms (SNPs) were scored using the TASSEL GBSv2 pipeline (Glaubitz et al. 2014) and the *Thinopyrum intermedium* draft genome reference sequence (prerelease access provided by *Thinopyrum intermedium* Genome Sequencing Consortium). The IWG draft genome reference includes three sets of seven chromosomes numbered 1-7 based on homology to the seven chromosomes of barley (Kantarski et al. 2016). Chromosomes corresponding to three homoeologous groups (subgenomes) of IWG were designated J1-J7, S1-S7, and V1-V7 based on homologies to possible diploid ancestors in the prereleased *Thinopyrum intermedium* draft genome reference sequence (unpublished data). A total of 123,423 putative SNPs were identified across the 6,824 genotyped genets. Filtering was done based on four criteria. First, each SNP aligned to one unique location on one of the 21 main chromosomes. Second, a minimum read depth of 4 tags was required to call a homozygous genotype, while heterozygotes could be called with a minimum two contrasting tags for each SNP. If the minimum read depth threshold was not met, the SNP site was set to missing. Third, the maximum data missing per SNP was 70% and individual genets could not have more than 95% missing data. Fourth, SNPs must have a minor allele frequency (MAF) greater than 0.01. After filtering, this dataset consisted of a total of 6,517 genets and 23,611 SNPs. Markers were imputed with Beagle version 4.1 using the default parameters (Browning and Browning 2016).

Linkage Disequilibrium and Genetic Parameters

Linkage disequilibrium (LD) was evaluated using TASSEL version 5.2.3 (Bradbury et al. 2007) for all pairwise comparisons within each chromosome. The Hill and Weir formula (Hill and Weir 1988) was fit using the *nls* function in R (R Core Team 2020) to describe the extent of genome and chromosome LD using r^2. The greater of the distance at which half of the maximum value of

the fitted value occurred, or r^2 =0.1 was considered the extent of LD (Flint-Garcia et al. 2003). The fixation index F_{ST} (Weir and Cockerham 1984) was used to evaluate population differentiation among cycles and was calculated using the *diveRsity* R package (Keenan et al. 2013). Values of F_{ST} > 0.15 were considered evidence of population differentiation whereas F_{ST} < 0.05 was considered as no evidence of population divergence (Hartl & Clark, 1997 pg 118-19).

Genome-wide Association Analysis and QTL Identification

The *GWAS* function in *rrBLUP* (Endelman 2011) was used to assess marker-trait associations for each set of phenotypic trait data both jointly and by cycle-year combination. The GWAS model is a mixed-linear model (Yu et al. 2006) with the form:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zg} + \boldsymbol{S\tau} + \boldsymbol{e} \qquad \text{Equation 2}$$

where $\boldsymbol{y}$ is an *n x 1* vector of phenotypic observations (BLUPs from Eqn. 1), $\boldsymbol{\beta}$ is a *p x 1* vector of fixed effects where *p* is the number of fixed effects for population structure, **X** is an *n x p* design matrix for fixed effects, $\boldsymbol{g}$ is an *n x 1* vector of random polygenic effects, $\boldsymbol{Z}$ is an *n x n* matrix that is the **GRM**, $\boldsymbol{\tau}$ is the fixed effect for a given marker being tested and $\boldsymbol{S}$ an *n x 1* vector of marker scores for the respective locus, $\boldsymbol{e}$ is an *n x 1* vector of random residuals. Population structure was accounted for by using the first six principal components (*p* = 6), and model compression used 'population parameters previously determined' (P3D) (Zhang et al. 2010)

A total of 23,611 markers were tested for each trait, and markers with a significance threshold above a 0.05 false discovery rate (FDR) (Storey and Tibshirani 2003) were considered significant. The FDR was calculated using a modified function from in *rrBLUP* R package (Endelman 2011). Plots were created using the qqman R package (Turner 2017). For each significant marker, marker effects were determined using the *lmekin* function from the *coxme* R package (Therneau 2020) following the analysis of Sehgal et al. (2020). Percent variance explained (PVE) was calculated following methods by Broman and Sen (2009 pg. 246). As there were often more than one significant marker on the same chromosome, we used a minimum gap threshold of 100MB between significant markers to distinguish and count unique QTL. Each unique QTL was identified by the marker with the highest logarithm of the odds (LOD) value

along with other significant markers not separated by at least 100 Mb, herein referred to as associated markers.

Power analyses were completed using scripts from Wang and Xu (2019) where β was set to 0.8. Minimum detectable QTL effect sizes were determined based on sample size, relationship between individuals, and heritability, where heritability was estimated from variance components of Eqn 1 as:

$$\frac{\sigma_g^2}{\sigma_p^2} \qquad \text{Equation 3}$$

where σ_g^2 is genet variance, σ_p^2 is phenotypic variance which is the sum of genet variance, variance due to multiple observations, and residual error variance. The total number of QTL per trait were estimated using a squared exponential distribution from Hall et al. (2016) according to the formula:

$$\boldsymbol{n_{QTL}} = \frac{\boldsymbol{h^2}}{\boldsymbol{\mu_d - \sqrt{2\mu_d\theta - \theta^2}}} \qquad \text{Equation 4}$$

where h^2 is the heritability calculated from Eqn 2, μ_d is the average percent variance explained by detected QTL, and θ is the lowest detectable QTL estimated from the power analysis. This formula worked for all traits with detected QTL, but a minimum number of QTL for each trait were estimated by dividing 1 by the smallest detectable QTL size. This provided a lower bound on the number of QTL, regardless of if a QTL for any trait had been observed. Combining the power analysis, which provided a minimum detectable PVE, the number of estimated QTL (Eqn. 4), and heritability we could estimate the size of the population required to detect QTLs explaining a given level of the total genetic variance (Lynch and Walsh 1998; Hall et al. 2016). For all analysis, we estimated the population size needed to detect QTLs accounting for 50% of the genetic variation.

Data Availability

All DNA sequence data has been deposited in the NCBI Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/bioproject/) as part of the umbrella BioProject PRJNA609325. All scripts for data analysis and phenotypic data have been placed in the Dryad Digital

Repository (https://datadryad.org/stash/share/A6qr_cVxvh7suuju4V0CwXNMAD_Wvg7gQxaKChFJY6E).

RESULTS

We assessed four TLI IWG breeding cycles that comprised approximately 4,200 genets and five years of phenotypic data to dissect quantitative traits and inform breeding decisions. A linear mixed model was used to account for multiple years of phenotypic observations and develop BLUPs, leveraging data collected within the breeding program to better understand IWG improvement. While a total of 34 different traits were observed across years (Supporting Information Table S1) the primary traits of selection were shattering, free-threshing seed, seed mass, and seed yield. Across cycles seed mass and shattering were positively correlated with spike yield, while a negative association was generally observed between free threshing and spike yield (Supporting Information Figures 1-4).

Linkage Disequilibrium Analysis

We evaluated the extent of linkage disequilibrium (LD) up to distance of 200 Mb, which is substantially less than the size of most IWG chromosomes. Within the population LD declined relatively rapidly, with genome-wide LD extending an average of 305 kb (Fig. 1). For individual chromosomes the half decay distance for r^2 ranged from less than 1kb up to 1.43 Mb with chromosome S3 having the shortest LD and chromosome J3 the longest LD. Even though average LD declined rapidly, there were numerous marker combinations that maintained LD at larger distances up to 50 Mb (Fig. 1).

