2016

Accuracy of Genomic Prediction in Switchgrass (Panicum virgatum L.) Improved by Accounting for Linkage Disequilibrium

Guillaume P. Ramstein  
University of Wisconsin-Madison, ramstein@wisc.edu

Joseph Evans  
Michigan State University

Shawn M. Kaeppler  
University of Wisconsin-Madison

Robert B. Mitchell  
University of Nebraska-Lincoln

Kenneth P. Vogel  
University of Nebraska-Lincoln

See next page for additional authors
Follow this and additional works at: http://digitalcommons.unl.edu/agronomyfacpub

Part of the Agricultural Science Commons, Agriculture Commons, Agronomy and Crop Sciences Commons, Botany Commons, Horticulture Commons, Other Plant Sciences Commons, and the Plant Biology Commons

http://digitalcommons.unl.edu/agronomyfacpub/895

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Accuracy of Genomic Prediction in Switchgrass 
(\textit{Panicum virgatum} L.) Improved by Accounting for Linkage Disequilibrium

Guillaume P. Ramstein,*†Joseph Evans,†‡Shawn M. Kaeppler,*§Robert B. Mitchell,**
Kenneth P. Vogel,** C. Robin Buell,†‡and Michael D. Casler*††

*Department of Agronomy, and §Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, WI 53706, †Department of Energy Great Lakes Bioenergy Research Center and ‡Department of Plant Biology, Michigan State University, East Lansing, MI 48824, **Grain, Forage, and Bioenergy Research Unit, Agricultural Research Service, United States Department of Agriculture, University of Nebraska, Lincoln, NE 68583-0937, and ††Agricultural Research Service, United States Department of Agriculture, Madison, WI 53706

ABSTRACT Switchgrass is a relatively high-yielding and environmentally sustainable biomass crop, but further genetic gains in biomass yield must be achieved to make it an economically viable bioenergy feedstock. Genomic selection (GS) is an attractive technology to generate rapid genetic gains in switchgrass, and meet the goals of a substantial displacement of petroleum use with biofuels in the near future. In this study, we empirically assessed prediction procedures for genomic selection in two different populations, consisting of 137 and 110 half-sib families of switchgrass, tested in two locations in the United States for three agronomic traits: dry matter yield, plant height, and heading date. Marker data were produced for the families’ parents by exome capture sequencing, generating up to 141,030 polymorphic markers with available genomic-location and annotation information. We evaluated prediction procedures that varied not only by learning schemes and prediction models, but also by the way the data were preprocessed to account for redundancy in marker information. More complex genomic prediction procedures were generally not significantly more accurate than the simplest procedure, likely due to limited population sizes. Nevertheless, a highly significant gain in prediction accuracy was achieved by transforming the marker data through a marker correlation matrix. Our results suggest that marker-data transformations and, more generally, the account of linkage disequilibrium among markers, offer valuable opportunities for improving prediction procedures in GS. Some of the achieved prediction accuracies should motivate implementation of GS in switchgrass breeding programs.

KEYWORDS genomic selection linkage disequilibrium exome capture bioenergy \textit{Panicum virgatum} L. GenPred Shared data resource

Genomic selection (GS) is the use of genome-wide marker information to predict genotype performance in breeding programs. Predictions in GS should be sufficiently accurate so that it is economically more viable to select individuals based solely on markers, rather than phenotypic measurements (Muir 2007; Lorenz \textit{et al.} 2011; Riedelsheimer and Melchinger 2013). In switchgrass (\textit{Panicum virgatum} L.), a perennial forage crop, this potential advantage derives from the fact that phenotypic measurements generally require 2–3 yr of field testing (one establishment year and 1–2 yr of trial; Casler and Brummer 2008), whereas acquiring genome-wide marker information would require less than a year (Resende \textit{et al.} 2014). The US Department of Agriculture and the US Department of Energy intend to make switchgrass a principal source of biofuel in the US by 2030, so as to meet the goal of displacing 30% of petroleum use with biofuel (Sanderson \textit{et al.} 1996; Perlack \textit{et al.} 2005). However, strong and rapid genetic gains for biomass yield to approximately 20 Mg/ha are required to meet this goal (Perlack \textit{et al.} 2005). Therefore, in the case of switchgrass breeding, GS is a technology that is not only economically attractive, but also strategically useful.
The first genomic prediction procedure, introduced by Meuwissen et al. (2001)—ridge regression-BLUP (RR-BLUP), equivalent to genomic BLUP (G BLUP) (Hayes et al. 2009)—assumed an infinitesimal genetic model (i.e., all loci in linkage equilibrium with additive effects of equal variance), and a homogeneous genetic architecture throughout individuals. Since then, improvements in genomic prediction models have been made, most notably through the account of (i) marker-effect heteroscedasticity, i.e., differential variances of marker effects (Meuwissen et al. 2001; Park and Casella 2008; Erbe et al. 2012; Shen et al. 2013); (ii) nonlinear marker effects and interactions between markers (Gianola and van Kaam 2008; Akdemir and Jannink 2015); (iii) genotype-by-environment interactions (Burgueño et al. 2012; Guo et al. 2013; Heslot et al. 2014); (iv) correlation among traits of interest (Calus and Veerkamp 2011; Jia and Jannink 2012); and (v) population heterogeneity (Harris and Johnson 2010; Makgahla et al. 2013; Karoui et al. 2012; Isidro et al. 2015).

Marker-data transformations have previously been used in GS to account for marker-effect heteroscedasticity, with weights on marker variables reflecting the significance of their relationship with the outcome of interest (de los Campos et al. 2013; Su et al. 2014; He et al. 2015). Transformations on marker data have also been used to account for redundancy in marker information due to linkage disequilibrium (LD), through principal component analysis (PCA) (Long et al. 2011); or weights on marker variables; so as to reflect the degree of tagging of loci by markers (Speed et al. 2012; Nishio and Satoh 2015). However, to date, few empirical studies in GS have been conducted to assess the usefulness of preprocessing marker data in order to account for LD.

In this study, we assessed the possibility of producing reliable predictions for GS in switchgrass. Our data consisted of relatively few observations on two distinct populations. A total of 247 individuals were genotyped by exome capture sequencing, and evaluated for important agronomic traits: biomass yield, plant height, and heading date. We assessed various types of prediction procedures, which differed from the standard procedure (G BLUP on nontransformed marker data) not only by the prediction model—which might account for heteroscedasticity and/or nonlinearity of marker effects—but also by the type of marker-data transformation—which might account for LD among markers. We also examined the potential of learning schemes (training-set designs) for improving genomic prediction procedures. While multiple-trait models may be useful to account for genotype-by-environment interactions, correlation among traits and/or population heterogeneity, such models were not found useful here because they either were not statistically efficient enough, or failed to effectively fit the data, certainly as a result of our small sample sizes.

MATERIALS AND METHODS

Populations assayed

Genomic selection (GS) in switchgrass was studied in two tetraploid genotypes. The first population comprised 137 half-sib (HS) families developed from WS4U, an upland-ecotype germplasm pool of 162 plants (Casler et al. 2006). The HS families were the progeny of genotypes produced in cycle 2 (C2) of selection on WS4U for high biomass yield (Casler 2010; Casler and Vogel 2014). The second population comprised 110 HS families developed from the cultivar Liberty, which is a stabilized lowland-upland hybrid cultivar (Vogel et al. 2014; Casler and Vogel 2014). The HS families were developed from the Liberty population by an additional breeding generation. They were the progeny of the genotypes selected in C2 for high biomass yield, excellent winter survival, excellent spring greenup, and no apparent diseases using the amond and within family breeding method. In both populations, the selected C2 genotypes were polycrossed in isolation. The two populations, hereafter referred to as WS4U-C2 and Liberty-C2, were tested in two locations in the United States: Arlington (WI) and Mead (NE) in 2012, 2013, and 2014. Families were assayed in a row-plot trial and replicated in a randomized complete block design (RCBD), with four family replicates for WS4U-C2, and three family replicates for Liberty-C2. There were up to five HS in each family replicate, with different HS between replicates (there was no vegetative propagation of individual plants). Rows were spaced 0.9 m apart, and plants were spaced 0.45 m apart within rows. Plots were established from greenhouse-grown seedlings in May 2011, and fertilized with 110 kg N ha⁻¹ in May of 2012 and 2013.

Phenotypic data and mixed-model analyses

In this study, trait measurements at different locations were considered different outcomes. There were six outcomes: dry matter yield (DMY), plant height (PH), and heading date (HD) in WI or NE. HD was scored on each individual plant as day-of-year when half of the panicles of a plant had fully emerged from the boot. PH was measured on each individual plant from the ground to the top of the tallest tiller after growth had ceased in late September. DMY was determined by harvesting each row plot with a flail harvester at a 10-cm cutting height, adjusted for moisture content using a 400-g sample of harvested tissue dried for 7 d at 60°C. Outcomes were measured in 2012 and 2013, with the exception of HD in WI, which was scored in 2013 and 2014 for WS4U-C2 and in 2012, 2013, and 2014 for Liberty-C2 (Table 1).

For PH and HD outcomes, measured on an individual-plant basis, the following linear mixed model was fitted:

\[ y_{ijkl} = \mu + g_i + b_j + t_k + (g \times b)_{ij} + (g \times t)_{ijk} + (b \times t)_{jk} + e_{ijkl} \]

where \( \mu \) is the grand mean; \( g_i, b_j, \) and \( t_k \) are the random effects of HS family, i, block j, and year k, respectively; \( \times \) indicates interactions; \( e_{ijkl} \) are residuals. For each term, the corresponding effects were modeled as independent and identically normally distributed. For PH (in both locations), an additional term \( plot_{ij} \) ~ Normal(0, \( \Sigma_p \sigma_{plot}^2 \)) was included in the model (on the basis of a lower Bayesian information criterion), where \( \Sigma_p \) is the Kronecker product of the first-order autoregressive covariance matrices on rows and on columns, respectively.

