G3: Genes|Genomes|Genetics Early Online, published on March 12, 2015 as doi:10.1534/g3.115.017533

Genome-wide association study based on multiple imputation with low-depth sequencing data: application to biofuel traits in reed canarygrass

Guillaume P. Ramstein[*], Alexander E. Lipka[§, †], Fei Lu[§], Denise E. Costich[‡], Jerome H. Cherney[**], Edward S. Buckler[§, §§, ††], Michael D. Casler[*, ‡‡]

[*] Department of Agronomy, University of Wisconsin-Madison, Madison, Wisconsin, United States of America

[§] Institute for Genomic Diversity, Cornell University, Ithaca, New York, United States of America

[†] Current address: Department of Crop Sciences, University of Illinois, Urbana, Illinois, United States of America

[‡] International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

[**] Soil and Crops Section, School of Integrative Plant Science, Cornell University, Ithaca, New York, United States of America

[§§] Agricultural Research Service, United States Department of Agriculture, Ithaca, New York, United States of America

[††] Department of Plant Breeding and Genetics, Cornell University, Ithaca, New York, United States of America

[‡‡] Agricultural Research Service, United States Department of Agriculture, Madison, Wisconsin, United States of America

SHORT RUNNING TITLE

Multiple imputation in association study

KEYWORDS

Genome-wide association study, multiple imputation, genotyping by sequencing, bioenergy, *Phalaris* spp.

CORRESPONDING AUTHOR

Guillaume Ramstein

UW-Madison

Department of agronomy

1575 Linden Drive

Madison, WI 53706

Tel.: + 1 608 890 0050

E-mail: ramstein@wisc.edu

ABSTRACT

Genotyping by sequencing allows for large-scale genetic analyses in plant species with no reference genome, but sets the challenge of sound inference in presence of uncertain genotypes. We report an imputation-based genome-wide association study (GWAS) in reed canarygrass (*Phalaris arundinacea* L., *Phalaris caesia* Nees), a cool-season grass species with potential as a biofuel crop. Our study involved two linkage populations and an association panel of 590 reed canarygrass genotypes. Plants were assayed for up to 5,228 single nucleotide polymorphism markers and 35 traits. The genotypic markers were derived from low-depth sequencing with 78% missing data on average. To soundly infer marker-trait associations, multiple imputation (MI) was used: several imputes of the marker data were generated to reflect imputation uncertainty and association tests were performed on marker effects across imputes. A total of nine significant markers were identified, three of which showed significant homology with the *Brachypodium dystachion* genome. Because no physical map of the reed canarygrass genome was available, imputation was conducted using classification trees. In general, MI showed good consistency with the complete-case analysis and adequate control over imputation uncertainty. A gain in significance of marker effects was achieved through MI, but only for rare cases when the amount of missing data was < 45%. In addition to providing insight into the genetic basis of important traits in reed canarygrass, this study presents one of the first applications of MI to genome-wide analyses and provides useful guidelines for conducting GWAS based on genotyping-by-sequencing data.

INTRODUCTION

Perennial crops, which include herbaceous energy crops (HEC), are increasingly studied as potential significant sources of energy, due to their environmental benefits and the increasing

prices of petroleum. In its 2005 billion-ton supply report, the US Department of Agriculture (USDA) and the US Department of Energy (USDOE) set the goal of a 30% replacement of U.S. petroleum consumption with biofuels by 2030. This goal implies a production of approximately one billion dry matter tons of biomass per year from forest and agricultural lands. According to the report's projections and assumptions, achieving that objective will require (i) the conversion of active croplands, pasture land and lands under the conservation reserve program (CRP) into perennial crop lands and (ii) achieving biomass yields ranging between 5.5 and 8 dry tons per acre from perennial crops (between 12.4 and 19.8 Mg ha^{-1}; USDA and USDOE Biomass R&D technical advisory group 2005). Plant breeding has an important part to play in achieving such yields. To efficiently select for higher biomass yield, selection may act on secondary traits such as plant height and flowering time (Price and Casler 2014), resistance to biotic stress, tiller density (Boe and Beck 2008), leaf area and plant architecture. Biomass quality considerations, for conversion into bioenergy, bring another suite of traits to bear in HEC breeding (e.g. Vogel et al. 2011). Common methods for transforming biomass feedstock into energy include direct combustion, pyrolysis and fermentation of sugars (soluble sugars, starch, cellulose and hemicellulose) into ethanol (Wrobel et al. 2009).

Reed canarygrass (*Phalaris arundinacea* L., *Phalaris caesia* Nees) is a promising HEC in North America. It belongs to the tribe *Avenae* (sub-family *Pooideae*, family *Poaceae*), which includes the oat genus *Avena* (Quintanar et al. 2007). Of the grass model species, i.e. *Brachypodium distachyon* (Brachypodium), *Oryza sativa* (Rice), *Zea mays* (maize) and *Sorghum bicolor* (Sorghum), Brachypodium is the most closely related to reed canarygrass (Bouchenak-Khelladi et al. 2008). Reed canarygrass is a species complex which comprises two chromosomal races: the tetraploid race, *Phalaris arundinacea* L. (2n = 4x = 28) thought to be native to Europe, Asia and

North America (Jakubowski et al. 2012), and the hexaploid race, *Phalaris caesia* Nees (2n = 6x = 42) (McWilliam and Neal-Smith 1962; Baldini 1995). Most cultivars and wild accessions of *P. arundinacea* found in North America are of European ancestry (Casler et al. 2009a, Jakubowski et al. 2011). Reports of breeding efforts in *P. arundinacea* trace back to the early 20th century in North America (Casler 2010) but this species was already cultivated in the 18th century in Europe (Always 1931). Some non-crop uses of reed canarygrass have been reported, e.g. phytoremediation (Picard et al. 2005), erosion control (Rice and Pinkerton 1993) and paper production (Pahkala and Pihala 2000). However, reed canarygrass has mostly been used as a forage crop. Consistently, most recent breeding efforts have focused on low alkaloid content for palatability to livestock, not on biomass yield (Wrobel et al. 2009).

Most of the research on HEC in the U.S. has been focused on switchgrass (*Panicum virgatum* L.; Sanderson et al. 1996), but reed canarygrass presents characteristics that may complement those of switchgrass: it is particularly tolerant to northern climates (Casler et al. 2009a), as well as to soil acidity, alkalinity and moisture content (Bittman et al. 1980) and high levels of metals and minerals (Cureton et al. 1990). However, biomass yields in reed canarygrass are not very high: based on trials in the U.S. Midwest, involving wild accessions and cultivars, genotype means for dry matter yield have been estimated to range from 7.6 to 10 Mg.ha^{-1} (Casler et al. 2009b). Besides, quality tends to be lower in reed canarygrass than it is in switchgrass (Cherney et al. 1988). Nonetheless, judging from the significant genotypic variation observed in both biomass yield (Asay et al. 1968; Casler et al. 2009b) and quality (Carlson et al. 1996; Olmstead et al. 2013), improvement by selection of these primary biofuel traits is feasible.

Because no linkage map and no genome sequence are available in reed canarygrass, association studies have not been conducted in this species complex. However, genotyping by sequencing

(GBS) provides the opportunity to call polymorphisms without prior marker development. With this technology, a genome-wide association study (GWAS) can be performed on reed canarygrass, but because no reference sequence is available, one limitation of such study is the inability to assign single nucleotide polymorphisms (SNP) to specific physical positions in the reed canarygrass genome. This has two important implications: (i) SNP-trait associations cannot be mapped to a particular region of the genome, and (ii) no imputation method based on marker sequences, e.g. hidden Markov models (HMM) or sliding-window algorithms, may be used. However, other methods are available for imputing unordered marker data (Poland et al. 2012; Rutkoski et al. 2013). Another limitation of using SNPs derived from GBS is the high amount of missing values in marker data, and, therefore the importance of accounting for uncertainty in the imputation of marker genotypes.

The purpose of this study was to make statistically sound inferences about associations between GBS markers with high frequency of missing values and biofuel traits (related to biomass yield and quality) in reed canarygrass. GWAS was performed under the assumption of disomic inheritance in reed canarygrass, which is suggested by the alloploidy of a closely-related species, *Phalaris aquatica* (Carlson et al. 1996). To perform statistically valid association tests, it was necessary to avoid false positives, not only by controlling for population structure and familial relatedness (Yu et al. 2006), but also by accounting for imputation uncertainty. Our study exemplifies the use of multiple imputation, initially proposed by Rubin (1987), to account for this source of variability when making inferences. Classification tree models, shown by Poland et al. (2012) and Rutkoski et al. (2013) to be promising, were used under this framework to impute missing values without information on markers' genomic location and order. After a description of the variability present in the panels under study, we examine the general behavior

of inferences in multiple imputation, compared to more traditional methods that would not fully account for imputation uncertainty. Then, we present our results of GWAS performed in a multiple-imputation framework and describe the significant markers, in light of where they map onto the *B. distachyon* genome.

MATERIAL AND METHODS

Populations

Three panels were assessed in this study: two randomly-segregating populations derived from biparental crosses (LP1 and LP2) comprising 177 and 189 clones respectively (Table S1); and one association panel (AP), comprising 590 clones originating from North America and Europe (Table S1). LP1 and LP2 were derived from two distinct crosses involving genotypes of WR00, a tetraploid cultivar originating from Wisconsin, USA. In AP, four populations (AR Upland, Superior, PI-284179 and PI-236525) were accessions of *P. caesia* (hexaploid). The AP subset consisting of only the 550 *P. arundinaceae* clones is referred to as AP-4x.

Clones were assessed in spaced-plant trials arranged in a Sets-in-Reps design with 2 replicates. The trials were performed in two locations: Arlington, WI (43.3°N, 89.4°W), and Ithaca, NY (42.6°N, 76.4°W). Soil type was Plano silt loam (fine-silty, mixed, mesic Typic Argiudoll) in Arlington and Niagara silt loam (fine-silty, mixed, active, mesic Aeric Endoaqualf) in Ithaca, NY.

Phenotypic data

Traits related to disease resistance (Ds), morphology (Standability, St; Leaf width, LW; Leaf length, LL; Total stem count, STC; Plant height, PH; Full height, FH), phenology (Heading date, HD; Anthesis date, AD) and quality (Dry matter percentage, DM; Neutral detergent fiber, NDF;

Acid detergent fiber, ADF; NDF digestibility, NDFD; Acid insoluble ash, AIA; Klason lignin, Lignin; Crude protein, CP; Ash, ASH; Calcium, Ca; Chlorine, Cl; Copper, Cu; Iron, Fe; Potassium, K; Magnesium, Mg; Manganese, Mn; Sodium, Na; Phosphorus, P; Sulfur, S; Zinc, Zn; Glucose, GLC; Galactose, GAL; Xylose, XYL; Arabinose, ARA; GLC conversion efficiency, GLC_Eff; XYL conversion efficiency, XYL_Eff; Energy content, BTU) were determined on plants grown at Ithaca in 2009-2011 and at Arlington in 2010-2011 (Table 1). Quality traits were predicted from near infrared reflectance spectroscopy (NIRS) measurements, using methodology described by Vogel et al. (2011). The NIRS prediction equations were developed on a diverse set of 110 reed canarygrass samples from the experiments described by Casler et al. (2009b). Wet-laboratory traits were determined on these samples using the procedures described by Vogel et al. (2011). The validity of predictions from the predictive models were verified by the extremely low frequency of outliers: 9 biomass-quality samples with Mahalanobis distance > 3 out of a total of >3300 samples (Shenk and Westerhaus 1991). The raw phenotypic data is available for download from http://dfrc.wisc.edu/sniper/.

Association analyses were performed on best linear unbiased estimations (BLUEs) of genotypes' performance for a given trait, inferred from the following linear mixed model:

$$y_{ijlmrc} = mean + genotype_i + location_j + year(location)_{jl} + rep.(location)_{jm} + \left(year \times rep.(location)\right)_{jlm} + (genotype \times location)_{ij} + \left(genotype \times year(location)\right)_{ijl} + plot\left(year(location)\right)_{jlrc} + \varepsilon_{ijlm}$$

where y_{ijlmrc} are measurements at one of the 35 traits considered; $genotype_i$ is the genotypic value of clone i, modeled as fixed (to guarantee convergence of the fitting algorithm and avoid assumptions about genotypes' sampling). For all other terms, the corresponding effects were considered random, independent and identically, normally distributed; $location_j$ is the effect of location j; $year(location)_{jl}$ is the effect of year l within location j; $rep.(location)_{jm}$ is the

effect of replicate m within location j; × indicates interactions; $plot(year(location))_{jlrc}$ is the effect of the plot at row r and column c, within an environment (year l within location j). Plot effects within environments were modeled as following a $\text{Normal}\left(0, (\boldsymbol{\Sigma}_r \otimes \boldsymbol{\Sigma}_c)\sigma^2_{plot}\right)$, where $\boldsymbol{\Sigma}_r \otimes \boldsymbol{\Sigma}_c$ is the Kronecker product of the 1st-order autoregressive covariance matrices on rows and on columns, respectively. The mixed models were fitted using ASREML-R (Butler et al. 2007).

The matrix of genotypes' BLUEs, computed as described above, can be downloaded from http://dfrc.wisc.edu/sniper/.

Marker data and quality control

Genome reduction, by *ApeKI* restriction, and sequencing were performed according to Elshire et al. (2011). Reduced DNA samples were sequenced on the Illumina HiSeq 2000, with 95 samples plus one negative control per lane. To simultaneously discover SNP markers and call genotypes, the UNEAK pipeline was used (http://www.maizegenetics.net/gbs-bioinformatics; Lu et al. 2013). This pipeline trims reads to a 64-bp length, to limit sequencing errors and speed up computation, and discards markers based on a network filter, designed to detect and eliminate markers showing complex relationships with others, which suggests paralogy and/or sequencing errors (Lu et al. 2013). A total of 29,313 SNPs were called out of the UNEAK pipeline. Marker genotypes were coded as allelic dosages: -1 for homozygotes at the reference allele, 0 for heterozygotes and 1 for homozygotes at the alternate allele, assuming disomic inheritance in reed canarygrass (Carlson et al. 1996). The raw genotypic data is available for download from http://dfrc.wisc.edu/sniper/.

Markers were selected based on proportion of missing values (PMV) < 0.90 in each of three panels, separately. It is typical to filter out GBS SNPs by a predetermined low missing rate (e.g. PMV < 0.20). However, we did not filter these markers to avoid removing potentially useful information and because a rigorous study of the behavior of the MI approach at different missing rates is merited. After this first filtering step, 18,818 markers were retained. Marker variables were then discarded if they met one of the following criteria: (i) being "constant", i.e., having a variance close to 0 (only 24 marker were discarded based on that criterion); (ii) being "collinear", i.e., being correlated by at least 0.999 (in absolute value) to some other marker variable with a smaller amount of missing values in the dataset. This filtering step was recommended by van Buuren and Groothuis-Oudshoorn (2011) and implemented in the `mice` R package. Not only did this avoid overly conservative tests in GWAS due to inadequate adjustment for multiple testing on highly correlated test statistics, but it also ensured that multiple imputations were not biased, as a result of strong correlations among predictors, and appropriately reflected imputation uncertainty. At this point, 6,138 markers were considered for further analyses. Figure 1 shows the distributions of PMV and MAF for marker data after the first filtering step (18,818 markers) and after the second filtering step (6,138 markers). As expected, filtering for variability and non-collinearity preferentially discarded markers with high PMV and low MAF. Note that filtering out markers on PMV < 0.80 across panels (which is still a very lenient criterion) would have resulted in only 3,419 markers being considered for subsequent analyses (Table S2). Also, for markers not meeting the criteria of van Buuren and Groothuis-Oudshoorn (2011), estimates of marker effects from averaged imputed data ((AD); see subsection "Association analyses") tended to show larger deviations from those obtained based on non-missing data only ((CC); see subsection "Association analyses"), no matter what the

imputation uncertainty was (Table S3, Figure S2). Finally, estimates of marker effects from multiple imputes ((MI); see subsection "Association analyses") tended to show excessively strong shrinkage in comparison with (AD) estimates, with little adjustment of estimates in response to imputation uncertainty (Table S4, Figure S3).