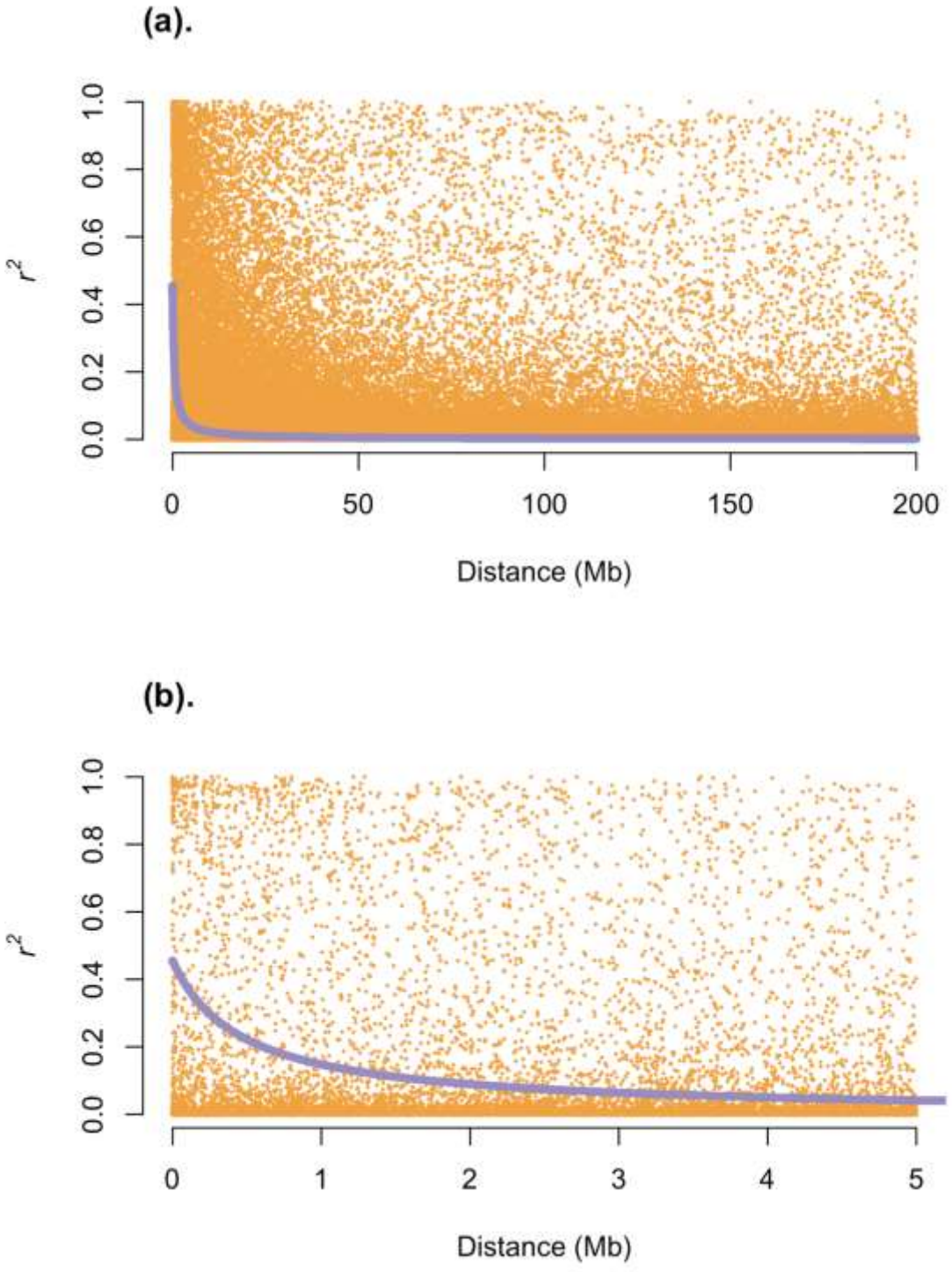

Fig. 1 genome-wide linkage disequilibrium (LD) for intermediate wheatgrass (*Thinopyrum intermedium*) for 200 Mb region (**a**) and 5 Mb region (**b**). Orange points represent individual marker combinations with a 250-marker sliding window. Average LD has been computed with the Hill and Weir formula (1988) and shown in blue

Genome-wide Association Analysis

We used a genome-wide association analysis to identify the location, number, and size of QTL underlying traits of interest to the IWG breeding program. Across all traits, the combined analysis found 93 marker-trait associations for 16 different traits, representing 37 separate QTL (Supporting Information Table 2). Of the traits of most interest to the breeding program—spike yield, free threshing, seed mass, and shattering—only QTL were identified for shattering (floret and brittle rachis) and free threshing. Both brittle rachis and free threshing had more than one QTL on the same chromosome (J3 and V2 respectively, Table 1 and Figure 2). These effects were small, explaining 1.0-2.7% of the observed variation with allele effects ranging from -0.13 to 0.11 units less shattering on a 5 point scale and up to 4.6 percentage points less free threshing on a 100 point scale for the alternate alleles compared to the reference genome (Table 1 and Supporting Information Table S2)

We found evidence for brittle rachis QTL which impacts shattering on chromosome J3 (Fig. 2), where the most significant marker (J3_122986862) was 5.6 Mb from IWG *brittle rachis 2* (*Btr2*) gene (Pourkheirandish et al. 2015) while another significant marker (J3_115931563, LOD = 8.47) was only 217 bp away from a *Btr2* gene. This QTL region was identified both in the combined analysis and analysis across individual cycles and years (Supporting Information Table 2) and was supported by up to 17 associated markers above the genome-wide threshold (Fig. 2).

Seed circularity had the most significant markers of any trait in the combined analysis, with 23 loci representing eight unique QTL located across seven different chromosomes (Fig. 2, Supporting Information Table S2). For the number of florets per spike and florets per spikelet, one QTL region overlapped with colocalized markers on chromosomes J5 having the same directional effects (Supporting Information Table S2). One QTL for FSU was identified on chromosome S5.

Along with analyzing the combined data, each cycle-year combination was analyzed independently. This resulted in 209 significant markers representing 67 unique cycle-year QTL being observed across 20 different traits (Supporting Information Table S2). Many of these QTL signals were similar to the combined analysis that the QTL had several, up to 26, associated

markers per QTL. Taken together, all analysis revealed QTL associations across 19 of the 21 chromosomes of IWG, with many chromosomes harboring QTL for multiple traits (Supporting Information Table S3).

Across 34 traits and up to nine cycle-year combinations, all of the identified loci using the joint analysis explained minimal variation, with 5.2% percent variation explained (PVE, stem diameter) being the maximum for any combined analysis with an average of 1.7% PVE per identified QTL. When considering markers identified by cycle-year analysis, the PVE were greater than the combined analysis, yet only 14 of the 209 markers had PVE > 10%.

Table 1 Significant quantitative trait loci (QTL) associations in The Land Institute intermediate wheatgrass (*Thinopyrum intermedium*) breeding program for priority traits

Trait	QTL #	Associated markers	Chr	Position	n	LOD	F_{ST}	PVE	Ref/ alt	MAF	Ref value	Ref SE	Alt value	Alt SE	Het value	Het SE
Brittle rachis	1	17	J3	122986862	1441	8.51	0.01	2.7	A/G	0.38 (A)	0.93	0.008	0.04	0.009	0.02	0.007
	2	0	J3	421522694	1050	4.71	0.08	2.0	G/T	0.37 (T)	0.98	0.008	-0.05	0.014	-0.02	0.008
	1	1	S3	132131062	1653	5.32	0.06	1.5	C/G	0.34 (G)	0.95	0.005	0.05	0.009	0.03	0.005
	1	0	S5	391487396	1659	4.58	0.00	1.3	T/C	0.34 (T)	0.98	0.008	-0.03	0.005	-0.01	0.005
Free threshing	1	0	V2	196007664	2867	5.95	0.08	1.0	G/T	0.20 (T)	55.53	0.158	-3.34	0.649	-1.50	0.351
	2	0	V2	357856894	1478	6.11	0.12	1.9	T/G	0.18 (G)	55.93	1.103	-4.65	0.963	-1.79	0.846
Shattering	1	0	J2	636725449	1359	4.58	0.00	1.5	C/T	0.17 (T)	2.19	0.010	-0.13	0.039	-0.08	0.021
	1	6	S2	441397840	3130	10.60	0.01	1.5	T/C	0.47 (T)	2.13	0.015	0.10	0.008	0.04	0.009
	1	1	S4	341952545	2204	5.90	0.22	1.2	G/C	0.46 (C)	2.20	0.023	-0.08	0.015	-0.02	0.013
	1	2	S5	380535059	1620	4.96	0.00	1.4	T/A	0.25 (A)	2.16	0.030	0.11	0.027	0.08	0.023

Marker with highest LOD reported for each QTL; QTL #, number of QTL per chromosome; Associated markers, number of significant markers associated with each QTL; Chr, chromosome; n, number of individuals observed; F_{ST} fixation index between The Land Institute (TLI) Cycle 6 and TLI Cycle 9; PVE, percent variance explained; Ref/alt reference and alternate allele respectively; MAF, minor allele frequency for combined TLI Cycles 6-9 with minor allele in parenthesis; Ref value, reference allele value for a trait, Ref SE, standard error of reference value; Het, heterozygous