For DMY outcomes, measured on a plot basis, the following linear mixed model was fitted:

\[ y_{ijk} = \mu + g_i + b_j + t_k + (g \times b)_{ij} + (g \times t)_{ik} + (b \times t)_{jk} + e_{ijk} \]

where effects are as described above, except for \( e_{ijk} \), which is the pooled error of plot if in year \( k \). The linear mixed models described here were all fitted using ASREML-R (Butler et al. 2007).

The predicted HS-family effects are best linear unbiased predictions (BLUPs) of the transmitting abilities of maternal parents. On the one hand, BLUPs have the property of being shrunk toward their mean (zero) differentially, depending on the relative amount of information available for their computation. As a result, estimates of marker effects based on BLUPs tend to be distorted compared to the estimates based directly on phenotypes (Garrick et al. 2009), which can be problematic in inferential studies such as QTL analyses, especially if reliabilities of BLUPs are highly variable among genotypes. So, in quantitative genetic analyses, it has been recommended to deregress BLUPs for subsequent use in weighted regression models, accounting for differential levels of uncertainty in the deregressed-BLUP estimates, rather than using nonweighted regression models on BLUPs directly. On the other hand,
BLUPs are generally more accurate estimates of the true values, and the approach based on BLUPs does not rely on (possibly suboptimal) weights in regression. Importantly, Guo et al. (2010) showed, in simulation studies, that GS models based on BLUPs predicted true performance of genotypes as, or more accurately than weighted GS models based on “daughter yield deviations”, equivalent to deregressed BLUPs (even when strong differences in available information were simulated), which suggests that BLUPs are acceptable alternatives to their deregressed counterpart as response variables in predictive studies. Consequently, here we chose not to deregress HS-family BLUPs, and use them directly as response variables for training and validating GS models. Nonetheless, all methods used in this study can be adapted to accommodate deregressed BLUPs with differential weights on observations, if needed.

The raw phenotypic data and the matrix of HS-family BLUPs are available online as Supplemental Material, File S1 and File S2, respectively, and from http://dfrc.wisc.edu/sniper/.

**Marker data and quality control**

Exome capture sequencing of HS-family maternal parents was performed using the Roche-Nimblegen EZ Developer libraries using the Roche-Nimblegen probeset ‘120911_Switchgrass_GLBRC_R_EZ_HX1’ as described previously (Evans et al., 2014, 2015). Capture was performed on the 247 individuals from WS4U-C2 and Liberty-C2, and sequencing was performed on the Illumina HiSeq2000 platform, generating 150-nt paired-end reads. Initial quality control was performed using FastQC (v0.10.0; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). PCR primers, adapter sequences, and bases with quality scores below 20 were trimmed using Cutadapt (v1.1; https://code.google.com/p/cutadapt/). Reads with lengths shorter than 35 nt were discarded. Cleaned reads were aligned to the hardmasked *P. virgatum* v1.1 reference genome (http://phytozone.jgi.doe.gov/pz/portal.html#info?alias=Org_Piortatum) using BowTie v0.12.7 (Langmead et al., 2009). Unanchored contigs were assigned to scaffolds (ChrUn1-ChrUn15) for more efficient alignment (Evans et al., 2015). Unique alignments were required, and only a single mismatched nucleotide was permitted in the first 35 bases of the read. Read alignments meeting the alignment criteria were processed using the index, sort, merge (default parameters), and mpileup (-BD –C 0 options) functions of the BAMTools package v0.1.18 (Li et al., 2009). Counts of reads corresponding to reference and alternate alleles were generated for each individual in the WS4U-C2 and Liberty-C2 populations at a subset of sites (2,179,164 loci; HapMap v2) previously determined to be polymorphic using exome capture sequencing data from two diversity panels, the Northern Switchgrass Panel (Evans et al. 2015) and a southern switchgrass panel (C. Brummer, unpublished data).

Then, marker genotypes were called, assuming disomic inheritance of tetraploid switchgrass based on previous genetic-mapping studies (Okada et al. 2010; Li et al. 2014). To infer marker genotypes while effectively accounting for genotype-calling uncertainty, read counts were converted to expected allelic dosages (values between 0 and 2 for the number of copies of the alternate allele); using the algorithm of Martin et al. (2010), fitted on each population separately. The algorithm of Martin et al. (2010) estimates the sequencing/alignment error rate and the population allele frequency for each marker separately, using an Expectation-Maximization (EM) algorithm. Then, for each combination of marker and individual, the posterior probability of each allelic dosage given the read-count data is obtained by Bayes’ rule, assuming Hardy-Weinberg equilibrium (HWE) to derive the prior probability of each allelic dosage, and a binomial distribution of read type (reference/alternate) to derive the likelihood of each allelic dosage. Expected allelic dosages were computed as the sum of possible allelic dosages weighted by their posterior probability for each combination of individual and marker.

In the resulting matrix of expected allelic dosages, marker variables were then filtered for (i) proportion of missing values (strictly lower than 5%); (ii) polymorphism (minor allele frequency across populations strictly greater than 1/2N, and variance higher than 2(1/2N)(1 − 1/2N)/N, with N the total number of genotypes across populations); (iii) HWE within each population (p-value for HWE, based on a χ²-test, strictly higher than 10⁻⁴ for each population considered individually); and (iv) availability of genomic-location information (available information on chromosome and position from the reference genome sequence, and annotation of *P. virgatum* v1.1; DOE-JGI, http://phytozone.jgi.doe.gov/). The resulting matrix M contained expected allelic dosages at q*= 141. 030 selected markers across populations, q*= 108, 077 in WS4U-C2 only and q*= 79, 543 in Liberty-C2 only.

To characterize LD between markers, we used the correlation matrix R consisting of Pearson correlation coefficients between allelic dosages: R_{ij} = Cor(M_i, M_j), where indexing on M refers to columns. For matters of efficiency, the q* × q* R matrix was made block-diagonal, with blocks corresponding to chromosomes (i.e., only local LD was accounted for, through R). This assumed sparsity in marker correlations allowed us to compute R with reasonable costs in time and memory, while potentially reducing noise in the estimations by assuming
zero correlation between markers from different chromosomes. Positive-definiteness of $\mathbf{R}$ (which implies invertibility) was ensured using the modified projection algorithm of Higham (2002) from the nearPD function of the R package Matrix.

Following Speed et al. (2012), for any marker $\mathbf{M}_i$, we define the degree of (local) tagging as the sum of squared correlations involving $\mathbf{M}_i$ and any other marker within the same chromosome, i.e., $\sum_j R_{ij}^2$, with $j$ indexing markers on the same chromosome as $\mathbf{M}_i$’s. This metric is supposed to depict the redundancy in information at $\mathbf{M}_i$, as reflected by $\mathbf{R}$.

The matrix $\mathbf{M}$ for markers selected across both populations is available online as a supplementary file (File S3, with values rounded to the fifth decimal digit) and from http://dfrc.wisc.edu/sniper/ (with non-rounded values).

### Prediction procedures

For each possible combination of population and outcome, we evaluated prediction procedures with respect to four components: (i) population learning scheme—set of parent genotypes to include for training; (ii) environment learning scheme—set of locations to include for training; (iii) marker-data transformation—type of transformation on the marker data used to produce marker features; and (iv) prediction model—method used to generate predictions on the outcome based on marker features. In this study, emphasis was placed on the last two components.

#### Prediction models: The standard statistical model for prediction was genomic BLUP (GBLUP; Habier et al. 2007; Hayes et al. 2009). For a sample of $n$ instances and $q$ marker features, we define GBLUP as follows:

$$g = \mu + Zu + e$$

where $g = \{g_i\}$ is the $n$-vector of HS-family BLUPs; $\mu$ is the $n$-vector of grand mean; $Z$ is the $n \times m$ design matrix attributing the $n$ observations to $m$ parent genotypes; $u \sim \text{Normal}(0, \mathbf{K}\sigma_u^2)$, $\mathbf{K}$ being the $m \times m$ genomic relationship matrix derived from marker features as $\mathbf{K} = \mathbf{X}\mathbf{X}^T$, with $\mathbf{X}$ the $m \times q$ matrix of marker features; $e \sim \text{Normal}(0, \mathbf{I}\sigma_e^2)$, with $\mathbf{I}$ the identity matrix. As explained in the next subsection, the marker features in $\mathbf{X}$ were not the expected allelic dosages, i.e., $\mathbf{X} \neq \mathbf{M}$. The normalizing factor in $\mathbf{K}$ was the sum of sample variances over marker features.

The GBLUP model is equivalent to the RR-BLUP model, where the assumptions of an infinitesimal genetic model are made: effects of marker features are assumed to be additive, linear, heteroscedastic (i.e., having equal variance), and independent (which implies no LD between markers). To accommodate genetic architectures that strongly deviate from the infinitesimal model, we considered eight additional models that were heteroscedastic and/or nonlinear.