The high values of PMV among selected markers (78% of missing values on average) show low sequencing depth of the GBS. From the quality control step, the marker-data matrix **X** was produced, in which missing values were coded as NA. Matrix **X**, used as input to the imputation procedure, is available for download from http://dfrc.wisc.edu/sniper/.

Marker data imputation

General principle of multiple multivariate imputation: Consider some incomplete data set **X**, with $\mathbf{X}_{obs}$ and $\mathbf{X}_{mis}$ denoting the observed and missing data, respectively. Given an estimand of interest θ (e.g., a model parameter), multiple imputation (MI) aims at producing a correct Monte Carlo (MC) approximation of $p(\theta|\mathbf{X}_{obs}) = \int p\left(\theta\middle|\mathbf{X}_{obs}, \widehat{\mathbf{X}}_{mis}\right) p\left(\widehat{\mathbf{X}}_{mis}\middle|\mathbf{X}_{obs}\right) d\widehat{\mathbf{X}}_{mis}$ from m imputes of **X**, with $\widehat{\mathbf{X}}_{mis}$ denoting some imputation of $\mathbf{X}_{mis}$**.** One implication of the MC approximation being correct is that, given an impute $\left\{\mathbf{X}_{obs}, \widehat{\mathbf{X}}_{mis}\right\}$, the distribution $p\left(\theta\middle|\mathbf{X}_{obs}, \widehat{\mathbf{X}}_{mis}\right)$ is well-approximated; i.e., the sampling model is correct. Another implication, which is of particular concern when performing MI, is that the distribution $p\left(\widehat{\mathbf{X}}_{mis}\middle|\mathbf{X}_{obs}\right)$ is also well approximated; i.e. the imputation procedure is proper (Rubin 1987), which is defined as follows: let $\dot{\theta}$ be an estimate of θ from a given impute, W be the within-impute estimation variance of $\dot{\theta}$, and B^* be the among-impute variance of $\dot{\theta}$ (corrected for a finite number of imputes). An imputation procedure that is proper meets the following criteria: (i) $\mathrm{E}[\bar{\theta}|\mathbf{X}] = \hat{\theta}$, i.e. the average of $\dot{\theta}$ over imputes ($\bar{\theta}$) is an unbiased estimator of $\hat{\theta}$, the estimate of θ from the

hypothetically complete dataset; (ii) $\mathrm{E}[\bar{W}|\mathbf{X}] = W$ and (iii) $\mathrm{E}[B^*|\mathbf{X}] \geq \mathrm{Var}(\bar{\theta})$ (Rubin 1987; van Buuren 2012). Condition (i) implies correct assumptions regarding the imputation model and the source of missingness in the data. Typically, the missing-at-random assumption (no factor other than those accounted for in the imputation model caused missingness; Little and Rubin 1987) is necessary to guarantee that condition (i) holds. Conditions (ii) and (iii) imply that inferences from MI are confidence valid (van Buuren 2012), i.e. that the total variance of $\dot{\theta}$ is realized in a conservative way: $\mathrm{E}\left[\widehat{\mathrm{Var}(\dot{\theta})}\right] \geq W + \mathrm{Var}(\bar{\theta})$. In this study, we will assess imputation models for unbiasedness, under specific assumptions (condition (i)), but we will not test imputation models for their ability to preserve variability in the data (condition (ii)) or to correctly reflect among-impute variability (condition (iii)).

When performing MI on data sets containing missing values at several variables (SNP markers), one must sample from the joint distribution of missing values at all variables. To achieve this, two strategies have been proposed: joint modelling (JM) and fully conditional specification (FCS). JM consists of sampling from the joint distribution directly, by ordinary MC (Rubin and Shafer 1990; Shafer 2010). This method is theoretically sound, but it requires complex model specifications and assumptions which, if they are not correct, may result in imputation bias. JM cannot accommodate tree-based approaches that have the advantage of being flexible and not requiring any model specification. FCS, on the other hand, implicitly samples from some hypothetical joint distribution by repeatedly sampling from the fully conditional distributions at each variable of interest, using a Gibbs sampling scheme (van Buuren et al. 2006; van Buuren 2007). Because it relies on Markov chain MC (MCMC), FCS can be computationally costly. Also, there is no guarantee that the joint distribution actually exists, so FCS is usually described as a pseudo-Gibbs sampling procedure (Gelman and Raghunathan 2001; van Buuren et al. 2006;

van Buuren 2007). However, simulation studies have suggested that FCS is robust to such issues (van Buuren et al. 2006). A critical incentive for using FCS rather than JM is that FCS allows for much more flexibility in the imputation procedure: each fully conditional distribution can be derived separately, using either parametric (e.g. linear regression) or non-parametric models (e.g. classification trees). This flexibility is particularly valuable when dealing with missing values at many variables, in which case JM may simply not be practical (Gelman and Raghunathan 2001).

Implementation of multiple imputation: In this study, we sampled missing values from $p(\widehat{\mathbf{X}}_{mis}|\mathbf{X}_{obs})$ by FCS. In the pseudo-Gibbs sampling process, missing values at the k^{th} SNP were sampled from $p(\widehat{\mathbf{X}}_{mis_k}|\mathbf{X}_{obs}, \widehat{\mathbf{X}}_{mis_{-k}})$ based on a classification and regression tree model (CART; Breiman et al. 1984). For unbiasedness, imputation of $\mathbf{X}_{mis}$ relied on the missing-completely-at-random (MCAR) assumption, stating that missingness occurred at random, with no factor causing SNP genotypes to be systematically missing (Little and Rubin 1987). CART presents the advantages of not requiring explicit model specification and conveniently accommodating non-linear effects.

The general algorithm was:

For impute $r = 1, \dots, m$:

1. For $k = 1, \dots, q$, fill in missing values at the k^{th} marker by random draws $\widehat{\mathbf{X}}_{mis_k}{}^{(r,0)}$ from $\mathbf{X}_{obs}$.

2. For $l = 1, \dots, L$:

    For $k = 1, \dots, q$:

    Sample $\widehat{\mathbf{X}}_{mis_k}{}^{(r,l)}$ from

    $$p\left(\widehat{\mathbf{X}}_{mis_k}\middle|\mathbf{X}_{obs}, \widehat{\mathbf{X}}_{mis_1}{}^{(r,l)}, \dots, \widehat{\mathbf{X}}_{mis_{k-1}}{}^{(r,l)}, \widehat{\mathbf{X}}_{mis_{k+1}}{}^{(r,l-1)}, \dots, \widehat{\mathbf{X}}_{mis_q}{}^{(r,l-1)}\right)$$

3. Return $\left\{\mathbf{X}_{obs}, \widehat{\mathbf{X}}_{mis}{}^{(r,L)}\right\}$ as $\dot{\mathbf{X}}^{(r)}$,

where m is the total number of imputes of $\mathbf{X}$, q is the number of marker variables, and L is the number of MCMC iterations.

The number of imputes m was set to 20, based on available computational and memory resources. The number of iterations L up to which the actual sampling occurred (i.e., the burn-in period) was set to 10, based on previous studies (Dai et al. 2006; Burgette and Reiter 2010) and available computational resources.

MI, based on CART, was implemented according to Doove et al. (2014) with packages `mice` (van Buuren and Groothuis-Oudshoorn 2011) and `rpart` (Therneau and Atkinson 1997) in the R programming language (R Development Core Team 2014). CARTs were fitted with pruning up to at least 5 donors per terminal node (i.e., no less than 5 observations at each leaf in the tree). For the MI algorithm to be computationally tractable, only a subset of SNPs was considered for fitting the CART models: for each SNP k, only the 500 SNPs showing the highest marginal mutual information with SNP k were considered as potential predictors.

MI implemented as described above took 1 day per imputation (chain of 10 iterations) on a workstation consisting of 24 Intel® Xeon® X7460 CPU processors @ 2.66 GHz with 264 Gb of RAM. The procedure was parallelized on 5 $(m/4)$ threads.

Two types of marker data were generated from the multiple imputation procedure: the MI set of matrices $\dot{\mathbf{X}} = \left\{\dot{\mathbf{X}}^{(1)}, \ldots, \dot{\mathbf{X}}^{(m)}\right\}$ and the average-dosage (AD) matrix $\overline{\mathbf{X}} = avg_r\left(\dot{\mathbf{X}}^{(r)}\right)$ (i.e., each element in $\overline{\mathbf{X}}$ was the mean of genotype codes across the $m = 20$ imputes). The array $\dot{\mathbf{X}}$ and the matrix $\overline{\mathbf{X}}$ from the imputation procedure can be downloaded from http://dfrc.wisc.edu/sniper/.

Association analyses

Data subsets: Only tetraploid samples were considered for GWAS, and the following subsets of the data were examined separately when testing marker-trait associations: LP1, LP2 and AP-4x (AP panel with only *P. arundinacea* samples), consisting of $n = 177, 189, 550$ individuals, respectively. After discarding marker variables with MAF < 0.05 (as estimated from $\mathbf{X}_{obs}$), in each subset separately, 5,024, 5,096 and 5,228 markers were assayed in association testing, respectively. Figure S1 shows, for each subset separately, the distribution of PMV and MAF for markers selected prior to conducting GWAS. The distributions are equivalent across subsets and similar to that observed with the larger set of 6,138 markers, with the exception that the average MAF increased from 0.28 to 0.32.

Association model: Within each of the three data subsets (LP1, LP2 and AP-4x), the linear mixed model of Yu et al. (2006) was fitted for each SNP k retained, using the P3D (population parameters previously determined) approximation for computational efficiency (Zhang et al. 2010; Kang et al. 2010):

$$\mathbf{g}_{obs} = \mu + \mathbf{X}_{k\,obs}\beta + \bar{\mathbf{Q}}_{obs}\mathbf{v} + \mathbf{Z}_{obs}\mathbf{u} + \mathbf{e}_{obs} \qquad \text{(CC)}$$

$$\mathbf{g} = \mu + \bar{\mathbf{X}}_k\beta + \bar{\mathbf{Q}}\mathbf{v} + \mathbf{Z}\mathbf{u} + \mathbf{e} \qquad \text{(AD)}$$

$$\mathbf{g} = \mu + \dot{\mathbf{X}}_k\beta + \bar{\mathbf{Q}}\mathbf{v} + \mathbf{Z}\mathbf{u} + \mathbf{e} \qquad \text{(MI)}$$

$$\mathbf{u} \sim \text{Normal}(0, \bar{\mathbf{K}}\sigma_u^2);\ \mathbf{e} \sim \text{Normal}(0, \mathbf{I}\sigma_e^2)$$

$$\mathbf{g} = \mu + \dot{\mathbf{X}}_k\beta + \dot{\mathbf{Q}}\mathbf{v} + \mathbf{Z}\mathbf{u} + \mathbf{e} \qquad \text{(MI*)}$$

$$\mathbf{u} \sim \text{Normal}\left(0, \dot{\mathbf{K}}\sigma_u^2\right);\ \mathbf{e} \sim \text{Normal}(0, \mathbf{I}\sigma_e^2)$$

where $\mathbf{g} = \{genotype_i\}$ is the n-vector of clones' genotypic values as described above; $\mathbf{X}_k$, $\overline{\mathbf{X}}_k$ and $\dot{\mathbf{X}}_k$ are the vectors of allelic dosage for SNP k and correspond to the three types of marker data described above; β is the effect of SNP k; **Q** is the matrix of the first t components used to account for population structure (Price et al. 2006) obtained from a principal component analysis (PCA) performed either on $\overline{\mathbf{X}}$ ($\overline{\mathbf{Q}}$) or $\dot{\mathbf{X}}$ ($\dot{\mathbf{Q}}$); **v** is the vector of their effects; **Z** is the design matrix for relating observations to clones; **u** is the vector of random polygenic effects; **K** is the realized genetic relationship matrix estimated either from $\overline{\mathbf{X}}$ ($\overline{\mathbf{K}}$) or $\dot{\mathbf{X}}$ ($\dot{\mathbf{K}}$); **e** is the vector of residuals; **I** is the identity matrix. σ_u^2 and σ_e^2 were estimated by restricted maximum likelihood (REML). The R package `rrBLUP` (Endelman 2011) was used to calculate the realized relationship matrix **K** and estimate σ_u^2 and σ_e^2. Subscript obs refers to the subset of individuals for which there were no missing value at SNP k.

(CC) is the analysis restricted to complete cases (non-missing values) for a given marker tested, with the approximation of **Q** and **K** estimated from $\overline{\mathbf{X}}$ used as fixed, to account for population structure and relatedness. In (AD), marker effects are assessed from $\overline{\mathbf{X}}$, considered as fixed. In (MI) and (MI*), marker effects are assessed from $\dot{\mathbf{X}} = \dot{\mathbf{X}}^{(1)}, \ldots, \dot{\mathbf{X}}^{(m)}$ and the variability across imputes $\dot{\mathbf{X}}$ is accounted for. Relying on the MCAR assumption, (CC) served as a reference for assessing the consistency of estimates from (AD) or (MI): (AD) estimates departing too much from (CC) estimates were considered unreliable, especially when imputation uncertainty (γ ; see below) was high. Estimates from (MI) were expected to show appropriate control over imputation uncertainty, i.e. shrinkage towards zero as γ increases. In (MI*), marker effects are assessed from $\dot{\mathbf{X}}$ as in (MI), but variability across imputes for estimates of **Q** and **K** is also accounted for, through $\dot{\mathbf{Q}}$ and $\dot{\mathbf{K}}$, thus allowing for full control over imputation uncertainty in inferences regarding β. So while (CC), (AD) and (MI) are convenient to assess the impact of

imputation uncertainty on β estimates, (MI*) should be the method permitting the most statistically sound inferences regarding marker-trait associations.

In AP-4x, population structure and relatedness were respectively accounted for by 1 principal component and by matrix $\overline{\mathbf{K}}$ calculated on AP-4x individuals. In LP1 and LP2, no principal component and no genetic-relationship matrix were included in the GWAS model, then equivalent to a single-marker analysis model.

Combination of parameter estimates in MI and calculation of p-values: In (CC) and (AD), significance of marker-trait associations was assessed by performing an F-test for the regression coefficient estimate β, as described in Kang et al. (2008). In (MI) and (MI*), an F-test was performed in which the p-value was (Rubin and Schenker 1986; van Buuren 2012):

$$\Pr\left(F_{1,\nu} > \frac{\bar{\beta}^2}{T}\right)$$

where $\bar{\beta} = \frac{1}{m}\sum_{r=1}^{m}\hat{\beta}^{(r)}$, is the average estimate of β over imputations, with $\hat{\beta}^{(r)}$ the estimate of β based on the r^{th} imputation; $T = \bar{W} + B^*$ is the (estimated) total variance of β estimates, partitioned into $\bar{W} = \frac{1}{m}\sum_{r=1}^{m}\mathrm{Var}\left(\hat{\beta}^{(r)}\right) \doteq \mathrm{E}\left[\mathrm{Var}\left(\hat{\beta}|\dot{X}^{(r)}\right)\right]$, the average within-impute variance of β estimates, and $B^* = \frac{m+1}{m}\mathrm{Var}_r\left(\hat{\beta}^{(r)}\right) \doteq \mathrm{Var}\left(\mathrm{E}\left[\hat{\beta}|\dot{X}^{(r)}\right]\right)$, the among-impute variance of β estimates, corrected by $\frac{m+1}{m}$ for unbiasedness ($\doteq$ means "asymptotically equal"). Barnard and Rubin (1999) derived a formula for the number of denominator degrees of freedom ν of the F-statistic:

$$\nu = \frac{\nu_m \nu_{obs}}{\nu_m + \nu_{obs}}$$

where $\nu_m = \frac{m-1}{\lambda^2}$, with $\lambda = \frac{B^*}{T}$; $\nu_{obs} = \frac{\nu_{com}+1}{\nu_{com}+3}\nu_{com}(1-\lambda)$, with $\nu_{com} = n - (2 + t)$ the number of degrees of freedom for the hypothetically complete dataset.