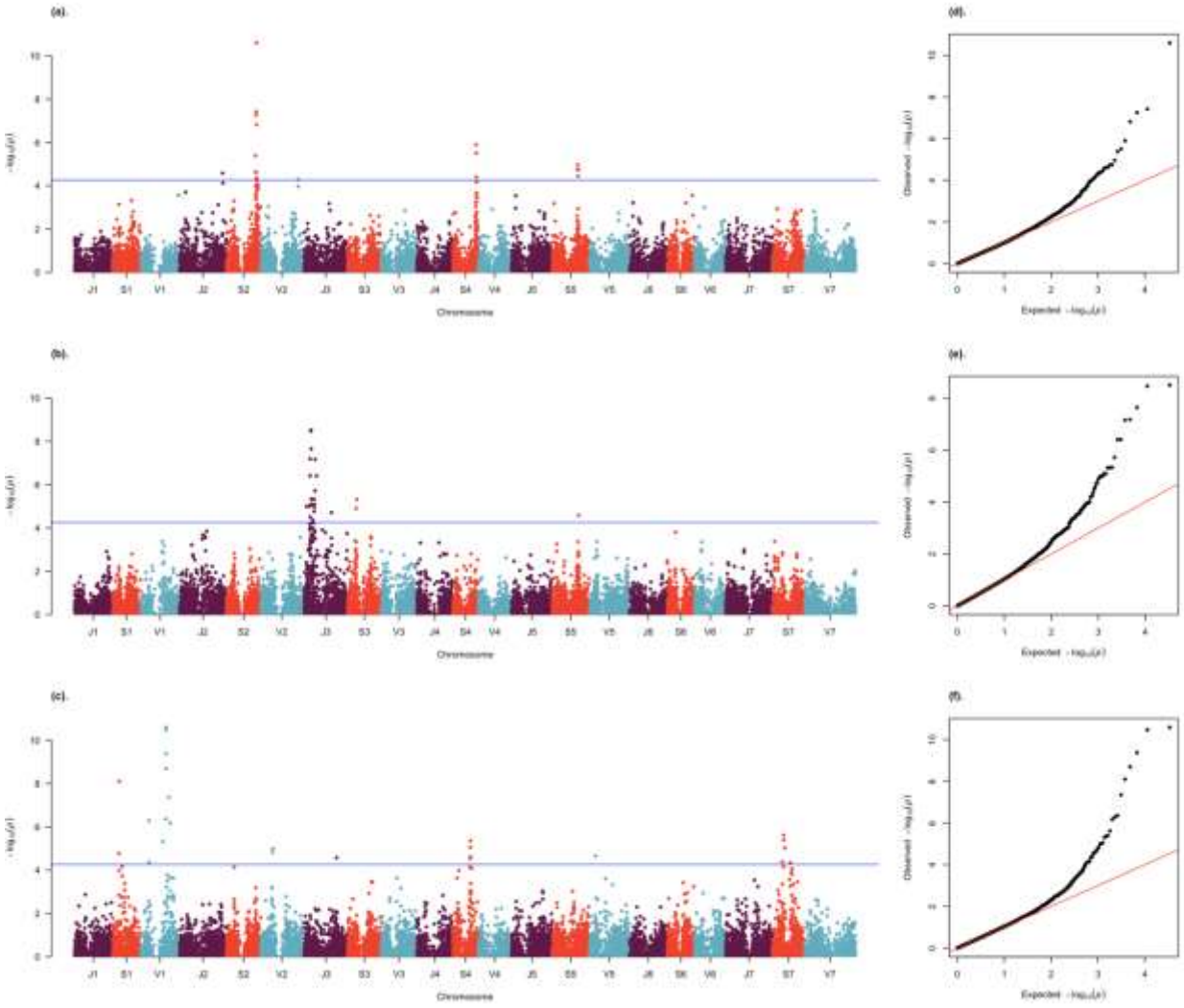

Fig. 2 Manhattan plots of (**a**) shattering, (**b**) brittle rachis, and (**c**) seed circularity in intermediate wheatgrass (*Thinopyrum intermedium*) with line indicating 0.05 false discovery rate. Panels **d**, **e** and **f** show quantile-quantile (QQ) plots for p values under the null hypothesis (no association) and observed p values for brittle rachis (d), shattering (e), and seed circularity (f), respectively

Number of Effective QTL

To evaluate the genetic architecture of these domestication and agronomic traits, we estimated the number of effective QTL for each trait using results from our power analysis, heritability, and QTL analysis. In general, our analysis of this breeding germplasm had the ability to detect small QTL, with PVE of the smallest detectable QTL ranging from 0.7 to 3.0% for each trait (Table 2). Determining the smallest detectable QTL also provided a lower bound estimate of the minimum number of QTL for each trait which ranged from 33 to 149 (Table 2) regardless of whether we had detected QTL. For traits with detected QTLs, we estimated the number of QTL for a given trait (using Eqn. 4) which ranged from 93 to 357 (Table 2) for combined analysis. Using each cycle-year combination, a range of QTL could be estimated for traits with detected QTL. For important traits such as shattering, the estimated number of QTL ranged from 97 to 258, brittle rachis could be controlled by up to 293 QTL, and free threshing could have as few as 39 QTL. While the reported number of QTL could vary greatly within and between traits, these estimates demonstrate that these traits are highly polygenic and controlled by many loci.

We also estimated the population size, that would be required to detect QTLs explaining 50% of the genetic variation. Population size differed between traits, ranging from a minimum of 98 up to 15,931 plants with an average population size of 1720 (Table 2). For priority breeding traits of spike yield, reduced shattering, and seed mass, the minimum population sizes, to detect QTLs explaining 50% of the genetic variance, were all > 2500 plants.

Table 2 Number of estimated quantitative trait loci (QTL) for phenotypic traits in intermediate wheatgrass, and the estimated population size to needed to detect QTLs explaining 50% of the genotypic variance

Trait	Smallest detectable QTL in percent variance explained	Minimum number of QTL	QTL estimate based on Hall et al. (2016)	Minimum QTL for individual cycle year combinations	Maximum number of QTL for individual cycle year combinations	Number of cycle-year combinations with observed QTL	h^2	Minimum population size	Maximum population size
Brittle rachis	0.9	111	166	18	293	4	0.44	624	7339
Flag leaf height (cm)	3.0	34					0.44	1042	
Flag leaf length (cm)	2.8	36		5		1	0.31	212	1522
Flag leaf width (mm)	2.8	36		84		1	0.62	755	1633
Free threshing	0.7	147	304	39		1	0.63	809	5343
Lodging	1.3	76					0.38	2433	
maturity	0.7	147		6	221	2	0.5	281	5055
Number of florets per spike	0.9	110	214				0.23	5592	10380
Number of florets per spikelet	0.9	110	177				0.25	5057	7919
Peduncle width (mm)	2.4	42	289	303		1	0.52	1004	6472
Floret site utilization	0.9	109	236	101		1	0.75	1572	3578
Plant height (cm)	0.7	149		14	558	3	0.38	627	15931
Seed area (mm2)	0.7	147		4	22	2	0.69	93	2428
Seed density	0.7	147	133				0.45	3398	3737
Seed image circularity	0.7	144	239	16	88	4	0.52	475	4948
Seed length (mm)	0.7	147	251	113		1	0.58	2298	4864
Seed mass (mg)	0.7	147					0.63	2732	
Seed perimeter (mm)	2.7	37					0.79	613	
Seed width (mm)	0.7	147	126	14	115	3	0.36	628	4550
Seeds per spike	0.7	148					0.42	3972	
Shattering	0.7	148	187	97	258	6	0.4	2823	7139
Spike dry weight (g)	0.9	110					0.36	3625	
Spike emergence (cm)	2.3	43	283	266		1	0.51	1064	6102
Spike emergence %	1.4	74	357				0.51	1770	7645
Spike harvest index	0.9	110					0.32	3961	
Spike length (cm)	1.3	78	268	547		1	0.49	1890	12297
Spike yield (g)	0.7	148					0.53	3231	
Spikelet density	2.7	37		21		1	0.63	482	777
Spikelets per spike	0.9	110	101	14		1	0.25	948	5153
Stem angle	1.3	79					0.59	1586	
Stem diameter (mm)	2.9	35	93	50		1	0.51	897	2197
Stem strength bottom	3.0	33		30		1	0.45	934	979
Stem strength middle	1.3	75		8		1	0.3	515	3049
Stem strength top	3.0	33					0.77	552	

Maximum population size only calculated if more than one number of QTL estimate available.

Allele Frequency and F_{ST}

Using all markers, fixation index, F_{ST}, was calculated to examine the potential impact of directional selection which would generation genetic differentiation between TLI-Cycle 6 and 9. Only 548 (2.3%) of these markers showed high or very high genetic differentiation, suggesting that many areas of the genome have not been under consistent selection pressure (Table 3). Of the 72 markers that showed very high genetic differentiation, they were distributed across 20 of the 21 chromosomes. For significant loci identified by GWAS, allele frequency and F_{ST} were used to further evaluate changes between TLI breeding cycles 6 and 9. As direct selection should alter allele frequency, we expected this analysis to provide evidence of selection pressure. Of the significant MTAs identified, only 10 markers (4.5%) had F_{ST} values > 0.15 (Table 4).