Heteroscedastic models were GBLUP-wG, GBLUP-sG, BayesA, and BayesB. The GBLUP-wG model, first used by de los Campos et al. (2013), consisted in weighting marker features by $-\log_2(p)$, where $p$ is the $p$-value for the effect of a marker feature on the outcome of interest. GBLUP-sG is a variation from GBLUP-wG, where are included in the model only the marker features with a false discovery rate (FDR) for their correlation with the outcome lower than some threshold, determined by tuning; the FDR was calculated using the qvalue package in R (Storey and Tibshirani 2003). BayesA and BayesB are Bayesian linear regression models, introduced by Meuwissen et al. (2001), which have the following specification:

$$g = \mu + Xb + e$$

where $g$, $\mu$, $X$ are as described above; $b \sim \text{Normal}(0, \mathbf{I}\sigma_b^2)$. In BayesA, $\sigma_b^2 \sim \chi^2(d_f, S_c^2)$; In BayesB, $\sigma_b^2 = 0$ with probability $\pi$, and $\sigma_b^2 \sim \chi^2(d_f, S_c^2)$ with probability $1 - \pi$; $\pi$ was chosen to follow a Beta(0.2, 1.8) in order to reflect relatively sparse distributions of causal variants across the genome while allowing uncertainty about $\pi$. In both BayesA and BayesB, $S_c^2 = \text{Gamma}(r_b, s_b)$, and $e \sim \text{Normal}(0, \mathbf{I}\sigma_e^2)$, with $\sigma_e^2 \sim \chi^2(d_f, S_c^2)$. The hyperparameters $d_f$, $r_b$, $s_b$, $d_f$, and $S_c^2$ were set through the heuristics described in Pérez and de los Campos (2014), based on a prior estimation of the proportion of variance explained by the model, which was here chosen to be $\sigma_e^2$, from a GBLUP model with an update on marker effects from the heteroscedastic effects model (HEM) of Shen et al. (2013). BayesA and BayesB were fitted by a Gibbs sampling algorithm with 5000 burn-in iterations, then 15,000 iterations for actual sampling of parameter values.

The one nonlinear model that we assayed was the reproducing kernel Hilbert space (RKHS) model described by Gianola and van Kaam (2008). In the implementation recommended by these authors, the RKHS model is made equivalent to the GBLUP model, where pairwise relationship coefficients in $\mathbf{K}$ are replaced by an appropriate nonlinear function of pairwise distances. The pairwise distances were Euclidean distances based on marker features, scaled by the maximum distance over pairs of individuals; the nonlinear function was the Gaussian kernel, with its scale parameter determined by tuning.

To account for both heteroscedasticity and nonlinearity, we extended the RKHS model to RKHS-wG and RKHS-sG, where marker features were weighted, as described above for GBLUP-wG and GBLUP-sG. One last heteroscedastic and nonlinear model that we considered was Random Forest (RF), which is a machine-learning method that combines results from several regression (or classification) trees, fitted to different variations of the data—bootstrap samples of instances and random subsets of features (Breiman 2001). The RF model was fitted with 200 trees, bootstrap samples of size $n$, and subsets of $q/3$ features.

Tuning for the scale parameter in RKHS (and its extensions -wG and -sG), and the FDR threshold in GBLUP-wG or RKHS-sG, was performed through minimization of the generalized-cross-validation criterion (GCV; Golub et al. 1979; Gu and Ma 2005) over a grid of values (strictly greater than 0 and lower than 1, with steps of 0.025 for the scale parameter in RKHS, and steps of 0.05 for the FDR threshold in GBLUP/RKHS-sG). The GCV criterion approximates the leave-one-out cross-validation mean squared error, based on one model fit to the whole training set; it is defined as:

$$GCV = \frac{(1/n)(g - \hat{g})^T (g - \hat{g})}{(1 - (1/n)tr(\mathbf{H}))^2},$$

where $tr$ refers to the trace (the sum of diagonal elements of a matrix); $\hat{g}$, the linear prediction of $g$, and $\mathbf{H}$, the “hat” (smoothing) matrix, such that $g = \mathbf{H}g$, depend on the parameter under tuning.

The GBLUP and RKHS models, as well as their extensions -wG and -sG, were fitted using the R package rrBLUP (Endelman 2011); the BayesA and BayesB models were fitted using the R package BGLR (Pérez and de los Campos 2014); the HEM of Shen et al. (2013) was fitted using the R package bigRR (Shen et al. 2013), and the RF model was fitted using the R package randomForest (Liaw and Wiener 2002).

### Marker-data transformations: As mentioned above, the input $X$ to prediction models were transformations of the marker-data matrix
For a given set of individuals, consisting of either WS4U-C2, Liberty-C2, or both populations combined, the following transformations of M were made: (i) Base, where features are centered allelic dosages, and correspond to the typical input to GS models: $X_{\text{Base}} = M - P$, with P the $m \times q$ matrix with uniform columns containing the mean allelic dosages within the population, $q = q^s$; (ii) PCA, where features are uncorrelated synthetic variables each contributed based on their relative degree of tagging (the more redundant information at a given marker, the lower its weight): $Q = \frac{1}{\sqrt{q^s}} \Phi^T \Phi^{-1}$, where $\Phi$ is the eigendecomposition of $\Sigma_X$; (iii) Cor, where features are marker variables scaled through a correlation matrix: $X_{\text{Cor}} = X_{\text{Base}} L$, with L the $q^s \times q^s$ lower-triangular matrix from the Cholesky decomposition of $\Theta = R^{-1}$, such that $LL^T = \Theta$, $R$ being the matrix of correlation between marker variables (as previously described, $q = q^s$); (iv) LD, where features are marker variables weighted based on their relative degree of tagging (the more redundant information at a given marker, the lower its weight): $X_{\text{LD}} = X_{\text{Base}} W^{1/2}$, with W the diagonal matrix of weights supposed to adjust for redundancy in marker information due to LD; $W = \text{diag}(w)$, and w was the least-absolute-error solution to $(R \# R)w = 1_q^s$, subject to $w_j \geq 0$, $j = 1, \ldots, q^s$, with $R \# R$ the matrix of squared correlation between marker variables and $1_q^s$ the $q^s$-vector of one values, $q = 16,555$ in WS4U-C2, and $q = 10,615$ in Liberty-C2 ($q < q^s$, as a result of some weights being exactly zero). The publicly available LDAK software was used to calculate w (http://dougspeed.com/lidak/; Speed et al. 2012).

Marker-data transformations were chosen so that they have well-defined equivalencies in the GBLUP and RKHS models. In a GBLUP model, PCA is exactly equivalent to Base (Figure S1): $(X_{\text{Base}} V)(X_{\text{Base}} V)^T = \Sigma_V V^T V^T V = \Sigma_V$, since $V^TV = I$, and $V^TV = I$, and $\Sigma_V$ is the eigen-decomposition of $X_{\text{Base}} X_{\text{Base}}^T$. Cor is equivalent to Base when local LD is accounted for through $\Theta$: $(X_{\text{Base}} L)(X_{\text{Base}} L)^T = X_{\text{Base}} L L^T X_{\text{Base}}^T = X_{\text{Base}} \Theta X_{\text{Base}}^T$; LD is equivalent to Base when local LD is accounted for by weights on marker features as in Speed et al. (2012): $(X_{\text{Base}}W^{1/2})(X_{\text{Base}}W^{1/2})^T = X_{\text{Base}} W X_{\text{Base}}^T$. The equivalences mentioned for cross-products in GBLUP also apply to Euclidean distances in RKHS. Interestingly, in a RR-GLM model (equivalent to GBLUP), Cor and LD correspond to Base when marker-feature effects are assumed to follow a Normal$(0, \Theta \sigma_2^2)$, and a Normal$(0, W \sigma_2^2)$, respectively, instead of a Normal$(0, I \sigma_2^2)$ ($\sigma_2^2$ is the variance of marker effects). For the more complex models accounting for heteroscedasticity (BayesA, BayesB, GBLUP-wG, GBLUP-sG, RKHS-wG, RKHS-sG, and RF), the transformed marker variables were considered features in their own right, whose contribution to a given outcome of interest may be weighted similarly to features in Base.

Throughout the article, prediction procedures are referred to by a combination of marker-data transformation and prediction model (e.g., Cor – RKHS-sG) for a given learning scheme (involving the grouping of populations and environments in a training set; see next subsections).

Matrices X for WS4U-C2 and Liberty-C2 (with transformations Base, PCA, Cor, and LD), and the corresponding relationship matrices and Euclidean distance matrices, are available online from http://dfrc.wisc.edu/sniper/, in .rds format readable in R.

Population learning schemes: Given a target population, we considered two types of training sets with regard to parent genotypes. The HS-family BLUPs used for training prediction models could be either from the target population only (within-population learning), or from both populations pooled together (across-population learning).

Environment learning schemes: Given a target outcome, we considered two types of training sets with regard to observations at each parent genotype. For example, with DMY in WI as the target outcome, the HS-family BLUPs used for training the prediction model could be either those from the target location only (within-environment learning; e.g., data on DMY in WI only), $n = m$, or from both locations considered jointly for the same trait (across-environment learning; e.g., data on DMY in WI and NE), $n = 2m$. In across-environment learning, whenever leaving out HS families from the dataset for validation (see next section), the data on the same HS families in both environments were used for testing.

Validation of prediction procedures: Prediction procedures were evaluated using prediction accuracy estimated in five-fold cross-validation. Given a random partition of
instances in five subsets of similar size, four subsets were used for training, and the remaining subset was used for testing. For each of the five subsets used sequentially for testing, prediction accuracy was computed as the Pearson coefficient of correlation between “observed” and predicted HS-family BLUPs. The significance of the difference in prediction accuracy between a given procedure and a standard procedure was assessed in replicated cross-validation by two-sided paired Dunnett tests, which are $t$-tests modified to account for multiple comparisons to a single control (Dunnett 1964). For each outcome and population, the standard (control) procedure was chosen to be Base – GBLUP with within-population and within-environment learning. In cross-validation, the overlap between training sets results in lower variability among estimates of prediction accuracy, compared to the hypothetical case where training sets are generated independently. So, in paired Dunnett tests, the $t$-statistic $T$ was adjusted to account for correlation among computed prediction accuracies, as described in Bouckaert and Frank (2004):

$$T = \frac{\bar{\Delta}}{SD(\Delta) \sqrt{1/K} + 1/K - 1},$$

where $K = 5$ is the number of “folds” in cross-validation, and $R = 10$ is the number of replications (five-fold cross-validation was repeated 10 times); $\Delta = \bar{z}(c_t) - z(c_0)$, with $c_t$ and $c_0$ the KR-vectors of prediction accuracies from the test procedure and the standard procedure, respectively, and $z$ the Fisher transformation (for normality of prediction accuracies); $\bar{\Delta}$ and $SD(\Delta)$ are the mean and SD of $\Delta$, respectively. The R package nCDunnett was used to obtain adjusted p-values for $T$.