For a particular coefficient β, the uncertainty in its estimate due to imputation is characterized by γ, the fraction of information about β missing due to missingness:

$$\gamma = \frac{s + \frac{2}{\nu + 3}}{1 + s}$$

with $s = \frac{B^*}{\bar{W}}$ (Rubin 1987; van Buuren 2012). γ reflects the dependency of the inferences about β upon the imputation procedure. Though quite complex, the γ-statistic may simply be interpreted as λ, the proportion of among-impute variance in inferences, adjusted for a finite number of imputes (van Buuren 2012). In this study, $\gamma = 1.026\lambda$ ($R^2 = 1$). According to Li et al. (1991), $0 \leq \gamma < 0.2$, $0.2 \leq \gamma < 0.3$ and $0.3 \leq \gamma < 0.5$ indicate a "modest", "moderately large" and "high" missing-data problem, respectively.

False discovery rate and significance for marker-trait association: The p-values obtained either from (CC), (AD) or (MI) were transformed into adjusted false discovery rates (FDR) using the method of Storey and Tibshirani (2003). Marker-trait associations for which FDR < 0.1 in (CC) or (MI*) were deemed significant and considered for further analyses, though evidence from (MI*) was preferred, because (CC) relied on the approximation of **Q** and **K** estimated from $\overline{\mathbf{X}}$ used as fixed in the GWAS model and also because (CC) was considered more prone to false positives or false negatives due to smaller sample sizes, $n_{obs} < n$ (see Discussion section). The R package `qvalue` (Dabney and Storey 2013) was used to compute FDR.

RESULTS

Genotypic variability in the panels

Phenotypic traits: Traits were measured in varying numbers of years and locations (Table 1). Some traits were measured in only one location (Ds, TSC, St and Quality traits in AP, measured/predicted in Ithaca only) and/or only one year (Ds, only in 2009, and Quality traits, only in 2011). As result, they would show an inflated genotypic variance. Ds, St, LL, LW and all quality traits showed a much higher ratio of genotypic-to-phenotypic variance (H) in AP compared to LP1 and LP2. PH showed a higher H in LP1 and AP, compared to LP2. AD showed a higher H in LP2 and AP, compared to LP1 (Table 2).

Population structure: To analyze population structure, principal component analysis (PCA) was performed on the average-dosage matrix $\overline{\mathbf{X}}$. For the whole dataset (panels LP1, LP2 and AP combined), the first four principal components (PC) were deemed relevant for describing population structure, based on proportions of variance explained and grouping patterns on PCs (Figure 2). Grouping patterns in the whole dataset are consistent with race (first PC, distinguishing accessions that are *Phalaris arundinacea* from those that are *Phalaris caesia*), panel (second and third PCs) and geographical origin: Eastern Europe versus North America or Western Europe (fourth PC). The fifth and sixth PCs do not seem to reflect any grouping and explain a small proportion of the total genotypic variation, compared with the first four PCs.

PCA was also performed for four other subsets of the data: LP1, LP2, AP and AP-4x (AP panel with only *P. arundinacea* samples). The number of selected PCs for those subsets was 0, 0, 2 and 1, respectively (data not shown). The relevant PCs in AP were consistent with race and geographical origin, respectively, in the same way as in the whole data set. The relevant PC in AP-4x was consistent with geographical origin.

Analysis of imputation uncertainty

Consistency of PCs and genetic-relationship coefficients across imputes: Consistency of a variable in MI is defined as the correlation with the same variable in (AD), in absolute value, averaged over imputes. Such value aims at assessing to what extent it is appropriate to use **Q** and **K**, estimated from $\bar{\mathbf{X}}$, as fixed, in (CC), (AD) and (MI) association analyses. In the whole dataset, **K** estimates seem very consistent over imputes (consistency of 0.90 ± 0.0002; Table 3), but while estimates of $\mathbf{Q_1}$ (first PC) are quite consistent, the estimates at the subsequent PCs lose coherence across imputes, with the fourth PC having a consistency of only 0.043 (± 0.032). This result suggests that accounting for imputation uncertainty with respect to population structure variables – when those are estimated from marker data – is important, which implies that (MI*) should be the most appropriate type of analysis for assessing marker-trait associations. In the AP-4x subset, the consistency of the only relevant PC is 0.076 (± 0.067) and the consistency of genetic-relationship coefficients is 0.934 (± 0.0002).

Consistency across imputation schemes of marker effects and significance: *From complete-case (CC) to average-dosage (AD).* Under the MCAR assumption, the relationship between $\mathbf{X}_{obs}$ and the trait of interest adequately reflects the effect of markers in the whole sample. That is, for any marker-trait association, the estimate $\hat{\beta}_{CC}$ based on complete cases is unbiased with respect to $\hat{\beta}$, the estimate of β based on the hypothetically complete data set. It follows from the MCAR assumption that imputations are unbiased if estimates of marker effects from (AD), $\hat{\beta}_{AD}$, are unbiased with respect to those from (CC), $\hat{\beta}_{CC}$. The plot (Figure 3a) and the regression analysis (Table 4) of $\hat{\beta}_{AD}$ on $\hat{\beta}_{CC}$, across traits and markers, strongly suggest that there is no bias from (CC) to (AD). For marker-trait associations with low imputation uncertainty ($\gamma < 0.2$), there is a close relationship between $\hat{\beta}_{CC}$ and $\hat{\beta}_{AD}$ (R^2=0.94). However, for marker-trait associations with

moderate-to-high imputation uncertainty ($\gamma > 0.2$), large differences between $\hat{\beta}_{CC}$ and $\hat{\beta}_{AD}$ can be observed, though at random. Such differences indicate noise in inferred values of β and opportunities for false positives or false negatives (Figure 3b).

From average-dosage (AD) to multiple-imputation (MI). The purpose of MI is to make sound inferences when one bases their statistical analyses on imputed data. Therefore, MI should produce more conservative estimates $\hat{\beta}_{MI}$ ($\bar{\beta}$) when imputation uncertainty regarding marker-trait associations is high, in order to avoid declaring as significant associations that are due to "fortuitous" imputations. The plot (Figure 4a) and the regression analysis (Table 5) of $\hat{\beta}_{MI}$ on $\hat{\beta}_{AD}$, across traits and markers, show that, as γ increases, estimates of β tend to shrink towards zero, from (AD) to (MI). This behavior results in higher p-values (lower significance) for those associations in which imputation uncertainty is high: whether associations inferred in (AD) are true or false, their p-values will tend to fall above the significance threshold in (MI) if their among-impute variability is too large, whereas there is good agreement between (AD) and (MI) for low values of γ (Figure 4b). Only for $\gamma = 0$ (no variation among impute) there would be no adjustment on $\hat{\beta}_{MI}$ and $\mathrm{Var}(\hat{\beta}_{MI})$. In other words, one should rely on (AD) for inferences only if it can be assumed that the situation is "close enough" to the case where $\gamma = 0$. Clearly this is not the case for these data. In Figure S4, the consistency of marker effects from (CC) to (MI) is shown in.

From multiple-imputation (MI) to multiple-imputation with full account of imputation uncertainty (MI).* In (MI*), the marker-data matrix **X**, but also the principal-component and the genetic-relationship matrices **Q** and **K** (both estimated from **X**) are allowed to vary over imputes. While considering **Q** and **K** as fixed (estimated by $\overline{\mathbf{Q}}$ and $\overline{\mathbf{K}}$**)** was convenient for studying the

behavior of inferences from (CC) to (AD) to (MI), (MI*) should permit the safest inferences, by appraising the imputation uncertainty in the imputed genotypes, as well as in the resulting **Q** and **K** estimates. Figure 5a suggests concordance of β estimates from (MI) to (MI*), despite the relatively low consistency of **Q** estimates across imputes (Table 3). Also, marker effects with low imputation uncertainty seem less shrunk towards zero in (MI*), where the variability in **Q** and **K** estimates is accounted for. As a result, more significant associations could be detected in (MI*) than in (MI) (Figure 5b). Values of γ were quite consistent from (MI) to (MI*), with increased noise around perfect concordance for values of γ above 0.2 (Figure 6).

Imputation uncertainty and potential gains in power: As Figure 7a suggests, for $\gamma < 0.2$ (i.e., markers with a modest missing data problem) there were generally small differences in significance ($-\log_{10}(p)$) from (CC) to (MI*) but some opportunity for a few outstanding gains in significance, with increases of up to 8.08 in $-\log_{10}(p)$ (for the association between marker TP87762 and TSC; see next subsection and Table 7). Assuming that the associations detected are true positives, this increase in significance may be considered an increase in power. This higher sensitivity in GWAS would logically come from a higher effective sample size due to imputation, with at the same time a modest missing-data problem. For $\gamma > 0.2$ however, there was little benefit from imputation, with very few opportunities for higher sensitivity from (CC) to (MI*): most values for the differential in $-\log_{10}(p)$ are actually negative. Those results show again that the threshold of 0.2 for γ made a good guideline for setting apart inferences with excessively high imputation uncertainty.

Potential factors influencing imputation uncertainty: Because there is good consistency between γ and the apparent usefulness of imputation for increasing power, it would be helpful to identify the factors that allow the analyst to effectively predict γ and determine how likely

imputation is to generate gains in power. Figures 7b, 7c and 7d respectively show the marginal relationships between γ and proportion of missing values (PMV), MAF or the average mutual information (AMI) between given markers and all other markers in the data set. There seems to be a strong positive relationship between γ and PMV, with values of γ that are below 0.2 for markers that have up to about 25% of missing values. As PMV increases, the slope for γ decreases and the variability around the conditional mean, as determined by smoothing splines, increases. There seems to be a weak correlation between γ and MAF, with markers having high MAF showing slightly higher imputation uncertainty, especially for MAF < 0.1. However, this relationship might be an artifact from markers with low PMV which tended, by happenstance, to have lower MAF. There seems to be some (non-linear) relationship between γ and AMI, but it's not clear whether it is due to the fact that markers with very low AMI tended to have lower PMV on average: for AMI < 0.02, average PMV is 0.45 (standard deviation: 0.32); for AMI > 0.02 average PMV is 0.80 (standard deviation: 0.071). A simple additive model fitted to the data, with $\arcsin(\gamma)$ regressed on PMV, MAF and AMI, suggests that that all three factors considered have a significant effect on imputation uncertainty: MAF (usually equivalent to marker-genotype variance) and – to a much larger extent – PMV generate imputation uncertainty, while AMI reduces imputation noise (Table 6). The model fitted has some predictive value (r^2_{CV}, prediction reliability in 10-fold cross-validation, was 0.21), but a rather large part of the variation could not be accounted for (Table 6).

Association analyses

Significance of marker-trait associations: Significant associations, for which FDR < 0.1 in (CC) or (MI*), were detected in the AP-4x subset and in the LP2 subset; no significant association was detected in LP1. These involved nine markers (thereafter "significant markers"):

TP140584, TP184396, TP191264, TP217634, TP268059, TP341988, TP477925, TP521945 and TP87762. The one association involving TP341988 was detected only in LP2 (TP341988 was not included in the analysis in AP-4x due to the threshold of MAF > 0.05) while the others were detected only in AP-4x. As shown in Table 7, behavior of inferences from (CC) to (MI*) was highly dependent on γ: for values of γ above 0.3 (associations involving TP184396, TP191264, TP217634, TP268059, TP341988 and TP521945), marker effects estimated from (MI*) were closer to zero and significance of associations decreased, certainly as a result of the shrinkage of marker effects as well as the high among-impute variance. These marker-trait associations that lost significance from (CC) to (MI*) had generally a high PMV (above 0.69). On the other hand, associations with $\gamma < 0.3$ (involving TP140584, TP477925 and TP87762), characterized by higher PMV (below 0.34), were more prone to show higher significance from (CC) to (MI*), with similar (sometimes larger) estimated marker effects. In some cases, such as TP140584-TSC, TP477925-ARA, -Mg, -Ds, -P, -TSC and TP87762-ARA and -Mg, associations were deemed significant in (MI*) but not in (CC). Such behavior suggests an increase in detection power despite the uncertainty associated with imputation. Those associations generally show consistency from (CC) to (MI*), which would bring evidence for unbiasedness of the imputation procedure. However, for associations TP477925-Mn ($\gamma = 0.17$) and –Cu ($\gamma = 0.09$), marker effects were estimated to be weaker in (MI*). Under the assumption that the imputation procedure is proper, such results suggest that the associations detected in (CC) were actually false positives.

The good overall concordance between observed and expected quantiles for (CC) suggests that potential confounders (due to population structure and relatedness in AP-4x) were well accounted for in the GWAS models (Figure 8). As to (MI*), the strong overall deflation of

$-\log_{10}(p)$ suggests that p-values simply do not follow a $Uniform(0,1)$ distribution under the null hypothesis (no significant marker-trait association), because of the extra variability due to imputation uncertainty, which was large for the majority of markers (Figure 6).

Correlation among significant markers: In the absence of haplotypic information, the squared correlation between markers' allelic dosages (r^2) was used to reflect linkage disequilibrium (LD) between markers. r^2 values were calculated based on $\overline{\mathbf{X}}$ from MI-CART. Among the nine significant markers, there seem to be one group of four markers that are in moderately strong association with each other (TP140584, TP217634, TP477925 and TP87762) and one group of two markers in mild association (TP521945 and TP268059) (Figure 9). This grouping is consistent with the associations inferred from GWAS (i.e. markers in one group tend to be associated to similar traits; Table 7), except for TP217634 which is not associated to the same traits as TP140584, TP477925 or TP87762.

Homology of SNP 64-bp sequences with the Brachypodium distachyon genome: For each marker, a 64-bp read containing the corresponding SNP was obtained from the UNEAK pipeline. The reads corresponding to the nine significant markers were analyzed for homology with the *Brachypodium distachyon* genome (v1.0) (Table 8). Three of the nine marker reads significantly match regions of the *Brachypodium distachyon* genome: (i) TP184396, shown to be negatively associated to Lignin in AP (the non-reference allele has a negative effect on the trait), is in a region homologous to the last intron of a Arginine-tRNA ligase gene (e-value = 1.2E-6); (ii) TP268059, shown to be negatively and positively associated to CP and XYL_Eff, respectively, is in a region homologous to the only exon in a phenylalanine/histidine-ammonia-lyase gene (e-value = 3.4E-12); (iii) TP341988, shown to be negatively associated to GLC in LP2, is in a region homologous to the second exon of a translation elongation factor G gene (e-value = 4.3E-

19). The moderate significance of TP184396's alignment to the *B. distachyon* genome may be due to the fact that the marker is located in an intron, which is usually more likely than exons to be under low selection pressure and diverge substantially across species.

DISCUSSION

Genotype calling uncertainty

Genotype calling uncertainty in GBS typically arises from (i) sequencing error, which generates miscalls in SNP alleles; and (ii) low sequencing depth, which causes heterozygotes to be miscalled as homozygotes. Genotype calling uncertainty may be accounted for by applying cut-offs on proportion of reads (e.g., calling individuals homozygous for a SNP if >80% of their reads is of a particular allele) or probabilistic methods (returning posterior probabilities of genotypes), which rely on some estimates of error rate and population allele frequency at SNPs (Nielsen et al. 2011). In this study, sequencing depth was very low; there were on average 0.29 reads per SNP called (standard deviation, across SNPs: 0.38). As a result, none of the above-mentioned methods could be used conveniently (probabilistic methods were not used because of the high amounts of missing values, particularly prohibitive for estimating population allele frequencies). Genotype calling uncertainty was therefore not accounted for, and the genotypes were used as returned by the UNEAK pipeline in the imputation procedure and association tests. Because genotype calling uncertainty translates into random error under a GWAS model, not accounting for it typically results in shrunk marker effects, and loss of power (Nielsen et al. 2011). But because there was presumably no systematic loss in precision (true discovery rate), any GWAS inference made from (CC) or (MI*) was still deemed valid in this study. Nonetheless, false positives may have arisen at random from noise (due to genotype-calling uncertainty or other sources), particularly in (CC) which usually had low sample size n_{obs}.