Of 24 traits with MTA, only three traits—seed circularity, shattering, and seed length—showed F_{ST} > 0.15 for any single trait associated marker. For shattering, a trait that has been under strong selection, two markers on chromosome S4 had high F_{ST} values (Fig. 3) and the other associated markers on chromosome S4 had moderate F_{ST} values. For this trait, all other significant markers on chromosomes J2, S1, S2, S3, S5, and V2 did not show any significant differentiation (F_{ST} < 0.05) after 3 cycles of selection based on F_{ST} values. The high F_{ST} value of marker S04_341952545 resulted in the alternate allele frequency increasing from 26% to 60%, corresponding to a -0.08 unit decrease in shattering over three cycles of selection (Fig. 3, Table 4). Compared to the 2.3% of genome-wide markers that showed high differentiation, up to 6.4% of the shattering markers had high F_{ST} values.

The marker J03_115931563, which had a strong association with brittle rachis had a moderate F_{ST} value of 0.11. Surprisingly, from TLI-cycle 6 to TLI-cycle 9 the frequency of the favorable (reference) allele actually decreased from 28% to 7.5%, in the opposite direction as expected by selection for reduced shattering. Evaluating this locus for all other traits showed that the reference allele while favorable for shattering and free threshing was detrimental to spike yield, seeds spike^{-1}, and spike dry weight. This suggest that there is tradeoff between spike yield and brittle rachis at this locus and that selection for increased yield could be driving the alternate allele.

In addition to shattering, high F_{ST} values were observed for seed circularity and seed length at the locus V01_438389996. This marker showed the reference allele frequency changed from 0.69 to 0.96 over the four cycles, corresponding to a decrease in seed circularity and increase in seed length, i.e. a more elongated seed. Even though significant markers were identified for number of florets per spike, florets per spikelet, spikelets per spike, and FSU—yield component traits—all markers except for one showed little differentiation across the four cycles of selection. Even for the priority trait of free threshing, only two of 10 markers (20%) showed moderate differentiation. The F_{ST} values for markers associated with plant height, which has not been targeted by selection, did not show any differentiation.

Table 3 Fixation index, F_{ST}, values for all single nucleotide polymorphisms (SNPs) in The Land Institute intermediate wheatgrass breeding population cycles 6 and 9

Level of differentiation	Range of F_{ST} values	Number of loci (%)			
		Genome-wide	**Shattering**	**Seed circularity**	**Brittle rachis**
Little	0.00 - 0.05	18258 (77.3)	26 (83.9)	26 (60.4)	44 (78.6)
Moderate	0.05 - 0.15	4805 (20.3)	3 (9.7)	16 (37.2)	12 (21.4)
High	0.15 - 0.25	476 (2.0)	2 (6.4)	1 (2.3)	0
Very High	> 0.25	72 (0.3)	0	0	0
Total		23611	31	43	56

Level of differentiation of F_{ST} values are from Hartl and Clark (1997).

Table 4 Biallelic reference allele frequency for four breeding cycles in The Land Institute breeding program that had Fixation index, F_{ST}, greater than 0.15

Trait	Marker	Cycle 6	Cycle 7	Cycle 8	Cycle 9
Shattering	S04_341952545	0.74	0.58	0.42	0.40
Shattering	S04_349886731	0.74	0.51	0.41	0.45
Seed circularity & seed length	V01_438389996	0.69	0.78	0.84	0.96

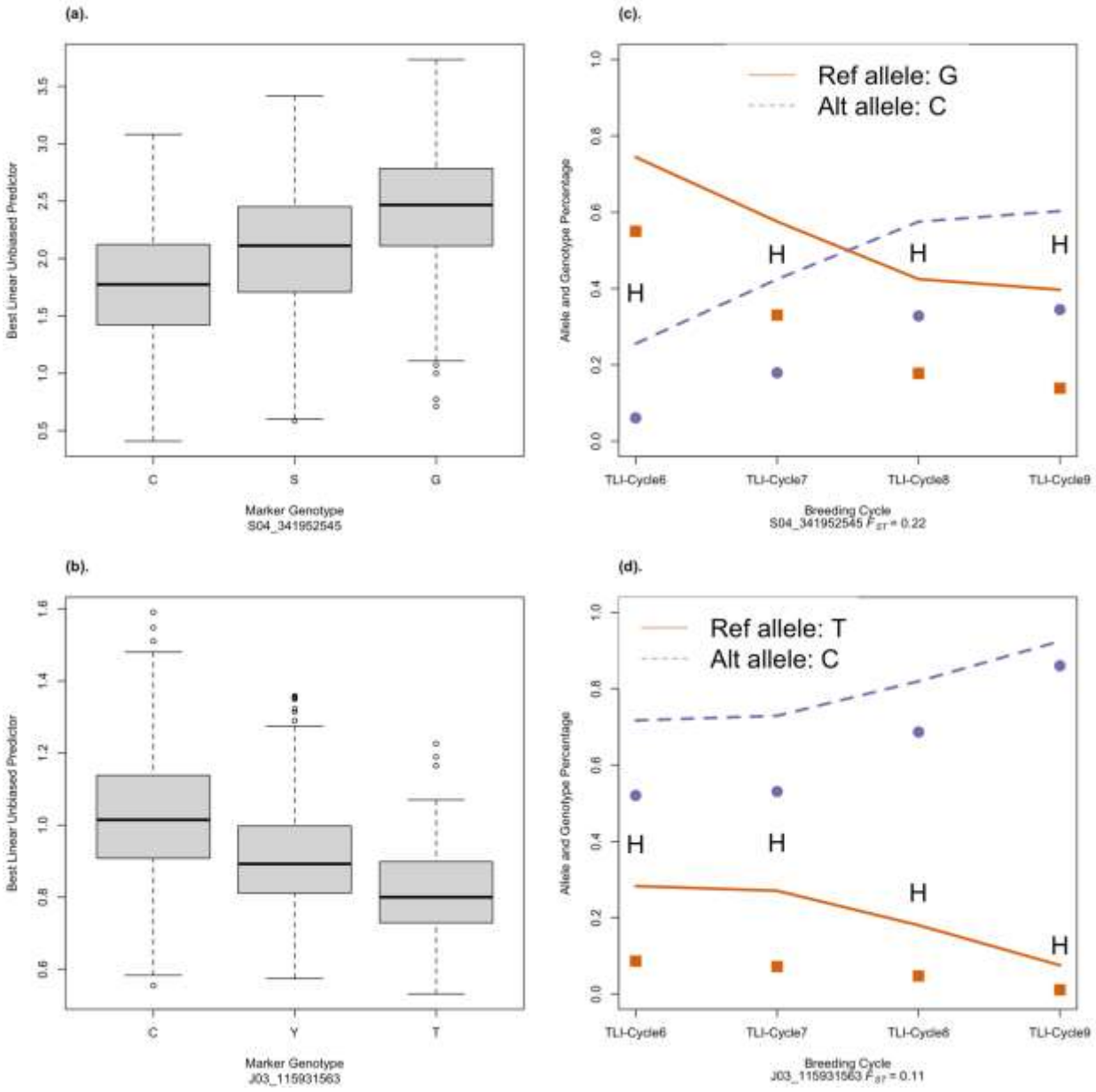

Fig. 3 Distribution of phenotypic values for shattering and brittle rachis (**a** and **b** respectively), where lower values are preferred, at the marker loci S04_341952545 and J03_115931563 in intermediate wheatgrass (*Thinopyrum intermedium*). Panels **c** and **d** displays the allele (line plots) and genotype (points, H is heterozygote) frequency change for the shattering marker in The Land Institute (TLI) Cycle 6 to 9 breeding populations, population differentiation expressed with F_{ST} between TLI Cycle 6 and TLI Cycle 9

DISCUSSION

GWAS in a Breeding Population

The TLI IWG breeding program has completed four cycles of breeding since 2017 using GS with current evidence suggesting that genetic gains can exceed 8% year^{-1} for spike yield and up to 14% year^{-1} for free threshing (Crain et al. 2021a). Given the magnitude of the challenges to domesticate a new crop, identifying genomic regions controlling traits should be a priority within breeding programs as a means to accelerate gains. As such, we used breeding data to complete GWAS analysis for 34 traits. While many markers were identified, there appeared to be little consistency from one cycle-year combination to the next, suggesting genotype by environment interaction between years. Additionally, the PVE explained by markers evaluating cycle by year combinations were generally higher than the combined analysis, ranging from 2 to 15%, suggesting that the exclusion of environmental variance within these models are upwardly biased because of truncation of QTL below the detectable threshold (Beavis 1994; Xu 2003).