In order to limit the number of possible combinations to assess, prediction procedures were first optimized with respect to learning

Figure 2 Distribution of degree of tagging in (A) WS4U-C2 and (B) Liberty-C2. Relationship between minor allele frequency (MAF) and degree of tagging in (C) WS4U-C2 and (D) Liberty-C2; the blue curve corresponds to the mean value (and its 95%-confidence interval) from a cubic-regression spline model assuming a Normal distribution for the degree of tagging. Cubic-regression spline models were fitted using the R package mgcv (Wood 2006).
schemes only, based on replicated cross-validation, and then optimized with respect to marker-data transformation and prediction model in a two-step process (intended to reduce computational burden in the optimization): for each outcome and population, the combination of marker-data transformation and prediction model with the highest prediction accuracy based on nonreplicated cross-validation was selected; then the selected procedure was compared to the standard procedure (Base – GBLUP) in replicated cross-validation. If the selected procedure differed from Base – GBLUP by both an alternate marker-data transformation and an alternate prediction model (e.g., Cor – BayesA), the alternate transformation was first compared to Base in a GBLUP model to assess the usefulness of transforming the marker data (e.g., Cor – GBLUP vs. Base – GBLUP). Then the alternate model was compared to GBLUP using the alternate transformation for both models to assess the benefit of a more complex prediction model (e.g., Cor – BayesA vs. Cor – GBLUP). Dunnett tests, with adjustments for the number of marker-data transformations or prediction models, were used in these comparisons in order to account for selection bias, i.e., the fact that the same data were used for both choosing the selected procedure and then comparing it with Base – GBLUP.

Genetic analysis of phenotypic traits

Genomic correlation: To characterize the correlation between genomic effects (u) at different outcomes, a multivariate GBLUP model was fitted on any pair of outcomes as in Burgueño et al. (2012); errors e were assumed independent within outcomes; the genomic relationship matrix was K × X_{base} X_{base}^{-T}; the covariances by outcome of genomic effects and errors were estimated by REML. The R package ASREML-R was used to fit the multivariate GBLUP models.

Association mapping: In order to assess the plausibility of heteroscedastic models, we conducted genome-wide association studies (GWAS) on all outcomes for both populations combined. For each marker in M having a minor allele frequency (MAF) higher than 0.05, the EMMAX linear mixed model of Kang et al. (2010), in which relatedness was accounted for through K × X_{base} X_{base}^{-T}, was fitted using the R package rrBLUP. The threshold used to declare significance of associations was a FDR (as from Storey and Tibshirani 2003) lower than 0.05. Significant markers were then selected altogether in one linear mixed model, with relatedness accounted for through K and fixed effects for markers, using a forward stepwise selection procedure based on the Bayesian information criterion. No covariate for population structure was included in the GWAS models, since the incentive for performing GWAS here was to investigate whether the ability of GBLUP to capture variation at outcomes could be significantly improved by including fixed marker effects. Even though between-population variability was adjusted for—because HS-family BLUPs were computed within each population separately—and genetic relatedness was captured through K in GWAS models, it cannot be ruled out that significant markers actually reflected some population structure.

Partition of genomic heritability: The genomic heritability is defined here as the proportion of variance explained by a GBLUP model, i.e., $\frac{\sigma^2_u}{\sigma^2_u + \sigma^2_e}$. To assess the relative contribution of markers with different degrees of tagging to the genomic heritability, a multiple-component GBLUP model was fitted on each outcome, as in Yang et al. (2011), with three nonoverlapping marker classes: genomic relationship matrices K_1, K_2, and K_3 were calculated from the distinct marker classes in X_{base}, and the associated variances $\sigma^2_{u_1}$, $\sigma^2_{u_2}$, and $\sigma^2_{u_3}$ were estimated by REML. The contribution of class j to the genomic heritability was defined as $\frac{\sigma^2_{u_j}}{\sum_j \sigma^2_{u_j} + \sigma^2_e}$. Marker classes were determined from tertiles on the markers’ degree of tagging. The R package ASREML-R was used to fit the multiple-component GBLUP models.

Data availability
The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

RESULTS

Genomic structure and relatedness in populations
The two populations considered, WS4U-C2 and Liberty-C2, had strong differences in their patterns of decay in local (within-chromosome) LD. WS4U-C2 had a rapid decline in local LD, with the squared correlation between expected gametic phases ($r^2$) being essentially zero for physical distances between markers above 1 Mb (Figure 1A). Conversely, in Liberty-C2, values of $r^2$ decayed more slowly (Figure 1B), in accordance with the fact that effective population size in this population, derived from a cross between two cultivars, is certainly lower than in WS4U-C2, derived from a diverse collection of 162 upland-ecotype plants. The concordance in LD from WS4U-C2 to Liberty-C2 was very low, as reflected by the low adjusted $R^2$ in a nonlinear regression for (y = r in Liberty-C2; x = r in WS4U-C2) on r in WS4U-C2. This low concordance seemed to be due to the rapid LD decay in WS4U-C2, with many values of r being close to zero in WS4U-C2 only (Figure 1C). However, when values of r departed from zero in both populations, there seemed to be some consistency in LD phase, as reflected by the sign of r being relatively consistent from one population to another. The distribution of minor allele frequency (MAF), over all 141,030 markers selected, differed by population. In WS4U-C2, as MAF decreased, the cognate number of markers increased exponentially (Figure S2A). In Liberty-C2, while a high proportion of markers appeared to be fixed, or singletons (likely due to de novo mutations, or possibly genotype miscalls), the remaining markers showed a relatively uniform distribution in their MAF (Figure S2B). The concordance in MAF from
WS4U-C2 to Liberty-C2 was low (adjusted $R^2 = 0.24$) but significant ($p < 0.0001$), with many markers being fixed in one population but not the other (Figure S2C).

Figure 2, A and B, shows the distribution in the degree of tagging, reflecting the duplication of marker signals along a given chromosome, in each population. The relative variability and skewness in degree of tagging were larger in WS4U-C2 than in Liberty-C2 (respectively: co-efficient of variation of 0.876 and 0.591; skewness of 3.99 and 1.31). In Liberty-C2, the relationship between MAF and degree of tagging was typical, with rare variants generally having a low degree of tagging (Figure 2D). Conversely, in WS4U-C2, degree of tagging was, on average, much higher for rare variants (MAF < 0.10) than for more common variants (MAF > 0.10) (Figure 2C). The high degree of tagging for rare variants in WS4U-C2 was due to extended local LD, observed in the LD-decay plots as lines of points for WS4U-C2 (Figure 1A) but not for Liberty-C2 (Figure 1B). We hypothesize that such regions of extended LD in WS4U-C2 are due to the presence, in WS4U (the original collection of 162 plants), of haplotypes consisting of rare alleles, and limited recombination during the two cycles of selection leading to WS4U-C2.

Genetic relationships, defined here as twice the coefficient of identity by descent (IBD) between genotypes, was generally lower within WS4U-C2 than within Liberty-C2, in accordance with the presumably higher effective population size in WS4U-C2 (Figure 3). Interestingly, the distribution in genetic relationships within Liberty-C2 was bimodal, with the first peak close to a HS relationship of 1/4, and the second peak close to a full-sib relationship of 1/2. This genetic structure within Liberty-C2 could have been generated by preferential mating of plants from the same ecotype (having similar flowering times), i.e., assortative mating during crosses between the upland and lowland cultivars. Relationships across populations, inferred to be exactly zero, indicated strong genetic dissemblance between WS4U-C2 and Liberty-C2. This dissemblance, along with the low consistency in LD and MAF across populations, suggests little benefit from pooling populations into one single training set for genomic prediction, as was observed in previous studies (e.g., Karoui et al. 2012; de los Campos et al. 2013).

Genetic variability and architecture of phenotypic traits

The outcomes considered for genomic prediction, described in Table 1, consisted of combinations of location and phenotypic trait, measured in 2 or 3 yr. Mean reliabilities (i.e., the average inferred squared correlation between true HS-family effects $g$’s and their BLUPs) were high ($\approx 0.61$) for PH and HD in both locations. Reliabilities for DMY were relatively high in NE (0.45 in WS4U-C2, and 0.53 in Liberty-C2), but low in WI (0.10 in WS4U-C2, and 0.21 in Liberty-C2).

Genomic correlations between outcomes of the same trait in different environments (based on multivariate GBLUP models) were positive, strong ($\approx 0.81$), and significant ($p \leq 0.05$), with PH and HD (Figure S3), suggesting some benefit from pooling data from different locations into one single training set, for genomic prediction on these two traits. Conversely, genomic correlations between outcomes were not significant ($p > 0.05$) than for more common, and therefore the corresponding genetic architecture presumably consisted of rare and/or small-effect causal variants.

For a given population, the relative contribution of different marker classes to the genomic variance at a given outcome was characterized through the partition of genomic heritability, from a multiple-component GBLUP model. The classes of markers were determined based on the degree of tagging. They consisted of markers with a degree of tagging below the 1/3-quantile, “Intermediate”: markers with a degree of tagging between the 1/3- and 2/3-quantiles; “High”: markers with a degree of tagging higher than the 2/3-quantile. For a given outcome and a given population, the height of the bar corresponding to class $j$ indicates its contribution to the total genomic heritability, i.e.,

$$
\sum_{j} \sigma^2_{gj} + \sigma^2_c \text{ estimated in a multiple-component GBLUP model, with X_{Base} as input, using ASREML-R (Butler et al. 2007).}
$$

Selection and validation of prediction procedures

Here, four components of prediction procedures for GS, on only one outcome at a time for each population, were assayed: (i) population learning scheme—training set consisting of data on either the target population only, or both populations; (ii) environment learning scheme—training set consisting of data on either the target trait-location only, or the same trait in both locations; (iii) marker-data
transformation—type of marker features used as input to prediction models; and (iv) prediction model—assumed relationship between the marker features and the outcome.