The use of tree models for imputation

Here, no reference genome or genetic map was available for imputing the GBS marker data Although it was possible to generate genetic maps on LP1 and LP2, the proportion of missing values was such that (i) in both populations, parental SNP genotypes could not been determined clearly (in particular, 37.6% and 25.6% of selected markers had missing values at both parents in LP1 and LP2, respectively), and (ii) imputation uncertainty would have to be accounted to produce reliable genetic maps, which would have been a serious statistical and computational challenge. Consequently, imputation methods based on HMM, which can be very accurate, were not used. In presence of unordered marker data with a general pattern of missingness and no reference panel, the strategy described here to impute missing values (i.e., FCS based on tree models), should be pertinent: though computationally intensive, the proposed imputation procedure was flexible, easy to implement, and appropriate for modelling marker data, as suggested by Dai et al. (2006).

Structure in the panels

Analysis of structure in the three panels combined revealed stratification by panels, race and geographical origin. The observed stratification by race and geographical origin was consistent with the results from Jakubowski et al. (2011), i.e. *P. caesia* being genetically distinct from *P. arundinacea* and, within *P. arundinacea*, East-European strains being distinct from West-European and North-American strains. Even though there would exist accessions native to North America with a genetic background distinct from that of European accessions (Jakubowski et al. 2012), most accessions of reed canarygrass found in North America, including all those evaluated in this study, share some common ancestry with West-European accessions.

Results of imputation-based association tests

In this association study, for the rare cases where γ was low enough ($\gamma < 0.30$), there were gains of significance from (CC) to (MI*) (and, presumably, a gain of power; Figure 7a, Table 7). Another interesting outcome from MI was the decreased significance, given one marker (TP477925), for only a subset of the associations detected in (CC) (Table 7). Such results suggest a possible gain of precision in MI-based association tests. Unfortunately, no novel significant markers could be detected from (CC) to (MI*). As expected, for high values of γ ($\gamma > 0.30$), MI had the desired property of decreasing significance of associations, for the sake of precision.

Values of γ lower on average than 0.3 would correspond to PMV lower than about 0.45, according to the model presented in Table 6, with values of MAF and AMI set to 0.1 and 0.05, respectively. PMV lower than 0.45 corresponded to 1.6%, 1.8% and 1.6% of the markers considered for GWAS, in the LP1, LP2 and AP-4x subsets, respectively. As discussed below, had a reference panel been available for imputation, there would have been much more opportunities for gains in power from MI.

GWAS results and similarity of marker sequences to *Brachypodium distachyon*

This GWAS revealed associations of markers with multiple traits, which may indicate pleiotropy of one single tagged causal variant, LD between distinct causal variants affecting different trait, or genetic correlation among traits. For example, TP87762 was negatively associated to TSC and positively associated to Ds, St, ARA and Mg. This result would probably suggest that, through lower tiller density, reed canarygrass plants carrying the non-reference allele at TP87762 were less prone to disease and lodging and had higher concentration of Arabinose and Magnesium.

Significant markers could not be mapped to a particular region of the reed canarygrass genome, because no reference map or genome sequence is available for that species. But the significant markers identified here may be used in further studies making use of DNA sequences in reed canarygrass. Potentially, our studies may bring in some insight about the function of genes in related species such as Brachypodium or oat. However, actual causal genes were probably not directly tagged by significant markers. Markers that did not map to the Brachypodium genome (TP140584, TP191264, TP217634, TP477925, TP521945 and TP87762; because of absence of homology or because of evolutionary divergence) were probably in LD with functional regions, which maybe could have been mapped to the Brachypodium genome. TP184396 and TP 341988 could be mapped to the Brachypodium genome, but within genes with very general purposes (both matched genes involved in mRNA translation). Such results could be "true hits", but this doesn't seem very likely. On the other hand, TP268059, negatively associated to CP and positively associated to XYL_Eff, was mapped to the exon of a gene coding for Phenylalanine/Histidine-Ammonia-Lyase. Although it is entirely possible that TP268059 did not directly tag the true causal gene, it is plausible that TP268059 is a true hit: Phenylalanine-Ammonia-Lyase (PAL) is involved in the very first step of the monolignol biosynthetic pathway leading to the synthesis of lignin (Boerjan et al. 2003), and a decreased ability to synthesize lignin is consistent with lower crude protein content and higher xylose conversion efficiency, since lignin inhibits fermentation (Vogel and Jung 2001). However, one would then expect to find significant associations involving Lignin (lignin content) and GLC_Eff (glucose conversion efficiency), which was not the case here. That said, the association between TP268059 and lignification still makes good sense from the homology of the marker in the Brachypodium genome and the relative directions of the detected effects.

A genome-wide association study based on multivariate MI

Historically, MI has been prevalent in epidemiological studies, probably because of the frequent and prohibitive occurrence of nonresponse in survey data (Rubin 1987; Rubin 1996; Klebanoff and Cole 2008; Sterne et al. 2009). Here, the large amount of missing values in our GBS data motivated us to use MI in order to account for imputation uncertainty and make sound inferences about marker-trait associations. There have been previous quantitative genetics studies which used MI to increase the precision of significance tests for marker-associated effects: Dai et al. (2006) used an imputation procedure similar to that presented here and compared it with expectation-maximization techniques for imputation accuracy. However, their study dealt with small marker data (10 SNPs in real data and 4 SNPs in simulated data) and limited PMV (10% or 20% of missing data). Bobb et al. (2011) used MI to gain precision in QTL mapping, but they generated multiple imputes of the phenotypic data, not the genomic data. To our knowledge, this study is the first report of MI applied in a genome-wide context (with thousands of markers over the genome). We believe the methodology presented here could be useful in large-scale genetic analyses involving hypothesis testing, in two ways. First, it exemplifies the use of tree models to impute unordered-marker data in GWAS and, in the context of multivariate MI, approximate the distribution $p(\widehat{\mathbf{X}}_{mis}|\mathbf{X}_{obs})$, by following the methodologies developed by Dai et al. (2006) and Burgette and Reiter (2010). As multiplexed GBS, which trades sequencing depth (and therefore genotyping costs) for uncertainty in genotype calling and imputation, is increasingly used for association mapping (Poland and Rife 2012), this type of imputation procedures in a MI context should be particularly useful in species where no or only part of a reference genome is available, like wheat or switchgrass. Second, it shows how estimates from different imputes about marker effects in the unified linear mixed model (Yu et al. 2006) could be pooled, using the rules

developed by Rubin (1987), to conveniently account for imputation uncertainty and perform statistically valid association tests. That said, MI is not the only method that has been developed to explicitly account for imputation uncertainty. A basic approach would be to use expected genotype counts, based on some distribution $p(\hat{\mathbf{X}}_{mis}|\mathbf{X}_{obs})$, and perform association tests as if the marker data was fixed, as was done in (AD) here (Guan and Stephens 2008; Zheng et al. 2011). As noted previously (Guan and Stephens 2008), such simplification may yield erratic behavior in association testing if imputation uncertainty is high *and* marker effects are large, which is consistent with the results obtained here (Figure 3b). A more sophisticated approach would be to use tests with explicit account for increased variance of parameter estimates due to imputation uncertainty, that are based on some approximation of $p(\hat{\mathbf{X}}_{mis}|\mathbf{X}_{obs})$ (Marchini et al. 2007; Guan and Stephens 2008; Zheng et al. 2011). In this framework, frequentist tests include score tests and likelihood-ratio tests, implemented in SNPTEST (Marchini et al. 2007). The Bayesian counterparts of this type of tests, implemented in SNPTEST and BIMBAM (Servin and Stephens 2007), feature some interesting advantages compared to the frequentist tests: in particular, they avoid an inflation in significance arising from high imputation uncertainty (Guan and Stephens 2008). Unfortunately, both the frequentist and Bayesian methods described above, in the current state of their implementations, are limited in the type of models that can be fitted to the data: the tests do not apply to linear mixed models (accounting for relatedness through the **K** matrix), and uncertainty in covariates cannot be conveniently accounted for, as was done in (MI*) here. Though in human GWAS, such models may be appropriate (e.g., Burton et al. 2007), the strong population stratification in plant GWAS panels call for mixed models that incorporate information on population structure and relatedness (Zhu et al. 2008). We believe the approach

used here was particularly useful in that it allowed to account for imputation uncertainty when using linear mixed models for testing associations.

Properness of imputations

An imputation procedure is proper if (i) it is unbiased and (ii) it is confidence-valid, i.e., the variability among imputed values is equivalent to what it would be had the data been complete. Here, we made the MCAR assumption stating that no factor (in particular, not the phenotypes of interest) influenced missingness at the marker data (Rubin and Little 1987; Sterne et al. 2009). Because marker effect estimates based on complete cases are unbiased under the MCAR assumption, (CC) analyses served as the reference to assess consistency of imputation-based tests. Judging from regression analyses on marker-effect estimates from (AD) (Table 4) and from (MI) (Table 5), it seemed that imputations did not generate bias, with shrinkage in marker-effect estimates from MI occurring only because of imputation uncertainty in (MI) (and (MI*)), which is desirable. Imputation bias may have occurred if the MCAR assumption – more generally, the MAR assumption – had not been valid. This may occur in survey data but is unlikely to occur, systematically, in GBS data: the probability of a marker read being available is unlikely to vary systematically across individuals. Confidence-validity could only be assumed here. For other imputation methods, which do not preserve the variability present in the original dataset, this assumption cannot be made. Such methods, which include mean imputation and major-allele imputation, should therefore be avoided in inference studies.

In the future, a simulation study might be conducted to characterize the unbiasedness and confidence validity of our imputation procedure in a setting where both marker effects and missing values are known *a priori*. Similar analyses were already performed on SNP data (e.g.

Dai et al. 2006), but assessing the properness of MI based on GBS data in a genome-wide context could certainly be useful.

Practical issues for MI in GWAS

Here, given the amount of missing data (78% on average) and the information available for imputation (no reference panel and no genetic map of markers), imputation uncertainty was too prohibitive for detecting novel significant markers when performing imputation-based association tests. Gains of significance were nonetheless possible, mostly for low values of γ (Figure 7a), corresponding roughly to PMV < 0.45 (Figure 7b). Pasaniuc et al. (2012) showed that a very substantial gain in imputation accuracy on low-depth GBS data in human could be achieved when using a reference panel (the 1000-genomes data). For example with 0.1x sequencing depth, they reported an increase in imputation accuracy from about 0.05 to more than 0.70 when this reference panel was used. In light of the results from Pasaniuc et al. (2012) and this paper, we would recommend imputation-based association studies on GBS data with sporadically missing values only when coverage is high enough to produce few missing values *or* when a reference panel is available for imputation.

When dealing with marker datasets that are much larger than the one considered here (numbers of markers q in the order of 10^5 or 10^6), implementing MI in GWAS would be a challenge. Computational and memory requirements will of course increase and parallelization should be devised accordingly. When dealing with unordered marker data, MI based on CART should be a good choice with regard to computational – and statistical – efficiency. However, the `mice` package uses a $q \times q$ predictor adjacency matrix to keep record of the candidate predictors for each marker. With many threads and/or high q, memory usage can be reduced by replacing the adjacency matrix with a $q \times q_{selected}$ adjacency list ($q_{selected}$: number of variables selected as

potential predictors; e.g. 500 in this study) or determining the set of candidate predictors at each iteration with no storage, hence trading computational efficiency for memory efficiency. Both of these measures would involve modifications of the `mice` code. Importantly, if a (completely typed) reference panel **H** is available for imputation, one can base imputations of $\mathbf{X}_{mis}$ on **H** only, i.e. $p(\widehat{\mathbf{X}}_{mis}|\mathbf{X}_{obs}, \mathbf{H})$ becomes $p(\widehat{\mathbf{X}}_{mis}|\mathbf{H})$. In such setting, when imputing a given variable k, there would be no uncertainty about values at the predictors to account for (supposedly, genotypes or haplotypes in **H** have been perfectly called). As a result, the multiple imputed datasets in MI could be sampled by ordinary MC instead of MCMC. This would dramatically reduce the computational and memory requirements when implementing MI. Note also that, when imputing at markers that are completely untyped in **X** (*in silico* genotyping), basing imputations on $p(\widehat{\mathbf{X}}_{mis}|\mathbf{H})$ rather than $p(\widehat{\mathbf{X}}_{mis}|\mathbf{X}_{obs}, \mathbf{H})$ may actually yield more accurate imputations (Guan and Stephens 2008). Thus, in presence of a reference panel, MI should not only be more useful (imputations being more accurate), but also more tractable (probably applicable when dealing with hundreds of thousands of markers).

ACKNOWLEDGEMENTS

The authors thank two anonymous reviewers for remarks and suggestions which greatly helped with improving the manuscript. This research was supported by U.S. Department of Energy & Department of Agriculture Plant Feedstock Genomics for Bioenergy Program Project Number DE-A-102-07ER64454, by Agriculture and Food Research Initiative Competitive Grant No. 2011-68005-30411 from the USDA National Institute of Food and Agriculture (CenUSA), and by USDA-ARS Congressionally allocated funds. Mention of commercial products and organizations in this manuscript is solely to provide specific information. The USDA is an equal

opportunity provider and employer. GPR was supported by the Gabelman-Shippo Wisconsin Distinguished Graduate Fellowship at the University of Wisconsin–Madison.

REFERENCES

Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-3402.

Alway, F.J., 1931 Early trials and use of reed canary grass as a forage plant. *Agronomy Journal* 23: 64-66.

Asay, K., I. Carlson, and C. Wilsie, 1968 Genetic Variability in Forage Yield, Crude Protein Percentage, and Palatability in Reed Canarygrass, Phalaris arundinacea L. *Crop Science* 8: 568-571.

Baldini, R.M., 1995 Revision of the genus Phalaris L. (Gramineae). *Webbia* 49 (2):265-329.

Barnard, J., and D.B. Rubin, 1999 Miscellanea. Small-sample degrees of freedom with multiple imputation. *Biometrika* 86: 948-955.

Bittman, S., J. Waddington, B.E. Coulman, and S.G. Bonin, 1980 *Reed canarygrass - a production guide*. Agriculture Canada, Ottawa (Ontario, Canada).

Boateng, A., H. Jung, and P. Adler, 2006 Pyrolysis of energy crops including alfalfa stems, reed canarygrass, and eastern gamagrass. *Fuel* 85: 2450-2457.

Bobb, J.F., D.O. Scharfstein, M.J. Daniels, F.S. Collins, and S. Kelada, 2011 Multiple imputation of missing phenotype data for QTL mapping. *Statistical Applications in Genetics and Molecular Biology* 10: 1-27.

Boe, A., and D.L. Beck, 2008 Yield components of biomass in switchgrass. *Crop Science* 48: 1306-1311.

Boerjan, W., J. Ralph, and M. Baucher, 2003 Lignin biosynthesis. *Annual Review of Plant Biology* 54: 519-546.

Bouchenak-Khelladi, Y., N. Salamin, V. Savolainen, F. Forest, M.v.d. Bank *et al.*, 2008 Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47: 488-505.

Breiman, L., J. Friedman, C.J. Stone, and R.A. Olshen, 1984 *Classification and regression trees*: CRC press.

Brummer, E., C. Burras, M. Duffy, and K. Moore, 2000 Switchgrass production in Iowa: economic analysis, soil suitability, and varietal performance. *Iowa State University, Ames, Iowa*.

Burgette, L.F., and J.P. Reiter, 2010 Multiple imputation for missing data via sequential regression trees. *American Journal of Epidemiology* 172: 1070-1076.