Brittle rachis had significant MTAs, but until 2019 this trait was not considered separate from shattering and not measured extensively in the population until 2020. As this marker is associated with a known gene—*Btr2*—and has been found in other IWG populations (Altendorf 2020), it is a prime candidate for the TLI breeding program to select and drive to fixation over the next few cycles. While our data did not show this locus associated with any other trait, work by Larson *et al.* (2019) identified traits including spikelets per spike, seeds per spikelet, seed area, and spike length on this chromosome which could explain the reduction of the favorable allele for brittle rachis.

Even though we evaluated 34 traits, very few significant associations were found, including for spike yield and seed mass, which have been a primary target of selection. Leveraging the large number of observations, we estimated the effective number of QTL in the population. While the estimations hinge on a number of assumptions including QTL distribution (Otto and Jones 2000), additive genetic effect, and random segregation (Hall et al. 2016), they provide an approximation of the complexity of given traits. Considering the ability to detect small QTL, that explained only 0.7% of the variation, and the large number of minimum QTL estimated, there appear to be no large effect QTL that can be targeted by marker-assisted selection for IWG improvement. The

apparent deficiency of large effect QTL could indicate that the intense selection bottleneck of IWG from 14 plants (Zhang et al. 2016) essentially fixed major allele effects in early generations or that large effect genes were not in the founder population. Even though research shows estimated genetic gains for spike yield of up to 8% year^{-1} in TLI-Cycle 7-8 (Crain et al. 2021a) it appears that these gains are from small effect loci. It should be noted that selection within the breeding program was based on GS values and not on any particular marker *per se*. This information suggest that the current breeding material has highly polygenic traits that follow an infinitesimal model (Fisher 1918; Barton et al. 2017). Interestingly, we identified that 2.3% (548 markers) had diverged between TLI-cycle 6 and 9, yet only four of the 217 MTAs showed high or very high F_{ST}. Since genomic selection has been exclusively used for the last four breeding cycles (Crain et al. 2021a), we evaluated the genome-wide estimated marker effects for yield traits. Of the 72 highly diverged markers, 60% and 56% (43 and 40 markers respectively) had coefficients with higher spike yield and seed mass and the mean of all these markers indicated selection was in the favorable direction. As the IWG breeding continues, improvement will be from selecting on many small effect alleles.

Linkage Disequilibrium and Allele Dynamics

Linkage disequilibrium was evaluated across the IWG genome. Within this IWG population, LD declines rapidly relative to closed breeding populations and half-decay rarely exceeded 1 Mb for any chromosome. In comparison to annual wheat which has LD estimated to be 50 Mb (Juliana et al. 2018), LD is 160 fold lower necessitating more markers to cover the genome. If markers were evenly spaced over the 12.75 Gb IWG genome (Vogel et al. 1999) a minimum of 42,000 markers would be needed to ensure all parts of the genome were in LD with a marker to increase mapping resolution. While we leveraged a high-quality draft genome reference sequence, it is possible that our results are influenced by the current genome assembly. It is possible that some of the significant marker positions will change, providing a more complete picture of trait observations and LD dynamics. As sequencing cost decreases and technology improves whole genome sequencing or skim-sequencing approaches could provide an exponential increase in markers to detect genomic associations (Jensen et al. 2020; Pavan et al. 2020).

While LD is estimated to decline rapidly, it should be noted that many of our significant GWAS hits had multiple markers extending beyond the expected LD. This could indicate that selection

has created larger linkage blocks. In maize, another outcrossing species, LD in diverse lines is estimated to extend less than 1 kb (Tenaillon et al. 2001) to over 100 kb in elite maize lines (Rafalski 2002), showing the extent that selection and narrow sets of germplasm can increase the LD block size. As GS has been the method of selection in cycles 7 through 9, this could be creating larger LD blocks for areas that are captured in the genome model. This would include both traits under direct selection such as spike yield as well as any trait that indirectly contributes to priority traits. It is also likely that there are different historical sources of LD, including LD that existed in nature, LD created between the time of collection and the initial bottleneck of selection at the Rodale Institute, and LD created in the TLI breeding program (Zhang et al. 2016; DeHaan et al. 2018).

Comparison to Other IWG GWAS Studies

Within IWG, several other studies have evaluated important domestication and agronomic traits, providing corroboration of key results. Studies have shown that IWG has strong collinearity with the barley genome, (Kantarski et al. 2016; Zhang et al. 2016) providing resources to identify candidate genes. Within a nested association mapping panel, Altendorf (2020) found the same marker as this study for brittle rachis. While this marker is closest to a *Btr2* gene, it is in a 7 Mb region with many *btr*-like genes (Pourkheirandish et al. 2015; Civáň and Brown 2017; DeHaan et al. 2020). Using a bi-parental IWG population, Larson *et al*. (2019) investigated QTL for domestication traits and found several overlaps with the current study. For seed shattering, Larson *et al*. (2019) discovered QTL on chromosome J2 and S4 that align with results found in this study. Chromosome S4 had the most significant seed shattering QTL (LOD > 15.0) in a full-sib family derived from C3_3471, which has been described as the first non-shattering and free-threshing IWG plant (Larson et al. 2019). There was also close alignment with free threshing QTLs located on chromosome V2.

One of the most unanticipated results from this study was the large number markers associated with seed circularity. One potential explanation is the effect of self-incompatibility genes, as these have been shown to have an impact on seed size and fertility in a full-sib mapping population of perennial ryegrass (*Lolium perenne* L.)(Studer et al. 2008). Self-incompatibility (SI) in grasses is controlled by a two locus (*S* and *Z*) system (Lundqvist 1954; Cornish et al. 1979; Baumann et al. 2000), which are located on homoeologous groups 1 and 2 of IWG,

respectively(Larson et al. 2019; Crain et al. 2020b). Self-incompatibility has been documented in IWG (Dewey 1978; Jensen et al. 1990), and previously reported markers for seed area, seed width, seed length, and seed weight by Zhang et al. (2017), Bajgain et al. (2019), and Larson et al. (2019) are located near putative *S* orthogenes on homoeologous group 1 of IWG. Although the mechanism of SI is not completely characterized, Manzanares et al. (2016) demonstrated that a domain of unknown function (*S-DUF247*) is involved in SI reactions. This region has also been associated with seed weight (Zhang *et al.*, 2017; Larson et al. 2019), seed length (Bajgain et al. 2019), and was identified by Crain et al. (2020b) as an active SI locus in IWG. Regardless of whether the loci related to seed circularity are related to potential SI activity or putative control of seed circularity, these loci could be beneficial to the breeding program because seed shape could have an impact on milling quality. Marshall et al. (1984) proposed that spherical seeds maximize volume to surface ratio. IWG has very long and thin seeds (Zhang et al. 2017), so selecting loci that alter seed shape could be used to both improve yield and end product use.

Application to Improving IWG

While we did not find any large effect QTL, our results suggest several potential applications within the breeding program. Our data support that continued use of GS models for breeding and selection is appropriate. While Bajgain *et al.* (2019) suggested using QTL as fixed effects in GS models to improve predictions, none of our detected QTL explained more than 10% of the variance, which would be large enough to be included as a fixed effect as suggested by Bernardo (2014). Second, based on the QTL effect size, genetic mapping studies will require large population sizes to accurately identify and estimate QTL. While the breeding program routinely analyzes c 4,000 plants, this is probably the smallest number of plants needed for genetic mapping based on our power analysis and assessment of genetic architecture for these key traits. Lastly, selection pressure on traits that would indirectly enhance yield, such as the number of florets per spike and florets per spikelet could be increased. Even though MTAs were identified for some of these traits, current results showed minimal allele differentiation between TLI-Cycle 6 and 9. The observed phenotypic variation of these traits suggest that GS can continue to be an effective tool to improve these traits. Along with FSU, biomass production (Vico et al. 2016) and seed set (Armstead et al. 2008) have been suggested as important steps in increasing perennial seed production.

By utilizing data generated within the breeding program, this study identified MTA for several agronomic and domestication traits, all of which had small effects suggesting that traits are highly polygenic. Even though no QTL were identified for spike yield and seed mass, several QTL were found for component traits of yield, suggesting that genetic control of these traits is from many small effect loci. Previous breeding efforts have increased spike yield by 77% and seed mass by 23% over two breeding cycles (DeHaan et al. 2014). These results, coupled with genetic gain estimates of over 8% $year^{-1}$ by Crain *et al*. (2021a), suggest that the current breeding program has considerable variability, providing opportunity for continued improvement well into the future. There appear to be few large effect QTL, with the substantial genetic progress having been made through selection of genome wide loci with small effects. These observations support a continued focus on classic breeding methods based on an underlying infinitesimal model of genetic architecture (Fisher 1918; Barton et al. 2017) and further implementation of genomic selection. The challenge of developing perennial grains is daunting, yet the knowledge generated from this study will help select high yielding and high performing genets, leading to large scale perennial grain crops.