**Learning schemes:** The importance of the first two components of prediction procedures (learning schemes) was assessed while using a GBLUP model with Base marker-data transformation (Base = GBLUP, i.e., using centered expected allelic dosages as input to the standard GBLUP model (Table 2 for DMY and Table S2 for other traits). Across all outcomes, there was no observed superiority of one particular learning scheme, based on mean prediction accuracies in five-fold cross-validation replicated 10 times. Also, none of the observed increases in prediction accuracy, compared to the within-population/within-environment scheme, were deemed significant (p > 0.10). Pooling data across populations was not beneficial, as could be expected from the low concordance in LD patterns (Figure 1C) and MAF (Figure S2C) between WS4U-C2 and Liberty-C2, as well as from the genetic dissemblance between the two populations (Figure 3). However, no significant increase in prediction accuracy was obtained with PH and HD from pooling data across environments in spite of the strong positive genomic correlations inferred for these two traits (Figure S3).

**Marker-data transformations and prediction models:** As described above, pooling data across populations and/or environments did not offer strong opportunities for improving prediction procedures. Thus, the importance of marker-data transformations and prediction models was investigated while using a simple within-population/within-environment learning scheme. Assessment of prediction procedures was performed in two steps. In step 1, a candidate prediction procedure (a combination of marker-data transformation and prediction model) was selected based on mean prediction accuracy in nonreplicated five-fold cross-validation (Table 3 for DMY, Table S3 for other traits, and Figure S4). In step 2, the selected candidate procedure was then compared to the standard—a GBLUP model on centered expected allelic dosages (Base = GBLUP)—in five-fold cross-validation replicated 10 times (Figure 5 for DMY, Figure S5 for other traits) so as to assess the benefit from the alternate marker-data transformation, and/or the more complex prediction model selected.

Over cases (combinations of outcome and population), average prediction accuracies across prediction procedures in nonreplicated cross-validation ranged from −0.012 (DMY in WI for Liberty-C2) to 0.545 (HD in WI for Liberty-C2), with generally higher accuracies with PH and HD than with DMY (Table 3, Table S3, and Figure S4). However, high accuracies were achieved with DMY in NE for Liberty-C2 (0.455 on average). With DMY in WI (for both populations), DMY in NE (for WS4U-C2), and PH in NE (for WS4U-C2), marker-data transformation seemed to offer more opportunities for improvement than prediction models. In other cases, prediction accuracies were generally not sensitive to marker-data transformations, except for PCA, with which heteroscedastic models often performed poorly. In general, across cases, the Base marker-data transformation was selected in only three out of 12 cases, which again indicates some potential benefit from accounting for LD through marker-data transformation. Heteroscedastic models were selected nine times out of 12, with strong increases observed with DMY in WI, as could be expected from the many GWAS signals detected with that trait (Table S1).

The statistical significance of the difference in prediction accuracies between Base = GBLUP and the candidate prediction procedure, selected based on nonreplicated cross-validation, was assessed in replicated cross-validation by paired Dunnett tests. Dunnett tests account for biases on significance due to multiple testing, which arose here from deriving a hypothesis from the same dataset as that used for testing (selection bias). Dunnett tests were further adjusted to account for the overlap between training sets in cross-validation, which caused the SD in prediction accuracy among “folds” to be an underestimate of the general SD (among hypothetical datasets). With DMY and PH in NE for WS4U-C2, and with DMY (in both locations) for Liberty-C2, the selected prediction procedures yielded higher mean prediction accuracies in replicated cross-validation. However, the observed differences were generally not deemed significant according to our tests, with the exception of DMY in NE for WS4U-C2 (Figure 5 and Figure S5). With DMY and PH in NE for WS4U-C2, the marker-data transformation apparently contributed more than the prediction model to the increase in prediction accuracy. Interestingly, for those cases, markers with a high degree of tagging tended to capture very little of the genomic heritability (Figure 4). The most notable case was DMY in NE for WS4U-C2, where a highly significant increase in accuracy (p < 0.01) was observed when comparing Base = GBLUP to Cor = GBLUP, but using a BayesA model (the selected alternate model) rather than a GBLUP model (the standard model) did not yield a significant increase in prediction accuracy (Figure 5). This one strong increase indicates that, with our data, some substantial and reliable increase in prediction accuracy could be achieved through marker-data transformation, and not so much by prediction models more complex than GBLUP. Accordingly, when going through the same process of selection and validation of prediction procedures while considering only Base for input to the various prediction models (which is a traditional protocol for prediction-procedure optimization in GS), no significant increase in prediction accuracy could be obtained in any case (Figure S6).

**DISCUSSION**

This study dealt with the optimization of prediction procedures in GS through four components: population learning scheme, environment learning scheme, marker-data transformation, and prediction model, with emphasis placed on marker-data transformations and prediction models. The distinction between marker-data transformations and
prediction models may seem arbitrary, especially for procedures like Cor — GBLUP or Cor — RKHS, which are equivalent to well-defined statistical models (see Material and Methods). However, ‘marker-data transformations’ and ‘prediction models’ account for different characteristics of marker loci, with marker-data transformations accounting for redundancy in marker information, and prediction models all assuming independence of marker-feature effects, and possibly accounting for heteroscedasticity and/or nonlinearity of marker-feature effects.

In GS studies, the type of prediction models has been by far the factor upon which optimization of genomic prediction procedures has most often acted (e.g., Moser et al. 2009; Crossa et al. 2010; Heslot et al. 2012). In this study, we demonstrate that applying linear transformations to the marker data to account for local LD among markers may be useful to achieve gains in prediction accuracy that are reliable based on replicated cross-validation and honest significance tests. In fact, the only highly significant increase in prediction accuracy that we achieved was due to Cor (the account of local LD among marker loci through a block-diagonal correlation matrix), with DMY in NE for WS4U-C2 (Figure 5). Some substantial increase in prediction accuracy could also be achieved by Cor with PH in NE for WS4U-C2, and PCA with DMY in WI for Liberty-C2, though the differences observed were not deemed significant ($p = 0.37$ and $p = 0.96$, respectively; Figure S5 and Figure 5). While optimizing prediction procedures through learning schemes has been useful in some studies (Rincent et al. 2012; Heslot et al. 2013), here they did not offer strong opportunities for improving prediction procedures. With PH and HD, for which genomic correlations between environments were high (Figure S3), the absence of consistent and significant increase in prediction accuracy from pooling environments may have been due to the high reliability of HS-family BLUPs for those two traits, causing additional measurements on one given genotype to contribute little to the quality of the signal in the data. Pooling data from different populations was not useful due to the strong genetic dissimilarities between WS4U-C2 and Liberty-C2 (Figure 1, Figure S2, and Figure 3), certainly owing to the differences in effective population sizes, as well as to local adaptation of switchgrass populations and the ancient divergence between upland and lowland ecotypes (Zhang et al. 2011).

In our study, both populations showed large variability in the degree of (local) tagging at marker loci (Figure 2, A and B), probably (to some extent) because markers were derived from exome capture, which target specific regions in the genome for sequencing. In WS4U-C2 specifically, there was also rapid LD decay along chromosomes with nevertheless extended LD (Figure 1A), presumably caused by sampling artifacts in WS4U, resulting in outstandingly high values of degree of tagging (Figure 2A). Also, variants with extreme degrees of tagging in that population tended to be rare (Figure 2C). Conversely, in Liberty-C2, LD decay within a chromosome was slower, and there was no extended LD (Figure 1B), resulting in relatively little skewness in the distribution of degree of tagging in that population (Figure 2B). Also, the relationship between MAF and degree of tagging was quite typical (Figure 2D), similar to that reported in human by Speed et al. (2012). Probably as a consequence of the differences in genomic structure across populations, the contribution of strongly tagged variants to the genomic heritability of outcomes was generally lower in WS4U-C2 than in Liberty-C2. The only cases in which there were substantial increases in prediction accuracy from marker-data transformations and prediction models, relatively to Base — GBLUP, were DMY and PH in NE for WS4U-C2, and DMY in WI for Liberty-C2: +75% from Cor-BayesA, +20% from Cor-BayesB, and +75% from PCA-RF, respectively (Figure 5). Quite remarkably, these were also the only cases in which proportions of genomic heritability explained by strongly tagged variants were very low (0%, 11%, and 0% of genomic heritability explained, in these three cases, respectively) (Figure 4). Importantly,

<table>
<thead>
<tr>
<th></th>
<th>DMY in WI</th>
<th>DMY in NE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GBLUP</td>
<td>GBLUP-wG</td>
</tr>
<tr>
<td>WS4U-C2 Base</td>
<td>0.151</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>0.151</td>
<td>-0.045</td>
</tr>
<tr>
<td></td>
<td>0.170</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>0.105</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.144</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liberty-C2 Base</td>
<td>0.016</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>0.016</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>0.004</td>
<td>-0.054</td>
</tr>
<tr>
<td></td>
<td>0.034</td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.018</td>
</tr>
</tbody>
</table>

Prediction accuracies were estimated with a within-population/within-environment learning scheme in five-fold cross-validation, with no replication. For a given population and outcome (trait-location combination), the highest average value across marker-data transformations is underlined; the highest value across prediction models all as-
previous simulation studies have investigated the effect of marker-data transformation LD on the accuracy of estimations (Speed et al. 2012) and predictions (Nishio and Satoh 2015) in genomic studies, and they have shown that the lower the relative degree of tagging at the causal variants for a given outcome, the more beneficial LD tends to be, compared to Base. Therefore, here we argue that cases where marker-data transformations might be useful are those where a small proportion of genomic heritability is captured by strongly tagged markers, which presumably derives from the fact that the causal variants are not strongly tagged. However, we are assuming here that Cor and PCA respond to features of genomic structure similarly to LD; simulation studies would be necessary to support such an assumption.