Burton, P.R., D.G. Clayton, L.R. Cardon, N. Craddock, P. Deloukas *et al.*, 2007 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661-678.

Butler, D., B.R. Cullis, A. Gilmour, and B. Gogel, 2007 ASReml-R reference manual. *Brisbane: Queensland Department of Primary Industries and Fisheries*.

Carlson, I.T., R.N. Oram, and J. Surprenant, 1996 Reed canarygrass and other Phalaris species, pp. 569-604 in *Cool-season forage grasses*, edited by LE Moser, DR Buxton, and MD Casler. American Society of Agronomy Inc., Madison (WI, USA).

Casler, M., 2009 Genetics, breeding, and ecology of reed canarygrass. *Intl J Plant Breeding*.

Casler, M., M. Phillips, and A. Krohn, 2009a DNA polymorphisms reveal geographic races of reed canarygrass. *Crop Science* 49: 2139-2148.

Casler, M.D., J.H. Cherney, and E.C. Brummer, 2009b Biomass yield of naturalized populations and cultivars of reed canary grass. *BioEnergy Research* 2: 165-173.

Cherney, J., K. Johnson, J. Volenec, and K. Anliker, 1988 Chemical composition of herbaceous grass and legume species grown for maximum biomass production. *Biomass* 17: 215-238.

Cureton, P., P. Groenevelt, and R. McBride, 1991 Landfill leachate recirculation: effects on vegetation vigor and clay surface cover infiltration. *Journal of Environmental Quality* 20: 17-24.

Dabney, A., J.D. Storey, and G.R. Warnes, 2013 qvalue: Q-value estimation for false discovery rate control. R package version 1.34.0.

Dai, J.Y., I. Ruczinski, M. LeBlanc, and C. Kooperberg, 2006 Imputation methods to improve inference in SNP association studies. *Genetic Epidemiology* 30: 690-702.

Dien, B.S., H.-J.G. Jung, K.P. Vogel, M.D. Casler, J.F. Lamb *et al.*, 2006 Chemical composition and response to dilute-acid pretreatment and enzymatic saccharification of alfalfa, reed canarygrass, and switchgrass. *Biomass and Bioenergy* 30: 880-891.

Doove, L., S. Van Buuren, and E. Dusseldorp, 2014 Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics & Data Analysis* 72: 92-104.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* 6: e19379.

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

Gelman, A., and T.E. Raghunathan, 2001 Using conditional distributions for missing-data imputation. *Statistical Science* 3: 268-269.

Guan, Y., and M. Stephens, 2008 Practical issues in imputation-based association mapping. *PLoS Genetics* 4: e1000279.

Jaiswal, P., J. Ni, I. Yap, D. Ware, W. Spooner *et al.*, 2006 Gramene: a bird's eye view of cereal genomes. *Nucleic Acids Research* 34: D717-D723.

Jakubowski, A.R., M.D. Casler, and R.D. Jackson, 2013 Genetic evidence suggests a widespread distribution of native North American populations of reed canarygrass. *Biological Invasions* 15: 261-268.

Jakubowski, A.R., R.D. Jackson, R. Johnson, J. Hu, and M.D. Casler, 2012 Genetic diversity and population structure of Eurasian populations of reed canarygrass: cytotypes, cultivars, and interspecific hybrids. *Crop and Pasture Science* 62: 982-991.

Kang, H.M., J.H. Sul, S.K. Service, N.A. Zaitlen, S.-y. Kong *et al.*, 2010 Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics* 42: 348-354.

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

Klebanoff, M.A., and S.R. Cole, 2008 Use of multiple imputation in the epidemiologic literature. *American Journal of Epidemiology* 168: 355-357.

Li, K.-H., T.E. Raghunathan, and D.B. Rubin, 1991 Large-sample significance levels from multiply imputed data using moment-based statistics and an F reference distribution. *Journal of the American Statistical Association* 86: 1065-1073.

Liaw, A., and M. Wiener, 2002 Classification and regression by randomForest. *R News* 2: 18–22.

Lu, F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney *et al.*, 2013 Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics* 9: e1003215.

Marchini, J., B. Howie, S. Myers, G. McVean, and P. Donnelly, 2007 A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics* 39: 906-913.

McWilliam, J., and C. Neal-Smith, 1962 Tetraploid and hexaploid chromosome races of Phalaris arundinacea L. *Crop and Pasture Science* 13: 1-9.

Nielsen, R., J.S. Paul, A. Albrechtsen, and Y.S. Song, 2011 Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* 12: 443-451.

Olmstead, J., M.D. Casler, and E.C. Brummer, 2013 Genetic variability for biofuel traits in a circumglobal reed canarygrass collection. *Crop Science* 53: 524-531.

Pahkala, K., and M. Pihala, 2000 Different plant parts as raw material for fuel and pulp production. *Industrial Crops and Products* 11: 119-128.

Pasaniuc, B., N. Rohland, P.J. McLaren, K. Garimella, N. Zaitlen *et al.*, 2012 Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nature Genetics* 44: 631-635.

Picard, C.R., L.H. Fraser, and D. Steer, 2005 The interacting effects of temperature and plant community type on nutrient removal in wetland microcosms. *Bioresource Technology* 96: 1039-1047.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S. Wu *et al.*, 2012 Genomic selection in wheat breeding using genotyping-by-sequencing. *The Plant Genome* 5: 103-113.

Poland, J.A., and T.W. Rife, 2012 Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome* 5: 92-102.

Price, A.L., N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick *et al.*, 2006 Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38: 904-909.

Price, D.L., and M.D. Casler, 2014 Divergent selection for secondary traits in upland tetraploid switchgrass and effects on sward biomass yield. *BioEnergy Research* 7: 329-337.

Quintanar, A., S. Castroviejo, and P. Catalán, 2007 Phylogeny of the Tribe Avenae (Pooideae, Poaceae) inferred from plastid Trn-T and nuclear ITS sequences. *American Journal of Botany* 94: 1554–1569.

R Development Core Team, 2014 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (Austria).

Rice, J., and B. Pinkerton, 1993 Reed canarygrass survival under cyclic inundation. *Journal of Soil and Water Conservation* 48: 132-135.

Rubin, D. B., 1987 *Multiple imputation for nonresponse in surveys*. John Wiley and Sons, New York (NY, USA).

Rubin, D.B., 1996 Multiple imputation after 18+ years. *Journal of the American Statistical Association* 91: 473-489.

Rubin, D.B., and R.J. Little, 2002 Statistical analysis with missing data. *Hoboken, NJ: J Wiley & Sons*.

Rubin, D.B., and J.L. Schafer, 1990 Efficiently creating multiple imputations for incomplete multivariate normal data, pp. 88 in *Proceedings of the Statistical Computing Section of the American Statistical Association*.

Rubin, D.B., and N. Schenker, 1986 Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association* 81: 366-374.

Rutkoski, J.E., J. Poland, J.-L. Jannink, and M.E. Sorrells, 2013 Imputation of unordered markers and the impact on genomic selection accuracy. *G3: Genes| Genomes| Genetics* 3: 427-439.

Sanderson, M., R. Reed, S. McLaughlin, S. Wullschleger, B. Conger *et al.*, 1996 Switchgrass as a sustainable bioenergy crop. *Bioresource Technology* 56: 83-93.

Schafer, J.L., 2010 *Analysis of incomplete multivariate data*: CRC press.

Servin, B., and M. Stephens, 2007 Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genetics* 3: e114.

Shenk, J., and M. Westerhaus, 1991 Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop Science* 31: 469-474.

Sterne, J.A., I.R. White, J.B. Carlin, M. Spratt, P. Royston *et al.*, 2009 Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj* 338.

Storey, J.D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 100: 9440-9445.

Therneau, T.M., and E.J. Atkinson, 1997 An introduction to recursive partitioning using the RPART routines.

Tilley, J., and R. Terry, 1963 A two-stage technique for the in vitro digestion of forage crops. *Grass and Forage Science* 18: 104-111.

Troyanskaya, O., M. Cantor, G. Sherlock, P. Brown, T. Hastie *et al.*, 2001 Missing value estimation methods for DNA microarrays. *Bioinformatics* 17: 520-525.

US Department of Agriculture and US Department of Energy, 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, USA).

Van Buuren, S., 2007 Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research* 16: 219-242.

Van Buuren, S., 2012 *Flexible imputation of missing data*. CRC press.

Van Buuren, S., and K. Groothuis-Oudshoorn, 2011 MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software* 45.

Van Buuren, S., J.P. Brand, C.G.M. Groothuis-Oudshoorn, and D.B. Rubin, 2006 Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation* 76: 1049-1064.

Van Keulen, J., and B.A. Young, 1977 Evaluation of acid-insoluble ash as a natural marker in ruminant digestibility studies. *Journal of Animal Science* 44: 282-287.

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91: 4414-4423.

Vogel, K. P., and H.J.G. Jung, 2001 Genetic modification of herbaceous plants for feed and fuel. *Critical Reviews in Plant Sciences* 20: 15-49.

Vogel, K. P., B.S. Dien, H.G. Jung, M.D. Casler, S.D. Masterson, and R.B. Mitchell, 2011 Quantifying actual and theoretical ethanol yields for switchgrass strains using NIRS analyses. *BioEnergy Research* 4: 96-110.

Vogel, K. P., J.F. Pedersen, S.D. Masterson, and J.J. Toy, 1999 Evaluation of a filter bag system for NDF, ADF, and IVDMD forage analysis. *Crop Science* 39: 276–279

Wood, S.N., 2003 Thin-plate regression splines. *Journal of the Royal Statistical Society (B)* 65: 95-114.

Wrobel, C., B.E. Coulman, and D.L. Smith, 2009 The potential use of reed canarygrass (Phalaris arundinacea L.) as a biofuel crop. *Acta Agriculturae Scandinavica, Section B - Plant and Soil Science* 59: 1–18.

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

Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari *et al.*, 2010 Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* 42: 355–60.

Zheng, J., Y. Li, G.R. Abecasis, and P. Scheet, 2011 A comparison of approaches to account for uncertainty in analysis of imputed genotypes. *Genetic Epidemiology* 35: 102-110.

Zhu, C., M. Gore, E.S. Buckler, and J. Yu, 2008 Status and prospects of association mapping in plants. *The Plant Genome* 1: 5-20.

TABLE 1 - *Summary of field traits and quality traits*

Trait	Code	Unit	Environments	Comment
Disease	Ds	1-9	I09	Resistance to biotic stress
Standability	St	1-9	I09, I10	Resistance to lodging
Leaf Width†	LW	cm	I09, I11, A11	Components of leaf area
Leaf Length†	LL	cm	I09, I11	
Total Stem Count	TSC	Count	I10, I11	Biomass yield component
Plant Height†	PH	m	I09, I11, A11	Biomass yield predictive trait
Full Height	FH	m	I09, I11, A11	
Heading Date	HD	DOY	I10, I11, A11	Length of vegetative stage
Anthesis Date	AD	DOY	I10, I11, A11	
Dry matter percentage‡	DM	%		Positively correlated with maturation and, therefore, with cellulose and lignin content (Dien et al. 2006)
Neutral Detergent Fiber‡	NDF	%DM		Lignin + Celluloses + Hemicelluloses
Acid Detergent Fiber‡	ADF	%DM		Lignin + Cellulose
NDF Digestibility‡	NDFD	%		Digestible fraction of NDF in vitro (Tilley and Terry 1963)
Acid Insoluble Ash‡	AIA	%DM		Positively correlated to DM digestibility (Van Keulen and Young 1977)
Klason Lignin‡	Lignin	%DM		Inhibits cellulosic fermentation (Vogel and Jung 2001) Increases conversion efficiency in thermochemical processes (Boateng et al. 2006)
Crude Protein‡	CP	%DM		Protein + Non-protein nitrogen (excluding nitrate)
Ash content‡	ASH	%DM		Fouling of bioreactors and disposal costs (Brummer et al. 2002)
Calcium‡	Ca	%DM		
Chlorine‡	Cl	%DM		
Copper‡	Cu	µg.g-1	I11	
Iron‡	Fe	µg.g-1		
Potassium‡	K	%DM		
Magnesium‡	Mg	%DM		
Manganese‡	Mn	µg.g-1		
Sodium‡	Na	%DM		
Phosphorus‡	P	%DM		
Sulfur‡	S	%DM		
Zinc‡	Zn	µg.g-1		
Glucose‡	GLC	mg.g-1		
Galactose‡	GAL	mg.g-1		
Xylose‡	XYL	mg.g-1		
Arabinose‡	ARA	mg.g-1		
GLC conversion efficiency‡	GLC_Eff	%		Expected fraction, on a mass basis, transformed into ethanol
XYL conversion efficiency‡	XYL_Eff	%		
Energy content‡	BTU	Btu/kg		1 Btu ≈ 1055 J

Units - *DOY: Day of the year; %DM: Percentage of dry matter; Btu: British thermal unit.* ***Environments –*** *A: Arlington (WI, USA); I: Ithaca (NY, USA); the two digits refer to the year. †: the trait was also measured in A10 and I10, in the LP1 and PL2 panels only. ‡: the trait was also measured in A10, in the LP1 and PL2 panels only.*

TABLE 2 - *Summary statistics on the field traits (9) and quality traits (26) in each panel*

Trait	Mean			Total standard deviation			H		
	LP1	**LP2**	**AP**	**LP1**	**LP2**	**AP**	**LP1**	**LP2**	**AP**
Ds	7.0	6.3	4.8	2.0	2.6	2.7	0.49	0.71	0.71
TSC	97.5	122.6	69.1	61.5	54.2	75.8	0.33	0.33	0.30
St	6.7	6.1	6.5	1.6	1.8	1.9	0.19†	0.10†	0.66
LW	21.0	18.6	17.0	6.0	3.9	3.8	0.16	0.16	0.29
PH	109	109	99	35.8	32.4	42.9	0.31	0.06	0.22†
LL	224	229	243	67.3	67.4	85.3	0.11	0.14	0.30
FH	162	160	143	31.2	28.5	36.3	0.31	0.24	0.32
HD	153	153	153	7.2	6.5	6.5	0.30	0.33	0.38
AD	157	158	158	7.4	6.5	7.0	0.27	0.45	0.43
DM	95.9	95.8	95.8	0.37	0.36	0.32	0.10	0.12	0.49†
CP	13.6	13.4	8.7	7.3	7.3	2.0	0.09	0.03	0.49
ADF	38.9	38.5	40.5	5.2	5.3	2.9	0.08	0.03	0.71
NDF	64.8	64.6	65.8	5.6	5.5	3.9	0.02	0.09	0.79
Lignin	5.4	5.6	6.5	1.83	1.84	0.58	0.12	0.05	0.55
NDFD	57.5	57.3	48.3	14.4	13.6	5.3	0.18	0.09	0.59
ASH	8.0	7.9	6.8	2.6	3.4	0.9	0.09	0.06	0.37
AIA	3.7	3.9	4.0	1.14	1.07	0.88	0.12	0.08	0.43
Ca	0.33	0.33	0.26	0.14	0.14	0.04	0.14	0.00	0.30†
P	0.24	0.25	0.2	0.151	0.07	0.031	0.11	0.01	0.55
Mg	0.24	0.25	0.18	0.162	0.151	0.058	0.04	0.05	0.48
K	1.9	1.9	1.6	0.93	0.35	0.35	0.26	0.18	0.69
Na	0.013	0.012	0.012	0.001	0.002	0.001	0.00†	0.00	0.45
S	0.25	0.24	0.20	0.20	0.07	0.05	0.20	0.02	0.48
Cl	0.69	0.67	0.62	0.32	0.26	0.18	0.00	0.03	0.38
Fe	122	163	156	69.9	69.5	92.1	0.17†	0.00	0.51
Mn	70.7	68.9	68.1	52.2	11.8	15.9	0.14	0.00	0.46
Zn	27.1	28.3	27.0	3.2	3.1	2.4	0.00†	0.00	0.52
Cu	4.9	4.3	3.7	4.2	3.7	1.9	0.21	0.19	0.51
BTU	7790	7823	7748	163.1	170.1	93.8	0.10	0.10	0.39
GLU	309	312	324	77.4	36.7	17.5	0.06	0.03	0.58†
XYL	216	216	216	6.4	7.2	7.7	0.00	0.05	0.72
ARA	34.7	33.4	33.3	6.6	6.3	3.6	0.22	0.05	0.54
GAL	26.2	26.1	23.6	6.4	4.9	1.6	0.16	0.03	0.43
GLC Eff.	55.6	56.4	48.4	11.4	12.6	5.9	0.05	0.01	0.41†
XYL Eff.	59.6	59.6	68.2	9.4	8.9	9.7	0.17	0.21	0.70