ACKNOWLEDGEMENTS

This work was funded in part by the Perennial Agriculture Project in conjunction with the Malone Family Land Preservation Foundation and The Land Institute. Shuangye Wu provided vital laboratory support, and Marty Christians provided invaluable field assistance. The *Thinopyrum intermedium* Genome Sequencing Consortium provided prepublication access to the IWG genome sequence. Computational work was completed on the Beocat Research Cluster at Kansas State University, which is funded in part by NSF grants CNS-1006860, EPS-1006860, and EPS-0919443. Contribution no. 22-046-J from the Kansas Agricultural Experiment Station.

DECLARATIONS

Funding This work was funded by The Perennial Agriculture Project, in conjunction with the Malone Family Land Preservation Foundation and The Land Institute.

Conflicts of interest/Competing interest The authors declare that they have no conflict of interest.

Availability of data and material Original sequence data has been uploaded to the NCBI sequence read archive (SRA) (https://www.ncbi.nlm.nih.gov/bioproject/) as part of the umbrella BioProject PRJNA609325.

Code availability R code generated for this study has been deposited in the Dryad digital Repository, https://doi.org/10.5061/dryad.rbnzs7hb7

Authors' contributions JC, SL, LD and JP planned and designed the research. JC, SL, LD and KD performed experiments, conducted fieldwork, and collected data. JC analyzed data. JC, SL, LD, and JP wrote the manuscript.

REFERENCES

Alonso MP, Mirabella NE, Panelo JS, et al (2018) Selection for high spike fertility index increases genetic progress in grain yield and stability in bread wheat. Euphytica 214:. https://doi.org/10.1007/s10681-018-2193-4

Altendorf KR (2020) Phenotypic and genetic characterization of domestication and yield component traits in the perennial grain crop intermediate wheatgrass. University of Minnesota

Altendorf KR, DeHaan LR, Heineck GC, et al (2021a) Floret site utilization and reproductive tiller number are primary components of grain yield in intermediate wheatgrass spaced plants. Crop Sci 61:1073–1088. https://doi.org/10.1002/csc2.20385

Altendorf KR, Larson SR, DeHaan LR, et al (2021b) Nested association mapping reveals the genetic architecture of spike emergence and anthesis timing in intermediate wheatgrass. G3 Genes|Genomes|Genetics 11:. https://doi.org/10.1093/g3journal/jkab025

Armstead IP, Turner LB, Marshall AH, et al (2008) Identifying genetic components controlling fertility in the outcrossing grass species perennial ryegrass (Lolium perenne) by quantitative trait loci analysis and comparative genetics. New Phytol 178:559–571. https://doi.org/10.1111/j.1469-8137.2008.02413.x

Bajgain P, Zhang X, Anderson JA (2020a) Dominance and G×E interaction effects improve genomic prediction and genetic gain in intermediate wheatgrass (Thinopyrum intermedium). Plant Genome 13:1–13. https://doi.org/10.1002/tpg2.20012

Bajgain P, Zhang X, Anderson JA (2019) Genome-wide association study of yield component traits in intermediate wheatgrass and implications in genomic selection and breeding. Genes|Genomes|Genetics g3.400073.2019. https://doi.org/10.1534/g3.119.400073

Bajgain P, Zhang X, Jungers JM, et al (2020b) 'MN-Clearwater', the first food-grade intermediate wheatgrass (Kernza perennial grain) cultivar. J Plant Regist 14:288–297. https://doi.org/10.1002/plr2.20042

Barton NH, Etheridge AM, Véber A (2017) The infinitesimal model: Definition, derivation, and implications. Theor Popul Biol 118:50–73. https://doi.org/10.1016/j.tpb.2017.06.001

Baumann U, Bian X, Langridge P (2000) Self-incompatibility in the grasses. Ann Bot 85:203–209. https://doi.org/10.1007/978-3-540-68486-2_13

Beavis WD (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: Proceedings of the forty-ninth annual corn and sorghum industry research conference. p 266

Becker R, Wagoner P, Hanners GD, Saunders RM (1991) Compositional, nutritional and functional evaluation of intermediate wheatgrass (Thinopyrum intermedium). J Food Process Preserv 15:63–77. https://doi.org/10.1111/j.1745-4549.1991.tb00154.x

Bernardo R (2014) Genomewide selection when major genes are known. Crop Sci 54:68–75. https://doi.org/10.2135/cropsci2013.05.0315

Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. Am J Hum Genet 98:116–126. https://doi.org/10.1016/j.ajhg.2015.11.020

Cattani D (2016) Selection of a perennial grain for seed productivity across years: Intermediate wheatgrass as a test species. Can J Plant Sci 524:516–524. https://doi.org/10.1139/cjps-

650 2016-0280
651 Civáň P, Brown TA (2017) A novel mutation conferring the nonbrittle phenotype of cultivated
652 barley. New Phytol 214:468–472. https://doi.org/10.1111/nph.14377
653 Cornish MA, Hayward MD, Lawrence MJ (1979) Self-incompatibility in ryegrass: I. Genetic
654 control in diploid lolium perenne L. Heredity (Edinb) 43:95–106.
655 https://doi.org/10.1038/hdy.1979.63
656 Crain J, Bajgain P, Anderson J, et al (2020a) Enhancing crop domestication through genomic
657 selection, a case study of intermediate wheatgrass. Front Plant Sci 11:1–15.
658 https://doi.org/10.3389/fpls.2020.00319
659 Crain J, DeHaan L, Poland J (2021a) Genomic prediction enables rapid selection of high-
660 performing genets in an intermediate wheatgrass breeding program. Plant Genome.
661 https://doi.org/10.1002/tpg2.20080
662 Crain J, Haghighattalab A, DeHaan L, Poland JA (2021b) Development of whole-genome
663 prediction models to increase the rate of genetic gain in intermediate wheatgrass
664 (Thinopyrum intermedium) breeding. Plant Genome
665 Crain J, Larson S, Dorn K, et al (2020b) Sequenced-based paternity analysis to improve breeding
666 and identify self-incompatibility loci in intermediate wheatgrass (Thinopyrum
667 intermedium). Theor Appl Genet 133:3217–3233. https://doi.org/10.1007/s00122-020-
668 03666-1
669 Crews TE, Carton W, Olsson L (2018) Is the future of agriculture perennial? Imperatives and
670 opportunities to reinvent agriculture by shifting from annual monocultures to perennial
671 polycultures. Glob Sustain 1:. https://doi.org/10.1017/sus.2018.11
672 Culman SW, DuPont ST, Glover JD, et al (2010) Long-term impacts of high-input annual
673 cropping and unfertilized perennial grass production on soil properties and belowground
674 food webs in Kansas, USA. Agric Ecosyst Environ 137:13–24.
675 https://doi.org/10.1016/j.agee.2009.11.008
676 Culman SW, Snapp SS, Ollenburger M, et al (2013) Soil and water quality rapidly responds to
677 the perennial grain Kernza wheatgrass. Agron J 105:735–744.
678 https://doi.org/10.2134/agronj2012.0273
679 DeHaan L, Christians M, Crain J, Poland J (2018) Development and evolution of an intermediate
680 wheatgrass domestication program. Sustainability 1–19.
681 https://doi.org/10.3390/su10051499
682 DeHaan L, Larson S, López-Marqués RL, et al (2020) Roadmap for Accelerated Domestication
683 of an Emerging Perennial Grain Crop. Trends Plant Sci 25:525–537.
684 https://doi.org/10.1016/j.tplants.2020.02.004
685 DeHaan LR, Ismail BP (2017) Perennial cereals provide ecosystem benefits. Cereal Foods World
686 62:278–281. https://doi.org/10.1094/CFW-62-6-0278
687 DeHaan LR, Wang S, Larson SR, et al (2014) Current efforts to develop perennial wheat and
688 domesticate Thinopyrum intermedium as a perennial grain. In: Batello C, Wade L, Cox S, et
689 al. (eds) Perennial Crops for Food Security Proceedings of the FAO Expert Workshop, 28–
690 30 Aug. 2013. pp 72–89
691 Dewey DR (1978) Intermediate wheatgrasses of Iran. Crop Sci 18:43.
692 https://doi.org/10.2135/cropsci1978.0011183x001800010012x
693 Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package
694 rrBLUP. Plant Genome J 4:250–255. https://doi.org/10.3835/plantgenome2011.08.0024
695 Endelman JB, Jannink JL (2012) Shrinkage estimation of the realized relationship matrix.