In Liberty-C2, there seemed to be some population structure, presumably caused by assortative mating, while there was little evidence for such structure in WS4U-C2 (Figure 3). Population structure may cause marker loci from different chromosomes to be correlated. Therefore, in Liberty-C2 particularly, with marker-data transformations Cor and LD, it might have been useful to account for global LD, i.e., correlations among marker loci within and across chromosomes. Interestingly, PCA does account for global LD in heteroscedastic models (in GBLUP and RKHS, the homoscedastic models, PCA is equivalent to Base), and this transformation was useful only in Liberty-C2, in one case (DMY in WI). However, despite being substantial, the increase in mean prediction accuracy in this case was not deemed significant in paired

Figure 5 Validation of selected prediction procedures for DMY in WI and NE. Prediction accuracies \( r_{g}^2 \) were estimated with a within-population/within-environment learning scheme in five-fold cross-validation, replicated 10 times. In each boxplot, up to two comparisons are made: (i) the candidate-transformation procedure (selected marker-data transformation according to nonreplicated five-fold cross-validation in a GBLUP model; Table 3) is compared to the standard procedure (Base — GBLUP), if relevant; and (ii) the candidate procedure (selected prediction procedure according to nonreplicated five-fold cross-validation; Table 3) is compared to the candidate-transformation procedure. The significance of differences in prediction accuracies was assessed by two-sided paired Dunnett tests, which accounted for multiple testing of data transformations, in (i), and of prediction models, in (ii). The t-statistics in Dunnett tests were adjusted to account for correlation among training sets in cross-validation, as described in Bouckaert and Frank (2004).
comparisons (Figure 5). Here, we did not account for global LD in Cor and LD, since the transformations would then have involved correlation matrices that are too large to process, or even store. Given the genomic features of Liberty-C2, pruning markers—which should not be too detrimental to prediction accuracy given the relatively large LD extent (Figure 1B)—and then accounting for global LD, might have proved useful with Cor and LD.

In general, the account of LD through marker-data transformations based on correlation matrices (Cor and LD) could be further improved by reducing the level of noise in estimates of marker correlations. Preprocessing correlation matrices for shrinkage and/or sparsity could then prove beneficial, with Gaussian graphical models (e.g., the graphical LASSO; Friedman et al. 2008) or generalized thresholding methods (e.g., the MCP method; Zhang 2010) being potentially useful tools. Furthermore, basing correlations on expected gametic phases (i.e., using haplotypic rather than genotypic correlations) would allow a more appropriate account of LD. However, in Cor (whenever preprocessing correlations or using expected gametic phases), one would then face the very serious computational challenge of ensuring that the resulting correlation matrix is positive definite (i.e., being a proper and invertible correlation matrix). Filtering out markers based on relatively stringent MAF thresholds could result in higher prediction accuracies if the outcome is affected mostly by common causal variants, but it could also increase the benefit from marker-data transformations such as Cor or LD, because correlations would probably be less prone to error if only common markers are considered—some correlations involving rare variants, estimated at low but nonzero values, might simply be spurious rather than effectively due to LD, and would then contribute to overall noise in the marker correlation matrix. However, preselection based on MAF would remove potentially useful information whereas the linear mixed models assayed here (i.e., all prediction models except RF) have the (supposedly desirable) property of downweighing the effect of markers with low variance (and therefore low MAF) as long as marker variables are not standardized (i.e., scaled through a covariance matrix). Besides, optimizing prediction procedures with respect to a MAF threshold would result in more intricate procedures: if optimization for MAF threshold were to be integrated into a prediction procedure (i.e., tuning for MAF threshold within the procedure), the resulting computational complexity would greatly increase, since tuning would then have to be performed within each "fold" during cross-validation for each outcome and population; if different MAF thresholds were to be considered for different procedures (i.e., the MAF threshold would be an additional component of prediction procedures), then the computational burden in the study would increase, but, more importantly, multiple testing would become a greater issue (accounting for the high number of candidates in comparisons of prediction procedures would cause significance tests to be highly conservative). In this study, we considered marker-data transformations that were relatively simple, but some of the treatments suggested above (regularization on correlation matrices, use of expected gametic phases for estimating correlations, and/or preselection based on MAF) may prove useful to better account for LD in GS.

Our results are based on two populations of switchgrass with three traits. The conclusions that we drew about the relative importance of prediction-procedure components are, of course, not generalizable to all GS contexts and genetic architectures. The limited number of genotypes (137 and 110 in WS4U-C2 and Liberty-C2, respectively) likely favored the most parsimonious (statistically efficient) models, regardless of traits’ genetic architecture. With larger sample sizes, the higher flexibility of more complex models may have been more beneficial with traits whose genetic architecture substantially deviates from the infinitesimal model. Also, not all genetic architectures were represented in our data. In particular, there was little apparent benefit from RKHS (the only nonlinear homoscedastic model) on prediction accuracy. This may indicate that the traits considered in our data are influenced mostly by additive effects, but the small sample sizes arguably limit such conclusions. Finally, assessments of GS through cross-validations is limited in that it does not test prediction procedures for persistency of accuracy over generations, a criterion by which prediction models can differ significantly, as was shown through simulations by Habier et al. (2007). It would be important to formally study the repercussion of accounting for LD in GS on the persistence of accuracy over generations, using simulation and/or empirical studies. Quite interestingly, Nishio and Satoh (2015) suggested that accounting for LD through the LD marker-data transformation would be beneficial in long-term GS when causal variants are unevenly tagged, because it would prevent strongly tagged causal variants from being quickly fixed relatively to weakly tagged ones, and therefore would result in genomic predictions from LD — GBLUP having accuracies that are more slowly deteriorated over generations, compared to predictions from Base — GBLUP.

We believe the relatively high prediction accuracies, particularly with DMY in NE, should motivate the implementation of GS breeding programs in switchgrass. Nevertheless, we may conduct future studies to compare GS with phenotypic selection for realized genetic gains, in programs that are run in similar conditions, so as to bring further evidence for the usefulness of GS technologies in perennial grass breeding.

ACKNOWLEDGMENTS

The authors thank two anonymous reviewers for remarks and suggestions that greatly helped with improving the manuscript. We are grateful to Jeremy Schnutz of the Department of Energy Joint Genome Institute and Hudson Alpha for his work on the switchgrass genome, and to Nick Baker and Joseph Halinar, USDA-ARS, Madison, WI, and Steve Masterson, USDA-ARS, Lincoln, NE, for assistance with field operations and data collection. This research was funded in part by the following agencies and organizations: the US Department of Energy Great Lakes Bioenergy Research Center, Department of the Environment (DOE) Office of Science BER DE-FC02- 07ER64494 (laboratory operations, genotyping, and bioinformatics), the US Department of Energy Joint Genome Institute, supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231 (sequencing), Agriculture and Food Research Initiative Competitive Grant No. 2011-68005-30411 from the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (GenUSA; field operations and phenotypic measurements), USDA-ARS Congressionally allocated funds (field operations, technical support, and logistics), and the University of Wisconsin Agricultural Research Stations (field operations). Mention of commercial products and organizations in this manuscript is solely to provide specific information. The USDA is an equal opportunity provider and employer. G.P.R. was supported by the Gabelman-Shippo Wisconsin Distinguished Graduate Fellowship at the University of Wisconsin–Madison. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

LITERATURE CITED

Erbe, M., B. Hayes, I. Matukumalli, S. Goswami, P. Bowman et al., 2012 Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95(7): 4114–4129.
Perlack, R. D., L. L. Wright, A. F. Turhollow, R. L. Graham, B. J. Stokes et al., 2005 Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply, Oak Ridge National Laboratory, Oak Ridge, TN.

Communicating editor: E. Huang
**Table S1 – Results of association mapping for DMY in WI across both populations**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Feature</th>
<th>Chromosome</th>
<th>Position (bp)</th>
<th>MAF</th>
<th>R²LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pavir.Ba03876</td>
<td>intronic</td>
<td>Chr02a</td>
<td>77,527,751</td>
<td>0.097</td>
<td></td>
</tr>
<tr>
<td>Pavir.Ga00641</td>
<td>exonic</td>
<td>Chr07a</td>
<td>8,145,980</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>Pavir.Ha00300</td>
<td>intronic</td>
<td>Chr08a</td>
<td>6,097,945</td>
<td>0.091</td>
<td></td>
</tr>
<tr>
<td>Pavir.Ia03558</td>
<td>intronic</td>
<td>Chr09a</td>
<td>69,972,787</td>
<td>0.081</td>
<td>0.42</td>
</tr>
<tr>
<td>Pavir.Bb00675</td>
<td>exonic</td>
<td>Chr02b</td>
<td>41,239,077</td>
<td>0.139</td>
<td></td>
</tr>
<tr>
<td>Pavir.Hb00247</td>
<td>exonic</td>
<td>Chr08b</td>
<td>4,667,943</td>
<td>0.053</td>
<td></td>
</tr>
</tbody>
</table>