H*: Broad-sense heritability (in LP1 and LP2) or proportion of genotypic variance to phenotypic variance (in AP), calculated on an individual-plant basis, as an estimate of* $\frac{\sigma_G^2}{\sigma_G^2+\sigma_{GE}^2+\sigma_e^2}$*, with* σ_G^2 *the genotypic variance,* σ_{GE}^2 *the variance of genotype-by-environment interactions and* σ_e^2 *the residual variance, from the model described in subsection "Phenotypic data" in "Material and Methods", but fitted within each panel separately and with* $genotype$ *as a random effect. †: the spatial-correlation model did not converge for the corresponding trait and panel. If the model did not converge, a simpler model, not including the* $plot(year(location))$ *effect, was fitted.*

TABLE 3 – *Consistency of population-structure and genetic-relationship variables across imputes*

Variable	Average correlation with (AD) estimate	Standard deviation across imputes
$\dot{\mathbf{Q}}_{\mathbf{1}}$	0.74	0.026
$\dot{\mathbf{Q}}_{\mathbf{2}}$	0.48	0.088
$\dot{\mathbf{Q}}_{\mathbf{3}}$	0.21	0.100
$\dot{\mathbf{Q}}_{\mathbf{4}}$	0.04	0.032
$\dot{\mathbf{Q}}_{\mathbf{5}}$	0.05	0.039
$\dot{\mathbf{Q}}_{\mathbf{6}}$	0.05	0.034
$\dot{\mathbf{K}}$	0.90	0.0002

Consistency of confounder variables across imputes is reflected here by $avg_r\left\{\left|Cor\left(\dot{Q}_{ij}^{(r)}, \bar{Q}_{ij}\right)\right|\right\}$ *for principal component j (*$j = 1, \ldots, 6$*) and* $avg_r\left\{\left|Cor\left(\dot{K}_{ii'}^{(r)}, \bar{K}_{ii'}\right)\right|\right\}$ *for realized genetic relationship (with* (i, i') *a given pair of genotypes).* Cor *is the Pearson correlation.*

TABLE 4 – *Regression analysis of the relationship between $\hat{\beta}_{CC}$ and $\hat{\beta}_{AD}$*

Selection on γ	**Intercept**		**Slope**		
	Coefficient estimate (± Standard error)	**p-value**	**Coefficient estimate (± Standard error)**	**p-value**	**$\hat{\sigma}_e^2$ (R^2)**
None	0.0010 (± 0.0011)	0.34	**1.00** (± 0.0012)	< 0.0001	0.238 (0.76)
$\gamma < 0.2$	-0.0024 (±0.0048)	0.61	**0.99** (± 0.0053)	< 0.0001	**0.048** (0.94)
$\gamma > 0.2$	0.0011 (± 0.0011)	0.33	**1.00** (± 0.0012)	< 0.0001	**0.240** (0.76)

Model: $\hat{\beta}_{CC} = \text{Intercept} + \text{Slope}.\hat{\beta}_{AD} + e$. *The estimated regression coefficients suggest no systematic bias from $\hat{\beta}_{CC}$ to $\hat{\beta}_{AD}$ (Slope = 1). However, for associations with values of γ above 0.2, inferences tend to be more erratic, as indicated by the substantially higher residual variance ($\hat{\sigma}_e^2$) at $\gamma > 0.2$. The model meets the assumptions of linearity but not normality of residuals. Though p-values are not exact, they are provided for information. β estimates are from analyses on the AP-4x subset.*

TABLE 5 – *Regression analysis of the relationship between $\hat{\beta}_{AD}$ and $\hat{\beta}_{MI}$*

Effect	Coefficient estimate (± Standard error)	p-value	R^2
Intercept	-0.0048 (± 0.0028)	0.0943	0.91
b_1: $\widehat{\beta}_{MI}$	**1.0073** (± 0.0069)	< 0.0001	
b_2: γ	0.0097 (± 0.0061)	0.1152	
b_3: $\widehat{\beta}_{MI} \times \gamma$	**5.7864** (± 0.0162)	< 0.0001	

Model: $\hat{\beta}_{AD} = \text{Intercept} + (b_1 + b_3\gamma).\hat{\beta}_{MI} + b_2.\gamma + e$. *The estimated regression coefficients suggest no difference between $\hat{\beta}_{AD}$ and $\hat{\beta}_{MI}$ for associations for which $\gamma = 0$ ($b_1 = 1$), and shrinkage of $\hat{\beta}_{MI}$ towards 0 as γ increases ($b_3 > 0$). The model meets the assumptions of linearity but not normality of residuals. Though p-values are not exact, they are provided for information. β estimates are from analyses on the AP-4x subset.*

TABLE 6 - *Regression analysis on potential factors affecting imputation uncertainty*

Effect	**Coefficient estimate (± Standard error)**	***p*-value**	**r^2_{CV} (± standard deviation)**	**MSE (± standard deviation)**
Intercept	0.0934 (± 0.0018)	< 2.2e-16	0.21 (± 0.0064)	0.00687 (± 8.7e-05)
b_1: **PMV**	0.4806 (±0.0029)	< 2.2e-16		
b_2: **MAF**	0.0099 (±0.0025)	8.3e-05		[Null model: 0.00873 (± 1.3e-04)]
b_3: **AMI**	-0.1789 (± 0.0325)	3.6e-08		

Model: $\arcsin(\gamma) = \text{Intercept} + b_1.PMV + b_2.MAF + b_3.AMI + e$. *PMV: Proportion (between 0 and 1) of missing values; MAF: Minor alle frequency (between 0 and 1); AMI: Average mutual information (base 2).* r^2_{CV}*: squared coefficient of correlation in 10-fold cross-validation*; MSE*: mean squared error in 10-fold cross-validation. The null model is* $\arcsin(\gamma) = \text{Intercept} + e$. *The* arcsin *transformation was used to account for* $\gamma \in [0; 1]$. *The model meets the assumptions of linearity and normality of residuals, but not the assumption of homoscedasticity: variance of residuals tends to increase as* $\hat{\gamma}$ *increases.* γ*-statistics are from analyses on the AP-4x subset.*

TABLE 7 - *Results of association analyses based on (CC) and (MI*) procedures*

SNP	Trait	*p*-value (FDR)		$\widehat{\beta}$		γ	MAF	PMV
		(CC)	(MI*)	(CC)	(MI*)			
TP140584	**Ds**	2.9e-08 (0.00011)	2.9e-07 (0.00075)	1.6	1.3	0.24	0.065	0.34
	Mg	1.2e-05 (0.060)	7.8e-07 (0.0014)	0.022	0.022	0.28		
	TSC	0.00029 (0.38)	3.6e-07 (0.00063)	-17	-22	0.25		
TP184396	**Lignin**	3.2e-05 (0.082)	0.15 (1)	-0.18	-0.046	0.58	0.41	0.73
TP191264	**AD**	4.5e-08 (0.00023)	0.27 (1)	2.1	0.19	0.43	0.12	0.83
	HD	3e-06 (0.016)	0.47 (1)	1.8	0.12	0.38		
	K	1.4e-05 (0.036)	0.2 (1)	0.21	0.031	0.49		
	CP	3e-05 (0.052)	0.33 (1)	0.95	0.096	0.35		
	Lignin	3e-05 (0.082)	0.47 (1)	-0.32	-0.027	0.46		
TP217634	**K**	3.3e-06 (0.017)	0.0033 (1)	-0.26	-0.17	0.61	0.08	0.74
TP268059	**CP**	2.5e-05 (0.052)	0.095 (1)	-0.74	-0.2	0.46	0.14	0.76
	XYL_Eff	3.1e-05 (0.081)	0.1 (1)	2.9	0.78	0.43		
TP341988	**GLC†**	1.6e-05 (0.084)	0.034 (1)	-13	-5.6	0.5	0.11	0.69
TP477925	Mn	1.1e-06 (0.0053)	0.00011 (0.56)	5.6	3.7	0.17	0.067	0.22
	Cu	7.8e-06 (0.041)	0.00029 (1)	-0.68	-0.44	0.09		
	St	7.9e-06 (0.040)	5.9e-10 (3.1e-06)	-0.64	-0.73	0.085		
	PH	2.7e-05 (0.069)	1.1e-06 (0.006)	-5.8	-5.3	0.063		
	ARA	9.9e-05 (0.25)	7.6e-09 (4e-05)	1	1.3	0.16		
	Mg	0.00022 (0.45)	8.1e-08 (0.00021)	0.015	0.02	0.21		
	Ds	0.00041 (0.26)	3.3e-06 (0.0057)	0.84	0.9	0.049		
	P	0.0023 (0.96)	8.4e-07 (0.0044)	0.0073	0.0096	0.055		
	TSC	0.15 (0.97)	6.7e-08 (0.00018)	-5.3	-17	0.063		
TP521945	**CP**	4.2e-07 (0.0022)	0.015 (1)	1.4	0.47	0.39	0.074	0.78
	GAL	5.5e-07 (0.0028)	0.017 (1)	1.1	0.41	0.5		
	XYL_Eff	8.9e-07 (0.0046)	0.018 (1)	-5.3	-2	0.49		
	P	9.2e-07 (0.0047)	0.03 (1)	0.021	0.0075	0.49		
	ARA	6.3e-06 (0.032)	0.045 (1)	2.2	0.81	0.54		
	AD	1.1e-05 (0.029)	0.024 (1)	1.9	0.73	0.41		
	PH	2.7e-05 (0.069)	0.048 (1)	-10	-3.7	0.46		
	K	5.5e-05 (0.096)	0.027 (1)	0.23	0.089	0.34		
TP87762	Ds	4.5e-08 (0.00011)	8e-10 (4.2e-06)	1.9	1.9	0.027	0.055	0.015
	TSC	1.7e-05 (0.088)	1.4e-13 (7.5e-10)	-23	-38	0.044		
	St	1.8e-05 (0.046)	1.6e-08 (4.3e-05)	-0.88	-1	0.026		
	ARA	0.00097 (0.50)	2.1e-07 (0.00056)	1.3	1.8	0.037		
	Mg	0.0013 (0.94)	8.7e-09 (4.6e-05)	0.019	0.031	0.036		

All results presented are based on the AP-4x subset of the data (AP panel, with only P. arundinacea samples). †*: One exception is the association between TP341988 and GLC, for which the results presented are based on the LP2 panel (in the AP panel, TP341988 has MAF < 0.05). For each marker, the trait in bold is the one for which there is the strongest evidence for association.* $\widehat{\boldsymbol{\beta}}$*: Estimated additive effect of the non-reference allele relatively to the reference allele.* ***FDR****: False discovery rate, as in Storey and Tibshirani (2003);* ***MAF****: Minor allele frequency;*

***PMV**: Proportion of missing values. **$\boldsymbol{\gamma}$**: Imputation uncertainty. Red values of γ indicate associations for which there is a high missing-data problem ($\gamma > 0.30$). p-value (FDR) – (MI*): For the associations with $\gamma < 0.30$, colors indicate increases (green) or decreases (red) in significance and magnitude of marker effects, from (CC) to (MI*).*

TABLE 8 – *Sequence information on significant markers*

SNP	Associated trait(s)	Sequence read	Location in *Brachypodium distachyon* (Transcript ID – Annotation: Putative protein function)	%Identity [%Coverage] (*e*-value)
TP140584	**TSC**, **Ds**, Mg	CAGCCCGGCAGTTTGGTCTTGGGCAAGTATCTC CCCATTTCTTCCTCCATCAC**C/T**TAACAGAGAG	-	-
TP184396	**Lignin**	CAGCCTTATTCACCCACAATTC**C/T**AAAAGTTGT GCATAAATTTGCACGCTCCTAGTGCTCAACTC	Chr. 2: 34773626-34777668 (BRADI2G34680.1 – Intron 17: Arginine-tRNA ligase)	91% [68%] (1.2E-6)
TP191264	**AD**, HD, K, CP, Lignin	CAGCGACAAAACTCTCAAGGA**C/T**CACTCGTGAT TTAGGCAACCACCACAGCACTTAGCTGAAAAA	-	-
TP217634	**K**	CAGCGCGTTCTCCTTCCTTCCTGCAACCTCTAG TAGCCTCCCTGCAAATCAATCCGACGG**A/T**AAC	-	-
TP268059	**CP**, XYL_Eff	CAGCTCAGAGCAATACGAGGCCATGGCGATTTC **C/G**GCTCCCTTCAAGCCATAGTCCAAGCTCGGG	Chr. 3: 48837936-48837999 (BRADI3G47120.1.1 – Exon 1: Phenylalanine/Histidine-Ammonia-Lyase)	89% [100%] (3.4E-12)
TP341988	**GLC**	CTGCAATTTGGAA**A/T**GCAAGGACACTTGAATCA ACATCATGGTAGGAGCCATCAACCAGCACTGA	Chr. 5: 20217343-20221405 (BRADI5G16980.1 – Exon 2: Translation Elongation Factor G)	94% [98%] (4.3E-19)
TP477925	**St**, ARA, TSC, Mg, P, Ds, PH	CTGCGGATTC**A/C**ACCCTTACTAGGCGATAGCTC TGATCTATACCTTTCCTAGGAGAGACCACTTC	-	-
TP521945	**CP**, GAL, XYL_Eff, P, ARA, AD, PH, K	CTGCTC**C/T**TGCCGGCGTGGTGCGTGGCTCCCGT TGCCGCTGAAAAAAAAAAAAAAAAAAAAAAA	-	-
TP87762	**TSC**, Ds, St, Mg, ARA	CAGCATTACTAGAACGTGTATACGGTGCCATCT TCGAAATAGA**A/C**CCAGAACCTTCGATGTATGG	-	-

*Sequence reads are the 64-bp reads about the SNP marker, as returned from the UNEAK pipeline; in bold: <reference allele>/<alternate allele>. Homologous sequences are significant matches to the Brachypodium distachyon genome sequence (v1.0), found by BLAST (Altschul et al. 1997) in the Gramene database (*http://www.gramene.org*; Jaiswal et al. 2006).*