696 Genes|Genomes|Genetics 2:1405–1413. https://doi.org/10.1534/g3.112.004259
697 Fisher RA (1918) The correlation between relatives on the supposition of Mendelian interitance.
698 Trans R Soc Edinburgh 52:399–433
699 Flint-Garcia SA, Thornsberry JM, Edward IV SB (2003) Structure of linkage disequilibrium in
700 plants. Annu Rev Plant Biol 54:357–374.
701 https://doi.org/10.1146/annurev.arplant.54.031902.134907
702 Gilmour AR, Gogel BJ, Cullis BR, et al (2015) ASReml User Guide Release 4.1 Functional
703 Specification. VSN International Ltd., Hemel, Hempstead, UK
704 Glaubitz JCJ, Casstevens TMT, Lu F, et al (2014) TASSEL-GBS: A high capacity genotyping
705 by sequencing analysis pipeline. PLoS One 9:e90346.
706 https://doi.org/10.1371/journal.pone.0090346
707 Glover JD, Reganold JP, Bell LW, et al (2010) Increased food and ecosystem security via
708 perennial grains. Science (80-.). 328:1638–1639
709 Hall D, Hallingbäck HR, Wu HX (2016) Estimation of number and size of QTL effects in forest
710 tree traits. Tree Genet Genomes 12:1–17. https://doi.org/10.1007/s11295-016-1073-0
711 Hartl DL, Clark AG (1997) Principles of population genetics
712 Hill WG, Weir BS (1988) Variances and covariances of squared linkage disequilibria in finite
713 populations. Theor Popul Biol 33:54–78. https://doi.org/10.1016/0040-5809(88)90004-4
714 Isik F, Holland J, Maltecca C (2017) Genetic data analysis for plant and animal breeding.
715 Springer
716 Jensen KB, Dewey DR, Zhang YFY, Dewey DR (1990) Mode of pollination of perennial species
717 of the Triticeae in relation to genomically defined genera. Can J Plant Sci 70:215–225.
718 https://doi.org/10.4141/cjps90-024
719 Jensen SE, Charles JR, Muleta K, et al (2020) A sorghum practical haplotype graph facilitates
720 genome-wide imputation and cost-effective genomic prediction. Plant Genome 13:1–15.
721 https://doi.org/10.1002/tpg2.20009
722 Juliana P, Singh RP, Singh PK, et al (2018) Genome-wide association mapping for resistance to
723 leaf rust, stripe rust and tan spot in wheat reveals potential candidate genes. Theor Appl
724 Genet 131:1405–1422. https://doi.org/10.1007/s00122-018-3086-6
725 Jungers JM, DeHaan LH, Mulla DJ, et al (2019) Reduced nitrate leaching in a perennial grain
726 crop compared to maize in the Upper Midwest, USA. Agric Ecosyst Environ 272:63–73.
727 https://doi.org/10.1016/j.agee.2018.11.007
728 K.W. B, S. S (2009) A Guide to QTL Mapping with R/qtl. Springer
729 Kantarski T, Larson S, Zhang X, et al (2016) Development of the first consensus genetic map of
730 intermediate wheatgrass (Thinopyrum intermedium) using genotyping-by-sequencing.
731 Theor Appl Genet 1–14. https://doi.org/10.1007/s00122-016-2799-7
732 Keenan K, Mcginnity P, Cross TF, et al (2013) DiveRsity: An R package for the estimation and
733 exploration of population genetics parameters and their associated errors. Methods Ecol
734 Evol 4:782–788. https://doi.org/10.1111/2041-210X.12067
735 Larson S, DeHaan L, Poland J, et al (2019) Genome mapping of quantitative trait loci (QTL)
736 controlling domestication traits of intermediate wheatgrass (Thinopyrum intermedium).
737 Theor Appl Genet 132:2325–2351. https://doi.org/10.1007/s00122-019-03357-6
738 Lindberg CL, Hanslin HM, Schubert M, et al (2020) Increased above-ground resource allocation
739 is a likely precursor for independent evolutionary origins of annuality in the Pooideae grass
740 subfamily. New Phytol 228:318–329. https://doi.org/10.1111/nph.16666
741 Lundqvist A (1954) Studies on self-sterility in rye, Secale cereale. L Hered 40:278–294.

https://doi.org/10.1111/j.1601-5223.1954.tb02973.x
Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Sunderland, MA
Manzanares C, Barth S, Thorogood D, et al (2016) A gene encoding a DUF247 domain protein cosegregates with the s self-incompatibility locus in perennial ryegrass. Mol Biol Evol 33:870–884. https://doi.org/10.1093/molbev/msv335
Marshall DR, Ellison FW, Mares DJ (1984) Effects of grain shape and size on milling yields in wheat. I. Theoretical analysis based on simple geometric models. Aust J Agric Res 35:619–630. https://doi.org/10.1071/AR9840619
Otto SP, Jones CD (2000) Detecting the undetected: Estimating the total number of loci underlying a quantitative trait. Genetics 156:2093–2107
Pavan S, Delvento C, Ricciardi L, et al (2020) Recommendations for choosing the genotyping method and best practices for quality control in crop genome-wide association studies. Front Genet 11:. https://doi.org/10.3389/fgene.2020.00447
Philipp N, Weichert H, Bohra U, et al (2018) Grain number and grain yield distribution along the spike remain stable despite breeding for high yield in winter wheat. PLoS One 13:. https://doi.org/10.1371/journal.pone.0205452
Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:. https://doi.org/10.1371/journal.pone.0032253
Pourkheirandish M, Hensel G, Kilian B, et al (2015) Evolution of the Grain Dispersal System in Barley. Cell 162:527–539. https://doi.org/10.1016/j.cell.2015.07.002
Pugliese JY, Culman SW, Sprunger CD (2019) Harvesting forage of the perennial grain crop Kernza (Thinopyrum intermedium) increases root biomass and soil nitrogen cycling. Plant Soil 437:241–254. https://doi.org/10.1007/s11104-019-03974-6
R Core Team (2020) R: a language and environment for statistical computing
Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100. https://doi.org/10.1016/S1369-5266(02)00240-6
Sehgal D, Mondal S, Crespo-Herrera L, et al (2020) Haplotype-based, genome-wide association study reveals stable genomic regions for grain yield in CIMMYT spring bread wheat. Front Genet 11:1–13. https://doi.org/10.3389/fgene.2020.589490
Smaje C (2015) The strong perennial vision: a critical review. Agroecol Sustain Food Syst 39:471–499. https://doi.org/10.1080/21683565.2015.1007200
Sprunger CD, Culman SW, Robertson GP, Snapp SS (2018) Perennial grain on a Midwest Alfisol shows no sign of early soil carbon gain. Renew Agric Food Syst 33:360–372. https://doi.org/10.1017/S1742170517000138
Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. 100:6. https://doi.org/10.1073/pnas.91.25.12091
Studer B, Jensen LB, Hentrup S, et al (2008) Genetic characterisation of seed yield and fertility traits in perennial ryegrass (Lolium perenne L.). Theor Appl Genet 117:781–791. https://doi.org/10.1007/s00122-008-0819-y
Tenaillon MI, Sawkins MC, Anderson LK, et al (2001) Patterns of diversity and recombination along chromosome 1 of maize (Zea mays ssp. mays L.). Proc Natl Acad Sci 98:9161–9166. https://doi.org/10.1093/genetics/162.3.1401
Therneau TM (2020) coxme: Mixed Effects Cox Models
Turner S (2017) qqman: Q-Q and Manhattan Plots for GWAS Data
Vico G, Manzoni S, Nkurunziza L, et al (2016) Trade-offs between seed output and life span - a

quantitative comparison of traits between annual and perennial congeneric species. New Phytol 209:104–114. https://doi.org/10.1111/nph.13574
Vogel KP, Arumuganathan K, Jensen KB (1999) Nuclear DNA content of perennial grasses of the Triticeae. Crop Sci 39:661–667. https://doi.org/10.2135/cropsci1999.0011183X003900020009x
Wagoner P (1990a) Perennial grain development: past efforts and potential for the future. CRC Crit Rev Plant Sci 9:381–408. https://doi.org/10.1080/07352689009382298
Wagoner P (1990b) Perennial grain new use for intermediate wheatgrass. J Soil Water Conserv 45:81–82
Wang M, Xu S (2019) Statistical power in genome-wide association studies and quantitative trait locus mapping. Heredity (Edinb) 123:287–306. https://doi.org/10.1038/s41437-019-0205-3
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution (N Y) 38:1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x
Würschum T, Leiser WL, Langer SM, et al (2018) Phenotypic and genetic analysis of spike and kernel characteristics in wheat reveals long-term genetic trends of grain yield components. Theor Appl Genet 131:2071–2084. https://doi.org/10.1007/s00122-018-3133-3
Xu S (2003) Theoretical Basis of the Beavis Effect. Genetics 165:2259–2268
Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–8. https://doi.org/10.1038/ng1702
Zhang X, Larson SR, Gao L, et al (2017) Uncovering the genetic architecture of seed weight and size in intermediate wheatgrass through linkage and association mapping. Plant Genome 10:1–15. https://doi.org/10.3835/plantgenome2017.03.0022
Zhang X, Sallam A, Gao L, et al (2016) Establishment and optimization of genomic selection to accelerate the domestication and improvement of intermediate wheatgrass. Plant Genome 9:1–18. https://doi.org/10.3835/plantgenome2015.07.0059
Zhang Z, Ersoz E, Lai CQ, et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–60. https://doi.org/10.1038/ng.546
Zhong J, van Esse GW, Bi X, et al (2021) INTERMEDIUM-M encodes an HvAP2L-H5 ortholog and is required for inflorescence indeterminacy and spikelet determinacy in barley. Proc Natl Acad Sci U S A 118:. https://doi.org/10.1073/pnas.2011779118