Gene names and marker localizations are based on the reference genome sequencing and annotation of Panicum virgatum v1.1 (DOE-JGI, http://phytozome.jgi.doe.gov/). R²LR: likelihood-ratio-based R² statistic (Sun et al. 2010). Association mapping was performed using a linear mixed model with the outcome being DMY, PH or HD in WI or NE, with one fixed additive effect for a given marker assayed and a random background effect accounted for through a realized genetic relationship matrix derived from marker data as K = X_baseX_base^T. The threshold used to declare significance of associations was a false discovery rate (as from Storey and Tibshirani 2003) lower than 0.05. The significant markers reported above were selected altogether in one linear mixed model, with a random background effect and fixed effects for markers, by forward stepwise selection based on the Bayesian information criterion. Significant associations were obtained only with DMY in WI (not with the eleven other outcomes).
Table S2 - Mean prediction accuracy across population and environment learning schemes for PH and HD in WI and NE

<table>
<thead>
<tr>
<th></th>
<th>PH in WI</th>
<th>PH in NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>WS4U-C2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within pop.</td>
<td>0.254</td>
<td>0.253</td>
</tr>
<tr>
<td>Across pop.</td>
<td>0.235</td>
<td>0.247</td>
</tr>
<tr>
<td>Liberty-C2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within pop.</td>
<td>0.329</td>
<td>0.364</td>
</tr>
<tr>
<td>Across pop.</td>
<td>0.334</td>
<td>0.367</td>
</tr>
<tr>
<td>HD in WI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WS4U-C2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within pop.</td>
<td>0.312</td>
<td>0.314</td>
</tr>
<tr>
<td>Across pop.</td>
<td>0.291</td>
<td>0.300</td>
</tr>
<tr>
<td>Liberty-C2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within pop.</td>
<td>0.530</td>
<td>0.519</td>
</tr>
<tr>
<td>Across pop.</td>
<td>0.506</td>
<td>0.478</td>
</tr>
</tbody>
</table>

Prediction accuracies were estimated with Base – GBLUP in five-fold cross-validation replicated ten times. The significance of differences in prediction accuracy was assessed by two-sided paired Dunnett tests, which accounted for multiple comparisons of learning schemes to a single reference (the within-population/within-environment scheme). The t-statistics in Dunnett tests were adjusted to account for correlation among training sets in cross-validation, as described in Bouckaert and Frank (2004). For a given population and trait-location combination, differences in prediction accuracy compared to the within-population/within-environment scheme were never deemed significant (p > 0.10 in paired Dunnett tests).
### Table S3 - Mean prediction accuracy across marker-data transformations and statistical models for PH and HD in WI and NE.

**PH in WI**

<table>
<thead>
<tr>
<th></th>
<th>GBLUP</th>
<th>GBLUP-wG</th>
<th>GBLUP-sG</th>
<th>RKHS-wG</th>
<th>RKHS-sG</th>
<th>BayesA</th>
<th>BayesB</th>
<th>RF</th>
<th>(Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WS4U-C2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>0.298</td>
<td>0.291</td>
<td>0.238</td>
<td>0.303</td>
<td>0.291</td>
<td>0.213</td>
<td>0.300</td>
<td>0.302</td>
<td>0.086</td>
</tr>
<tr>
<td>PCA</td>
<td>0.298</td>
<td>0.082</td>
<td>0.040</td>
<td>0.303</td>
<td>0.106</td>
<td>0.078</td>
<td>0.196</td>
<td>0.170</td>
<td>0.194</td>
</tr>
<tr>
<td>Cor</td>
<td>0.289</td>
<td>0.259</td>
<td>0.153</td>
<td>0.279</td>
<td>0.263</td>
<td>0.205</td>
<td>0.292</td>
<td>0.306</td>
<td><strong>0.313</strong></td>
</tr>
<tr>
<td>LD</td>
<td>0.288</td>
<td>0.302</td>
<td>0.196</td>
<td>0.293</td>
<td>0.291</td>
<td>0.212</td>
<td>0.286</td>
<td>0.294</td>
<td>0.064</td>
</tr>
<tr>
<td>(Mean)</td>
<td>0.293</td>
<td>0.234</td>
<td>0.157</td>
<td>0.2945</td>
<td>0.238</td>
<td>0.177</td>
<td>0.269</td>
<td>0.268</td>
<td>0.164</td>
</tr>
<tr>
<td><strong>Liberty-C2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>0.344</td>
<td>0.315</td>
<td>0.236</td>
<td>0.319</td>
<td>0.299</td>
<td>0.136</td>
<td>0.322</td>
<td>0.315</td>
<td>0.132</td>
</tr>
<tr>
<td>PCA</td>
<td>0.344</td>
<td>0.195</td>
<td>0.085</td>
<td>0.319</td>
<td>0.24</td>
<td>0.141</td>
<td>0.328</td>
<td>0.344</td>
<td>-0.046</td>
</tr>
<tr>
<td>Cor</td>
<td>0.299</td>
<td>0.249</td>
<td>0.225</td>
<td>0.282</td>
<td>0.253</td>
<td>0.203</td>
<td>0.277</td>
<td>0.267</td>
<td>0.270</td>
</tr>
<tr>
<td>LD</td>
<td><strong>0.371</strong></td>
<td>0.320</td>
<td>0.196</td>
<td>0.348</td>
<td>0.315</td>
<td>0.187</td>
<td>0.361</td>
<td>0.37</td>
<td>0.173</td>
</tr>
<tr>
<td>(Mean)</td>
<td>0.340</td>
<td>0.270</td>
<td>0.186</td>
<td>0.317</td>
<td>0.277</td>
<td>0.167</td>
<td>0.322</td>
<td>0.324</td>
<td>0.132</td>
</tr>
</tbody>
</table>

**PH in NE**

<table>
<thead>
<tr>
<th></th>
<th>GBLUP</th>
<th>GBLUP-wG</th>
<th>GBLUP-sG</th>
<th>RKHS-wG</th>
<th>RKHS-sG</th>
<th>BayesA</th>
<th>BayesB</th>
<th>RF</th>
<th>(Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WS4U-C2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>0.326</td>
<td>0.312</td>
<td>0.321</td>
<td>0.326</td>
<td>0.320</td>
<td>0.328</td>
<td>0.320</td>
<td>0.326</td>
<td>0.127</td>
</tr>
<tr>
<td>PCA</td>
<td>0.326</td>
<td>0.130</td>
<td>0.130</td>
<td>0.326</td>
<td>0.159</td>
<td>0.159</td>
<td>0.198</td>
<td>0.188</td>
<td>0.187</td>
</tr>
<tr>
<td>Cor</td>
<td>0.372</td>
<td>0.356</td>
<td>0.360</td>
<td>0.374</td>
<td>0.365</td>
<td>0.385</td>
<td>0.386</td>
<td><strong>0.396</strong></td>
<td>0.288</td>
</tr>
<tr>
<td>LD</td>
<td>0.290</td>
<td>0.280</td>
<td>0.227</td>
<td>0.309</td>
<td>0.295</td>
<td>0.269</td>
<td>0.291</td>
<td>0.291</td>
<td>0.003</td>
</tr>
<tr>
<td>(Mean)</td>
<td>0.329</td>
<td>0.270</td>
<td>0.260</td>
<td>0.334</td>
<td>0.285</td>
<td>0.285</td>
<td>0.299</td>
<td>0.300</td>
<td>0.151</td>
</tr>
<tr>
<td><strong>Liberty-C2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td><strong>0.500</strong></td>
<td>0.482</td>
<td>0.418</td>
<td>0.469</td>
<td>0.466</td>
<td>0.295</td>
<td>0.500</td>
<td>0.481</td>
<td>0.381</td>
</tr>
<tr>
<td>PCA</td>
<td>0.500</td>
<td>0.377</td>
<td>0.377</td>
<td>0.469</td>
<td>0.398</td>
<td>0.398</td>
<td>0.496</td>
<td>0.499</td>
<td>0.237</td>
</tr>
<tr>
<td>Cor</td>
<td>0.474</td>
<td>0.456</td>
<td>0.410</td>
<td>0.453</td>
<td>0.449</td>
<td>0.417</td>
<td>0.473</td>
<td>0.470</td>
<td>0.379</td>
</tr>
<tr>
<td>LD</td>
<td>0.477</td>
<td>0.464</td>
<td>0.426</td>
<td>0.453</td>
<td>0.452</td>
<td>0.317</td>
<td>0.479</td>
<td>0.472</td>
<td>0.380</td>
</tr>
<tr>
<td>(Mean)</td>
<td>0.488</td>
<td>0.445</td>
<td>0.408</td>
<td>0.461</td>
<td>0.441</td>
<td>0.357</td>
<td>0.487</td>
<td>0.481</td>
<td>0.344</td>
</tr>
</tbody>
</table>
### HD in WI