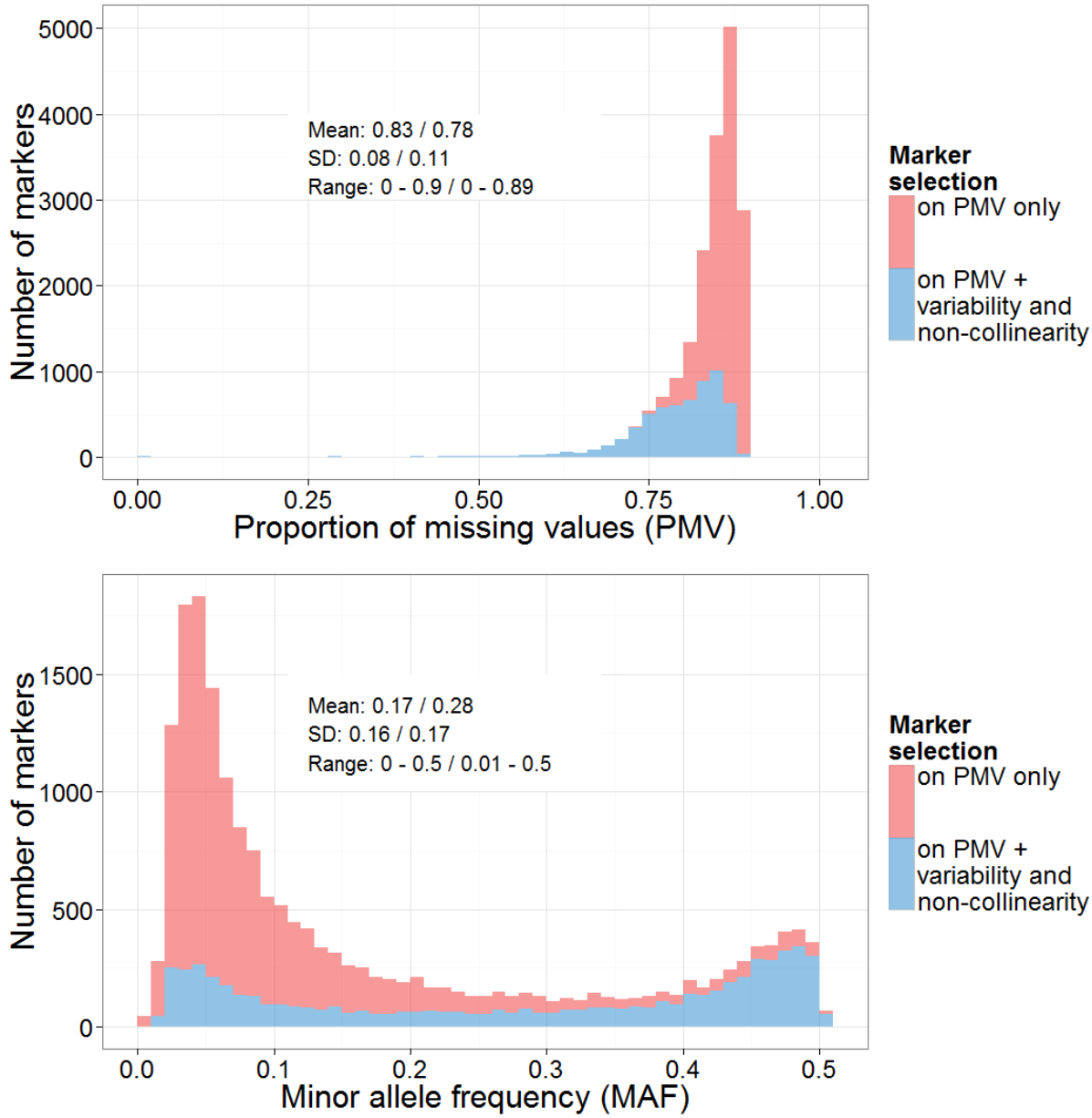

FIGURE 1 – *Summary statistics and distributions of proportions of missing values (PMV) and minor allele frequency (MAF), on markers retained after the first filtering step ("on PMV only", i.e. PMV ⩽ 0.9 by panel; 18,818 markers) or after both filtering steps ("on PMV + variability and non-collinearity", i.e. PMV ⩽ 0.9 by panel + filtering step recommended by van Buuren and Groothuis-Oudshoorn (2011) in which collinear and constant marker variables are discarded; 6,138 markers). Statistics are the mean, standard deviation (SD) and range for "on PMV only" / "on PMV + variability and non-collinearity". Filtering for variability and non-collinearity preferentially discarded markers with high PMV and low MAF.*

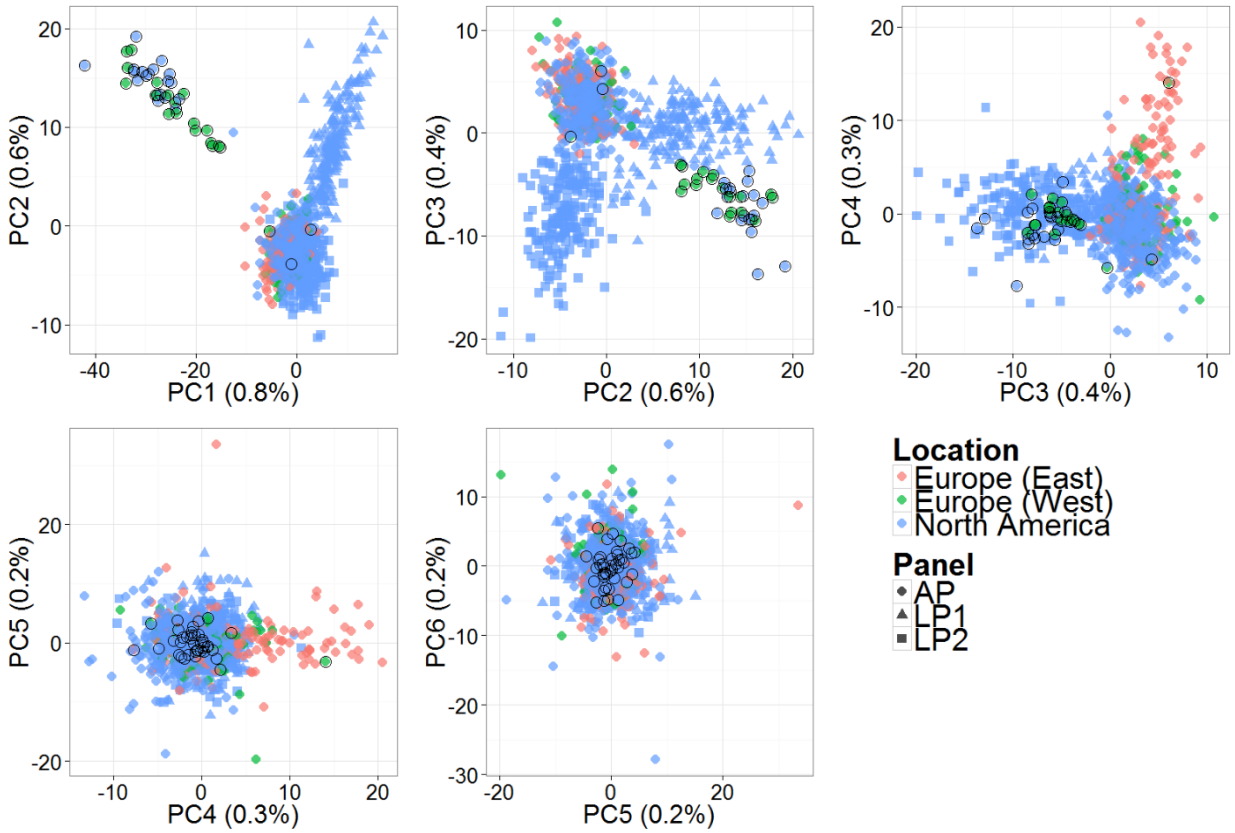

FIGURE 2 – *Principal Component Analysis (PCA) on the three panels combined: first six components and the proportion of marker variation explained, in parentheses on axis labels. Colors refer to location of origin (as in Table 1). Shapes refer to the panel (as in Table 1). Data points for P. caesia clones are circled in black.*

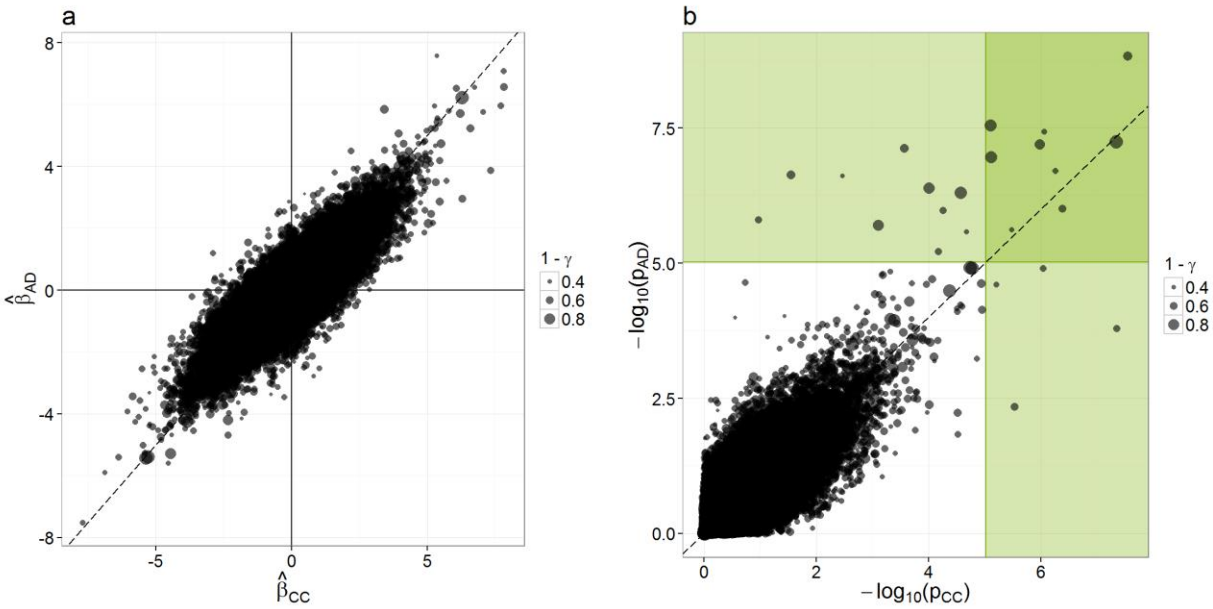

FIGURE 3 – *Concordance in inferences from (CC) to (AD) in the AP-4x subset of the data. The points' size refers to the stability of inferences across imputes* ($1 - \gamma$). ***(a)*** *Concordance across procedures and over all traits in marker effect estimates, standardized within each trait.* ***(b)*** *Concordance across procedures and over all traits in significance* ($-log_{10}(p)$)*; green areas correspond to significant associations according to a Bonferroni correction on a single-trait basis* ($p < 9.56 \times 10^{-6}$).

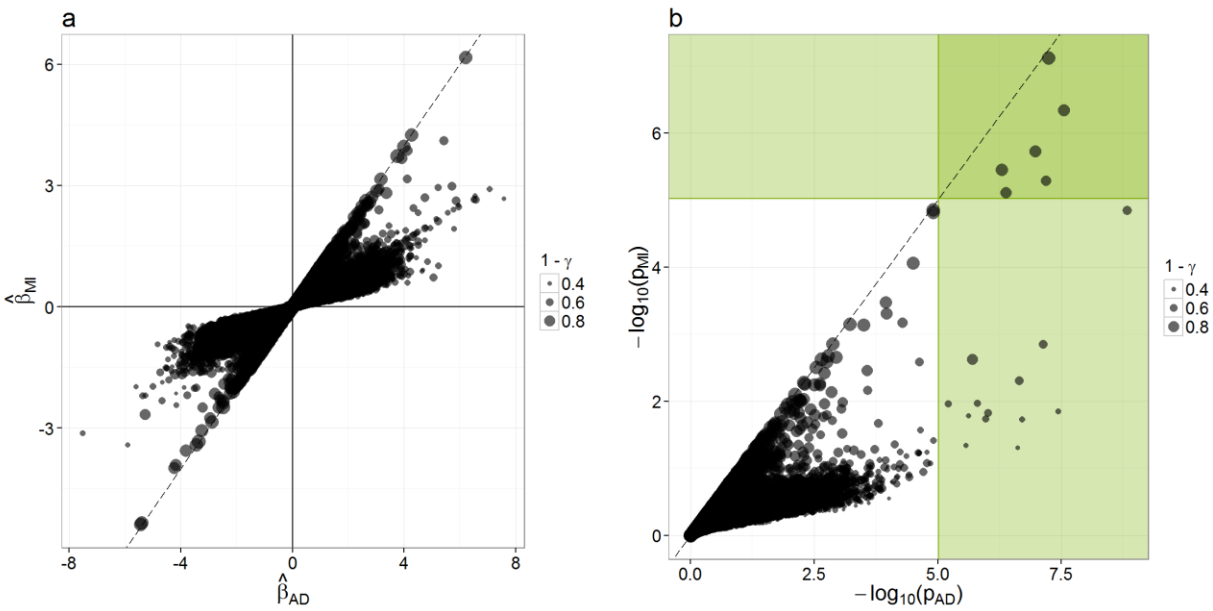

FIGURE 4 – *Concordance in inferences from (AD) to (MI) in the AP-4x subset of the data. The points' size refers to the stability of inferences across imputes* ($1-\gamma$). ***(a)*** *Concordance across procedures and over all traits in marker effect estimates, standardized within each trait.* ***(b)*** *Concordance across procedures and over all traits in significance* ($-log_{10}(p)$)*; green areas correspond to significant associations according to a Bonferroni correction on a single-trait basis* ($p < 9.56 \times 10^{-6}$).

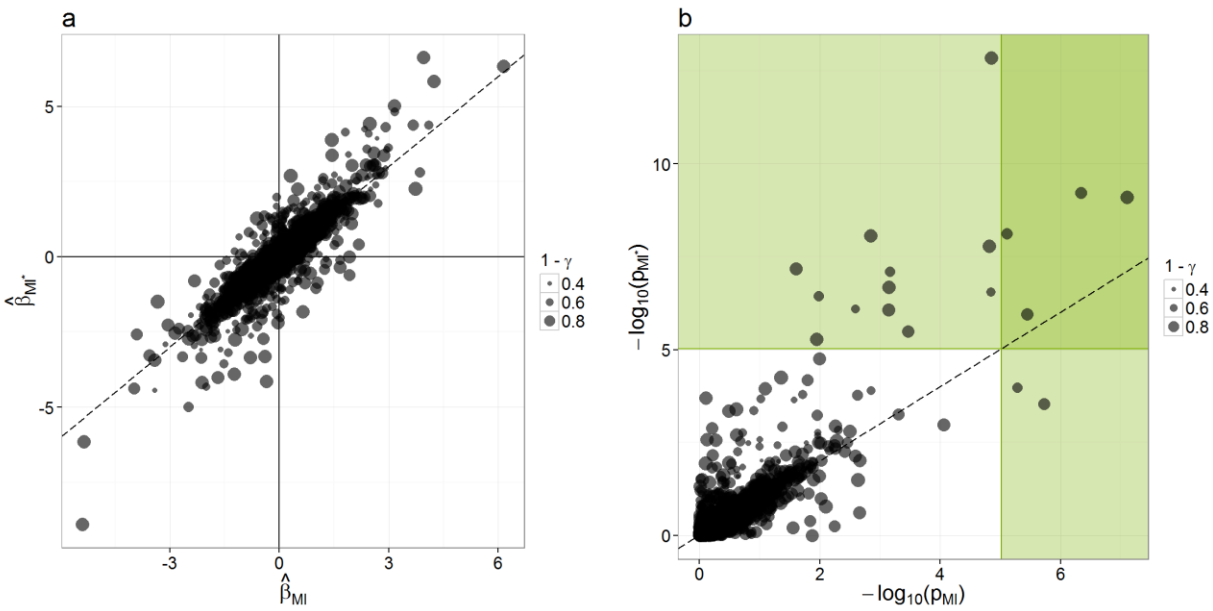

FIGURE 5 – *Concordance in inferences from (MI) to (MI*) in the AP-4x subset of the data. The points' size refers to the stability of inferences across imputes* ($1-\gamma$). ***(a)*** *Concordance across procedures and over all traits in marker effect estimates, standardized within each trait.* ***(b)*** *Concordance across procedures and over all traits in significance* ($-log_{10}(p)$)*; green areas correspond to significant associations according to a Bonferroni correction on a single-trait basis* ($p < 9.56 \times 10^{-6}$).