FIGURES

Fig. 1 genome-wide linkage disequilibrium (LD) for intermediate wheatgrass (*Thinopyrum intermedium*) for 200 Mb region (**a**) and 5 Mb region (**b**). Orange points represent individual marker combinations with a 250-marker sliding window. Average LD has been computed with the Hill and Weir formula (1988) and shown in blue

Fig. 2 Manhattan plots of (**a**) shattering, (**b**) brittle rachis, and (**c**) seed circularity in intermediate wheatgrass (*Thinopyrum intermedium*) with line indicating 0.05 false discovery rate. Panels **d**, **e** and **f** show quantile-quantile (QQ) plots for p values under the null hypothesis (no association) and observed p values for brittle rachis (d), shattering (e), and seed circularity (f), respectively

Fig. 3 Distribution of phenotypic values for shattering and brittle rachis (**a** and **b** respectively), where lower values are preferred, at the marker loci SS04_341952545 and SJ03_115931563 in intermediate wheatgrass (*Thinopyrum intermedium*). Panels **c** and **d** displays the allele (line plots) and genotype (points, H is heterozygote) frequency change for the shattering marker in The Land Institute (TLI) Cycle 6 to 9 breeding populations, population differentiation expressed with F_{ST} between TLI Cycle 6 and TLI Cycle 9

TABLES

Table 1 Significant quantitative trait loci (QTL) associations in The Land Institute intermediate wheatgrass (*Thinopyrum intermedium*) breeding program for priority traits

Table 2 Number of estimated quantitative trait loci (QTL) for phenotypic traits in intermediate wheatgrass, and the estimated population size to needed to detect QTLs explaining 50% of the genotypic variance

Table 3 Fixation index, F_{ST}, values for all single nucleotide polymorphisms (SNPs) in The Land Institute intermediate wheatgrass breeding population cycles 6 and 9

Table 4 Biallelic reference allele frequency for four breeding cycles in The Land Institute breeding program that had Fixation index, $F_{ST,}$ greater than 0.15

SUPPORTING INFORMATION

Supporting Information Table S1 Descriptive statistics, including total number and number of genets by cycle, of the best linear unbiased predictors of 34 traits measured in The Land Institute intermediate wheatgrass breeding program cycles 6-9 during 2016-2020

Trait	C6 n	C7 n	C8 n	C9 n	Total n	Min.	Max.	Mean	STD
Brittle Rachis	0	1186	947	955	3088	0.49	1.63	0.99	0.17
Seed Image Circularity	1013	1151	940	926	4030	0.37	0.55	0.45	0.02
Spike Emergence (cm)	0	1180	0	0	1180	1.8	36.69	19.79	6.13
Seed Length (mm)	1013	1183	976	926	4098	5.07	8.13	6.68	0.41
Seed Width (mm)	1013	1183	976	926	4098	1.29	1.8	1.55	0.07
Flag Leaf Length (cm)	982	0	0	0	982	20.18	36.03	27.87	2.23
Flag Leaf Width (mm)	982	0	0	0	982	8.83	23.79	15.75	1.91
Flag Leaf Height (cm)	924	0	0	0	924	37.51	78.12	63.86	5.71
Number of Florets per Spike	1019	1176	873	0	3068	123.17	230.38	174.13	14.49
Number of Florets per Spikelet	1019	1176	873	0	3068	6.41	9.79	7.66	0.44
Spike Emergence %	918	1130	0	0	2048	0.05	1.31	0.64	0.18
Lodging	0	1182	938	0	2120	2.42	8.89	5.98	1.05
Free Threshing	1014	1184	979	937	4114	2.3	98.37	57.17	17.1
Peduncle Width (mm)	0	1150	0	0	1150	7.71	11.99	9.9	0.69
Floret Site Utilization	1012	1173	864	0	3049	0.12	0.39	0.22	0.03
Plant Height (cm)	1020	1191	992	954	4157	82.87	136.54	111.03	5.92
Seed Area (mm2)	1013	1183	976	926	4098	5.39	10.59	8.08	0.68
Seed Density	1013	1183	976	926	4098	0.97	1.67	1.31	0.08
Seeds per Spike	1016	1186	980	958	4140	10.7	62.24	38.21	6.8
Seed Mass (mg)	1013	1183	977	937	4110	5.89	13.34	9.46	1.08
Seed Perimeter (mm)	1013	0	0	0	1013	11.2	17.55	14.73	0.95
Shattering	1019	1188	981	948	4136	0.17	3.79	2.13	0.57
Spikelet Density	1017	0	0	0	1017	0.52	1.06	0.72	0.06
Spike Dry Weight (g)	0	1167	955	959	3081	0.83	1.75	1.31	0.13
Spiklets per Spike	1019	1176	873	0	3068	18.02	27.36	22.69	1.13
Spike Harvest Index	0	1156	955	958	3069	0.11	0.46	0.3	0.05
Spike Length (cm)	1017	1141	0	0	2158	22.59	45.88	34.78	2.86
Spike Yield (g)	1019	1188	981	958	4146	0.14	0.58	0.36	0.06
Stem Angle	1022	1179	0	0	2201	19.26	82.33	53.86	9.99
Stem Strength Bottom	914	0	0	0	914	1214.48	3424.35	2181.99	344.09

Stem Diameter (mm)	946	0	0	0	946	2.24	3.89	3.01	0.21
Stem Strength Middle	915	1174	0	0	2089	747.15	1910.92	1197.3	166.77
Stem Strength Top	915	0	0	0	915	288.65	615.32	418.66	53.05
Maturity	1017	1189	944	953	4103	52.21	69.9	63.14	1.48

859

Supporting Information Table S2 Significant marker trait associations for The Land Institute intermediate wheatgrass breeding program analyzed by combined data and cycle-year combinations

QTL#, number of QTL; Associated markers, number of markers associated with each QTL; Chr, chromosome; n, number of individuals observed; F_{ST} fixation index between The Land Institute (TLI) Cycle 6 and TLI Cycle 9; PVE, percent variance explained; Ref/Alt reference and alternate allele respectively; MAF

Supporting Information Table S3 Chromosome location of genome-wide associations by trait for combined analysis (C) and individual cycle combinations (6-9) for The Land Institute intermediate wheatgrass breeding program

Trait	Chromosome																
	J2	J3	J4	J5	J6	J7	S1	S2	S3	S4	S5	S7	V1	V2	V3	V5	V6
Brittle rachis		C,7,8,9							C,7,9		C,7	9	7			7	
Flag leaf length		6															
Flag leaf width													6			6	
Floret site utilization											C,7						
Free threshing														C,9			
Maturity	7				6												
Number of florets per spike				C											C		
Number of florets per spikelet	C			C													
Peduncle width	C,7																
Plant height					7,8												
Seed area						6						8					
Seed density							C					C					
Seed image circularity		C					C,7,8			C,6		C,6,7,8	C,6,7	C		C,6,7	
Seed length												7	C				C
Seed width	C		C,7			9								C,6			
Shattering	C						7	C,7,8,9	7	C,7,8,9	C,7,8			7			
Spike emergence	C,7																
Spike emergence %	C																
Spike length											C		7				
Spikelet density									6						6		
Spikelets per spike								C				6					
Stem strength bottom									6								
Stem diameter						C,6											
Stem strength middle									6								

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportingFigureS1.png
- SupportingFigureS2.png
- SupportingFigureS3.png
- SupportingFigureS4.png
- SupportingInformationTable2.xlsx