<table>
<thead>
<tr>
<th></th>
<th>GBLUP</th>
<th>GBLUP-(wG)</th>
<th>GBLUP-(sG)</th>
<th>RKHS-(wG)</th>
<th>RKHS-(sG)</th>
<th>BayesA</th>
<th>BayesB</th>
<th>RF (Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WS4U-C2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>0.254</td>
<td>0.262</td>
<td>0.207</td>
<td>0.271</td>
<td>0.271</td>
<td>0.224</td>
<td>0.271</td>
<td>0.217</td>
</tr>
<tr>
<td>PCA</td>
<td>0.254</td>
<td>0.070</td>
<td>0.041</td>
<td>0.271</td>
<td>0.114</td>
<td>0.091</td>
<td>0.204</td>
<td>0.179</td>
</tr>
<tr>
<td>Cor</td>
<td>0.224</td>
<td>0.200</td>
<td>0.101</td>
<td>0.232</td>
<td>0.209</td>
<td>0.094</td>
<td>0.212</td>
<td>0.224</td>
</tr>
<tr>
<td>LD</td>
<td>0.234</td>
<td>0.247</td>
<td>0.234</td>
<td>0.249</td>
<td>0.257</td>
<td>0.238</td>
<td>0.280</td>
<td>0.192</td>
</tr>
<tr>
<td>(Mean)</td>
<td>0.242</td>
<td>0.195</td>
<td>0.146</td>
<td>0.256</td>
<td>0.213</td>
<td>0.162</td>
<td>0.242</td>
<td>0.203</td>
</tr>
<tr>
<td><strong>Liberty-C2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>0.577</td>
<td>0.598</td>
<td>0.599</td>
<td>0.581</td>
<td>0.595</td>
<td>0.587</td>
<td>0.565</td>
<td>0.555</td>
</tr>
<tr>
<td>PCA</td>
<td>0.577</td>
<td>0.446</td>
<td>0.352</td>
<td>0.581</td>
<td>0.509</td>
<td>0.428</td>
<td>0.570</td>
<td>0.569</td>
</tr>
<tr>
<td>Cor</td>
<td>0.528</td>
<td>0.550</td>
<td>(\textbf{0.604})</td>
<td>0.527</td>
<td>0.547</td>
<td>0.599</td>
<td>0.528</td>
<td>0.517</td>
</tr>
<tr>
<td>LD</td>
<td>0.563</td>
<td>0.564</td>
<td>0.583</td>
<td>0.571</td>
<td>0.567</td>
<td>0.579</td>
<td>0.565</td>
<td>0.560</td>
</tr>
<tr>
<td>(Mean)</td>
<td>0.561</td>
<td>0.540</td>
<td>0.535</td>
<td>0.565</td>
<td>0.555</td>
<td>0.548</td>
<td>0.557</td>
<td>0.550</td>
</tr>
</tbody>
</table>

### HD in NE

<table>
<thead>
<tr>
<th></th>
<th>GBLUP</th>
<th>GBLUP-(wG)</th>
<th>GBLUP-(sG)</th>
<th>RKHS-(wG)</th>
<th>RKHS-(sG)</th>
<th>BayesA</th>
<th>BayesB</th>
<th>RF (Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WS4U-C2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>0.246</td>
<td>0.233</td>
<td>0.230</td>
<td>0.239</td>
<td>0.232</td>
<td>0.240</td>
<td>0.193</td>
<td>0.226</td>
</tr>
<tr>
<td>PCA</td>
<td>0.246</td>
<td>0.151</td>
<td>0.114</td>
<td>0.239</td>
<td>0.202</td>
<td>0.187</td>
<td>0.230</td>
<td>0.229</td>
</tr>
<tr>
<td>Cor</td>
<td>0.212</td>
<td>0.179</td>
<td>0.161</td>
<td>0.230</td>
<td>0.197</td>
<td>0.161</td>
<td>0.214</td>
<td>0.271</td>
</tr>
<tr>
<td>LD</td>
<td>0.213</td>
<td>0.158</td>
<td>0.162</td>
<td>0.208</td>
<td>0.166</td>
<td>0.195</td>
<td>0.213</td>
<td>0.216</td>
</tr>
<tr>
<td>(Mean)</td>
<td>0.229</td>
<td>0.180</td>
<td>0.167</td>
<td>0.229</td>
<td>0.199</td>
<td>0.196</td>
<td>0.213</td>
<td>0.236</td>
</tr>
<tr>
<td><strong>Liberty-C2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>(\textbf{0.420})</td>
<td>0.413</td>
<td>0.351</td>
<td>0.396</td>
<td>0.398</td>
<td>0.295</td>
<td>0.406</td>
<td>0.399</td>
</tr>
<tr>
<td>PCA</td>
<td>0.420</td>
<td>0.298</td>
<td>0.073</td>
<td>0.396</td>
<td>0.330</td>
<td>0.169</td>
<td>0.387</td>
<td>0.366</td>
</tr>
<tr>
<td>Cor</td>
<td>0.368</td>
<td>0.358</td>
<td>0.331</td>
<td>0.353</td>
<td>0.354</td>
<td>0.345</td>
<td>0.356</td>
<td>0.360</td>
</tr>
<tr>
<td>LD</td>
<td>0.403</td>
<td>0.397</td>
<td>0.343</td>
<td>0.376</td>
<td>0.383</td>
<td>0.328</td>
<td>0.400</td>
<td>0.409</td>
</tr>
<tr>
<td>(Mean)</td>
<td>0.403</td>
<td>0.367</td>
<td>0.275</td>
<td>0.380</td>
<td>0.366</td>
<td>0.284</td>
<td>0.387</td>
<td>0.384</td>
</tr>
</tbody>
</table>

Prediction accuracies were estimated with a within-population/within-environment learning scheme in five-fold cross-validation, with no replication. For a given population and outcome (trait-location combination), the highest average value across marker-data transformations is underlined; the highest value across prediction procedures is underlined and bolded.
Figure S1 – Genetic relationship coefficients in WS4U-C2 (upper panel) and Liberty-C2 (lower panel), based on marker
features from alternate data transformations (PCA, Cor or LD; see section Material and methods), compared to Base (only
centering of expected allelic dosages).
Figure S2 — Distribution of minor allele frequency (MAF) in (a) WS4U-C2 and (b) Liberty-C2; (c) Concordance of MAF from WS4U-C2 to Liberty-C2; the blue curve corresponds to the mean value (and its 95%-confidence interval) from a cubic-regression spline model assuming a Normal distribution for MAF in Liberty-C2. Cubic-regression spline models were fitted using the R package mgcv (Wood 2006).
Figure S3 – Heatmaps of genomic correlations between trait BLUPs for traits measured (a) in WS4U-C2; (b) in Liberty-C2.

White cells show unavailable estimates (due to non-convergence of the fitting algorithm). Significance of inferred genotypic correlations was assessed by likelihood-ratio tests; *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. Genomic correlations were estimated in a multivariate GBLUP model, with $\mathbf{X}_{\text{Base}}$ as input, using ASREML-R (Butler et al. 2007).
Figure S4 – Bar plots of mean prediction accuracies from non-replicated five-fold cross-validation for all outcomes. The values showed correspond to those in Tables 3 and S3. For a given outcome, only the values for Base and the selected marker-data transformation are shown.
Figure S5 – Validation of selected prediction procedures for PH and HD in WI and NE. Prediction accuracies ($r_{gb}$) were estimated with a within-population/within-environment learning scheme in five-fold cross-validation, replicated ten times. In each boxplot, up to two comparisons are made: (i) the candidate-transformation procedure (selected marker-data transformation according to non-replicated five-fold cross-validation in a GBLUP model; Table S3) is compared to the standard procedure (Base – GBLUP) – if relevant; and (ii) the candidate procedure (selected prediction procedure according to non-replicated five-fold cross-validation; Table S3) is compared to the candidate-transformation procedure. The significance of differences in prediction accuracies was assessed by two-sided paired Dunnett tests, which accounted for multiple testing of data.
transformations, in (i), and of prediction models, in (ii). The t-statistics in Dunnett tests were adjusted to account for correlation among training sets in cross-validation, as described in Bouckaert and Frank (2004).
Figure S6 – Validation of selected prediction procedures for all outcomes, ignoring all alternate marker-data transformations (i.e., other than Base). Prediction accuracies ($r_{gg}$) were estimated with a within-population/within-environment learning scheme in five-fold cross-validation, replicated ten times. In each boxplot, the selected procedure (best prediction procedure according to non-replicated five-fold cross-validation; Tables 3 and S3) is compared to the standard procedure (Base – GBLUP). The significance of differences in prediction accuracies was assessed by two-sided paired Dunnett tests, which accounted for multiple testing of data transformations of prediction models. The t-statistics in Dunnett tests were adjusted to account for correlation among training sets in cross-validation, as described in Bouckaert and Frank (2004).
File S1: Raw phenotypic data

Excel (.xls) file consisting of two sheets: one for traits measured at the plant level (Heading date – in day-of-year – and Plant height – in cm) and the other for traits measured at the plot level (Fresh biomass yield, Sample fresh matter weight and Sample dry matter weight – in grams).

Population: WS4U-C2 or Liberty-C2
Genotype: genotype index for female parent within Population
Row: row number (as coordinate)
Col: column number (as coordinate)
Rep: replicate
Plant (for plant-level traits): plant index within plot
Number_of_plants (for plot-level traits): number of (surviving) plants within plot
Location: ARL (Arlington, WI) or MEAD (Mead, NE)
Year: 2012, 2013 or 2014

Plots are combinations of Location, Genotype and Rep.

In this study, dry matter yield was calculated as:
(Fresh biomass yield) * (Sample dry matter weight) / (Sample fresh matter weight).

File S2: Half-sib (HS) family BLUPs

Excel (.xls) file containing best linear unbiased predictions (predicted random effects) of HS family effects as obtained from the models described in the article, fitted for each population and outcome separately.

Population: WS4U-C2 or Liberty-C2
Genotype: genotype index for HS-family parent within Population
PH_WI, HD_WI, DMY_WI: Plant height, Heading date and Dry matter yield, respectively, in Arlington, WI.
PH_NE, HD_NE, DMY_NE: Plant height, Heading date and Dry matter yield, respectively, in Mead, NE.

File S3: Untransformed genotype calls at marker loci

Compression (.rar file) of a .csv file with expected allelic dosages (expected number of alternate alleles, as per posterior probabilities from genotype calling), rounded to the fifth decimal digit, at selected markers for each genotype (HS-family parent). Rows correspond to the q* = 141,030 marker loci selected across both populations (see main text) and columns correspond to all 247 genotypes across both populations. Note that a .xlsx version of this file, with non-rounded values, is available from http://dfrc.wisc.edu/sniper/.

Row name: Chromosome index + SNP index (e.g. ‘Chr05bSNP1187770’ refers to SNP 1187770 in chromosome 5, subgenome B)
Column name: Population name + genotype index (e.g. ‘Liberty-C2_31804’ refers to genotype 31804 in population Liberty-C2)