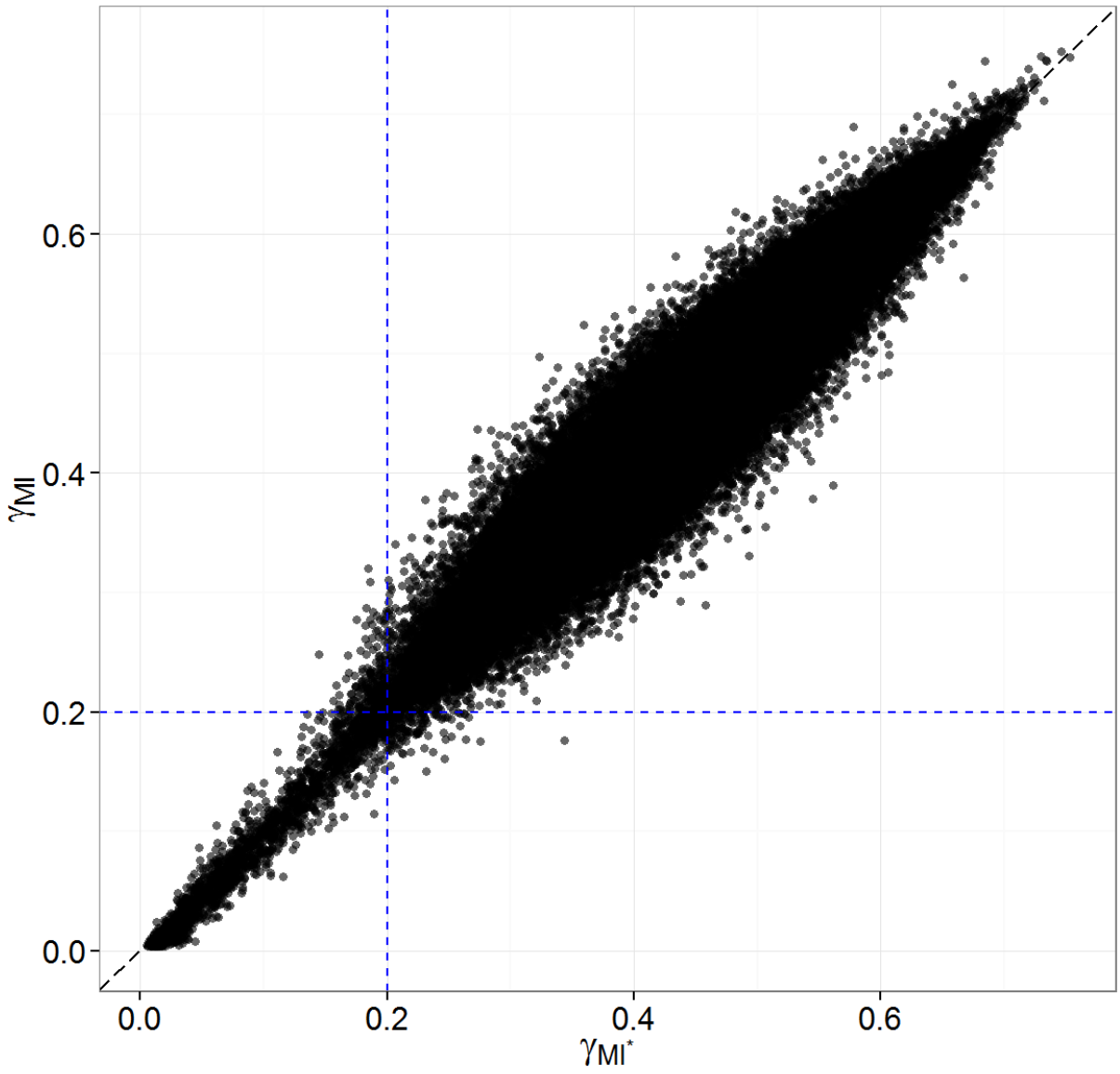

FIGURE 6 – *Concordance in imputation uncertainty, as reflected by the γ-statistic, over all traits, in the AP-4x subset, from (MI*) to (MI). The blue dashed lines correspond to the seemingly critical threshold of* $\gamma > 0.2$*, beyond which the γ-statistic loses coherence across procedures.*

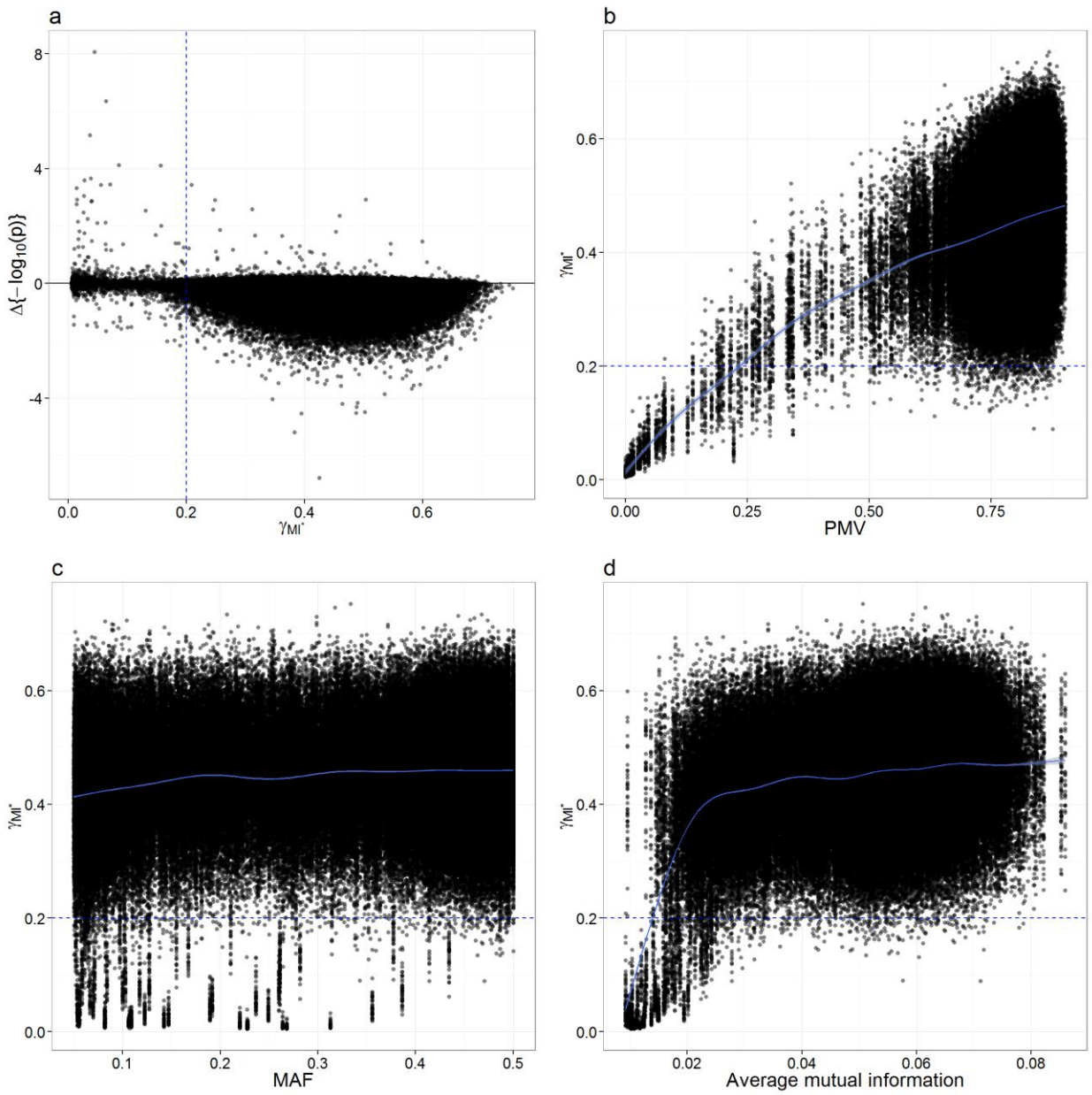

FIGURE 7 – ***(a)*** *Relationship between γ, in (MI*), and the difference in significance from (CC) to (MI*) ($\Delta\{-\log_{10}(p)\}$). Decrease in significance ($-\log_{10}(p)$) seems to occur more often and with higher intensity for $\gamma > 0.2$; presumably, there would to be more opportunities for gaining detection power with $\gamma < 0.2$.* ***(b-d)*** *Potential factors affecting imputation uncertainty (γ): relationship between γ, in (MI*), and* ***(b)*** *PMV (proportion of missing values),* ***(c)*** *MAF (minor allele frequency), and* ***(d)*** *the average mutual information between one given marker and all other markers in the dataset. In purple are the smoothed curves obtained from thin plate regression (default smoother in* `mgcv` *R package; Wood 2003). γ-statistics and p-values are from analyses on the AP-4x subset.*

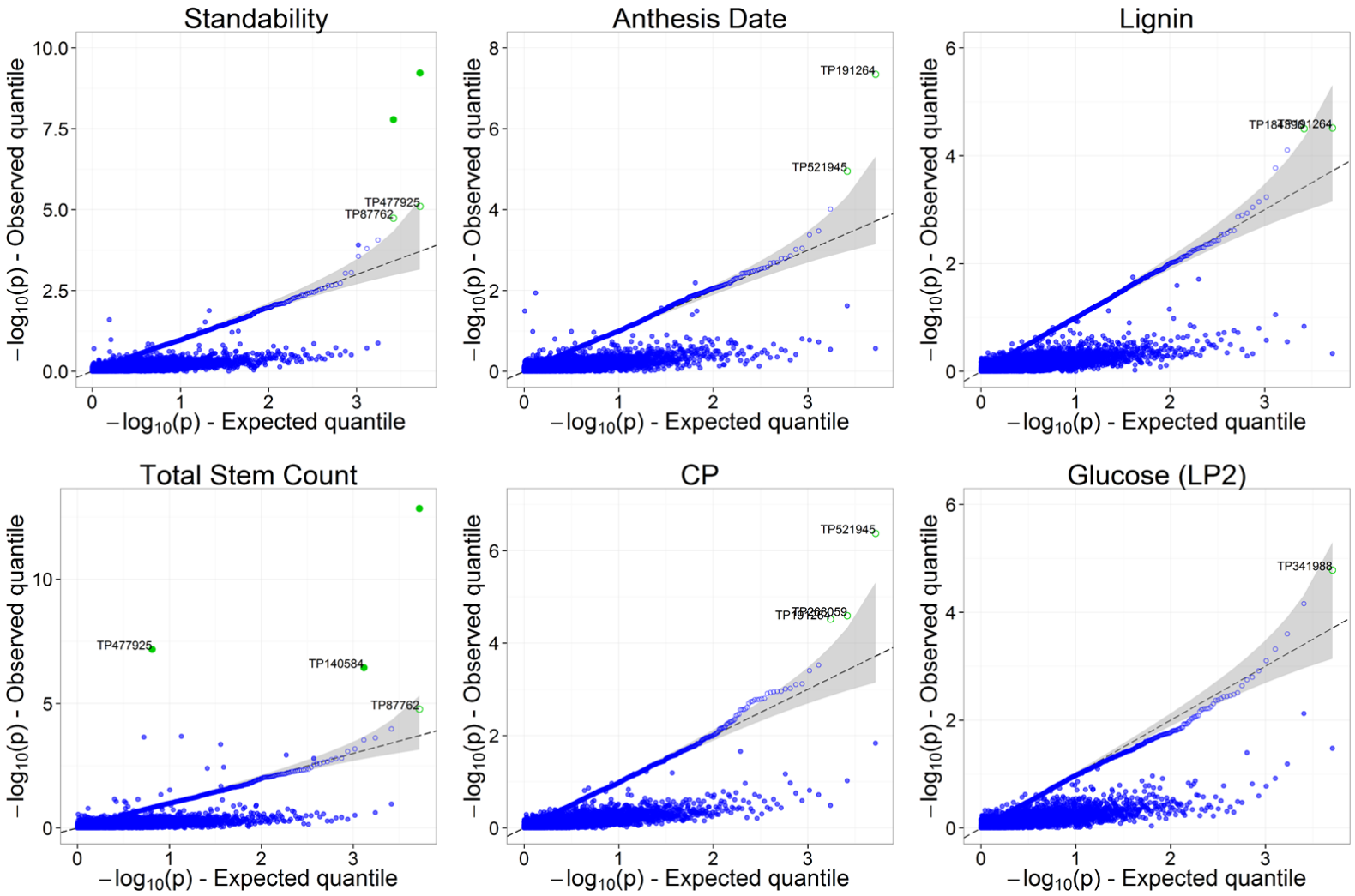

FIGURE 8 - *Q-Q plot of p-values from (CC), for St, TSC, AD, CP, Lignin (in AP-4x subset) and GLC (in LP2 subset), with corresponding p-values from (MI*). Expected quantiles are from the (CC) analysis; the corresponding values from (MI*) are shown unordered to reflect how consistent quantiles are, across*

analyses. Open circles: $-\log_{10}(p)$ *based on (CC); Full circles:* $-\log_{10}(p)$ *based on (MI*). Green symbols indicate associations for which FDR < 0.1. Grey areas correspond to the 95% confidence interval of quantiles under the null hypothesis that* p*-values follow a* $Uniform(0,1)$*.* p*-values from (CC) seem to follow a* $Uniform(0,1)$*, which indicate good control for potential confounders (population structure and relatedness). However,* p*-values from (MI*) do not follow a* $Uniform(0,1)$*, due to losses of significance caused by imputation uncertainty.*

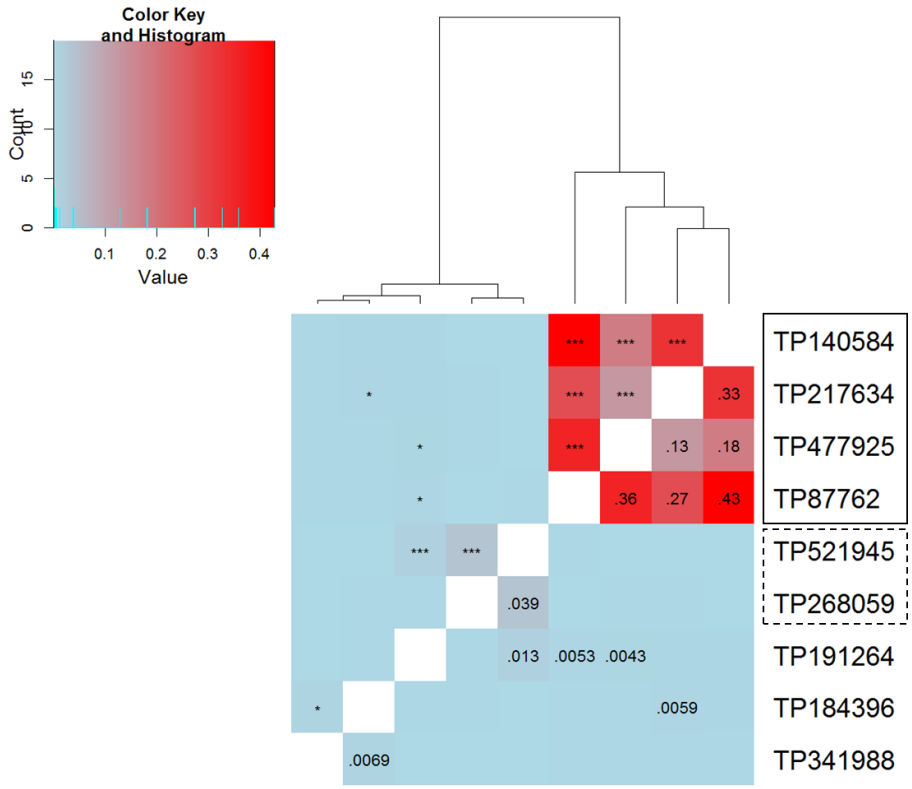

FIGURE 9 – *Squared coefficients of correlation among significant markers, based on* $\overline{\mathbf{X}}$ *(below-diagonal elements). Above-diagonal elements indicate significance; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. TP140584, TP217634, TP477925 and TP87762 form a group of fairly correlated marker loci. TP521945 and TP268059 are in very mild correlation with each other. TP191264, TP184396 and TP341988 each seem independent of other significant markers.*