Theor Appl Genet (2012) 125:1181–1194
DOI 10.1007/s00122-012-1905-8

ORIGINAL PAPER

Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects

Frank Technow · Christian Riedelsheimer · Tobias A. Schrag · Albrecht E. Melchinger

Received: 7 March 2012 / Accepted: 16 May 2012 / Published online: 26 June 2012

Abstract Identifying high performing hybrids is an essential part of every maize breeding program. Genomic prediction of maize hybrid performance allows to identify promising hybrids, when they themselves or other hybrids produced from their parents were not tested in field trials. Using simulations, we investigated the effects of marker density (10, 1, 0.3 marker per mega base pair, Mbp^{-1}), convergent or divergent parental populations, number of parents tested in other combinations (2, 1, 0), genetic model (including population-specific and/or dominance marker effects or not), and estimation method (GBLUP or BayesB) on the prediction accuracy. We based our simulations on marker genotypes of Central European flint and dent inbred lines from an ongoing maize breeding program. To simulate convergent or divergent parent populations, we generated phenotypes by assigning QTL to markers with similar or very different allele frequencies in both pools, respectively. Prediction accuracies increased with marker density and number of parents tested and were higher under divergent compared with convergent parental populations. Modeling marker effects as population-specific slightly improved prediction accuracy under lower marker densities (1 and 0.3 Mbp^{-1}). This indicated that modeling marker effects as population-specific will be most beneficial under low linkage disequilibrium. Incorporating dominance effects improved prediction accuracies considerably for convergent parent populations, where dominance results in major contributions of SCA effects to the genetic variance among inter-population hybrids. While the general trends regarding the effects of the aforementioned influence factors on prediction accuracy were similar for GBLUP and BayesB, the latter method produced significantly higher accuracies for models incorporating dominance.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1905-8) contains supplementary material, which is available to authorized users.

Communicated by M. Sillanpää.

F. Technow · C. Riedelsheimer · T. A. Schrag · A. E. Melchinger (✉)
Department of Applied Genetics, Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, 70599 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de

Introduction

While genetic progress in maize breeding is made through development of improved inbred lines, the main focus is on the F_1 hybrid progeny between two such lines, as the final, marketable product. Identifying high performing hybrids is therefore an integral part of every maize breeding program. However, because field evaluation of all potential hybrids is way too resource intensive, only a small subset can actually be tested in field trials.

Bernardo (1996) proposed best linear unbiased prediction (BLUP) for performance prediction of untested hybrids. This is achieved by exploiting the genetic covariance between tested and untested hybrids. The covariance can be estimated from pedigree (Bernardo 1996) or from molecular marker data (Maenhout et al. 2010). The latter approach can be seen in close analogy to genomewide BLUP (GBLUP), first proposed by Meuwissen et al. (2001) for estimation of marker effects for prediction of breeding values. Later, GBLUP was shown to be equivalent to traditional BLUP of breeding values when the pedigree-derived relationship matrix is replaced with one derived from marker data (Goddard 2009). Meuwissen et al. (2001) proposed two Bayesian methods, named Bayes A and B, as powerful alternatives to GBLUP. While their superiority over GBLUP for highly polygenic traits could so far not

conclusively be demonstrated, there is clear evidence for their advantage when the trait is controlled by a finite number of loci (Hayes et al. 2010; Clark et al. 2011; Meuwissen and Goddard 2010). Recently, Yang and Tempelman (2012) proposed improvements and extensions to Bayes A and B. The utility of these Bayesian methods was so far not investigated for hybrid prediction.

An important component of hybrid performance is the specific combining ability (SCA) between the parental lines of a hybrid. Thus, not only additive but also dominance effects of markers have to be estimated to account for the entire genetic variance. A further complication is that the parental lines in hybrid breeding are taken from genetically distant populations to maximize exploitation of heterosis. In Central Europe these are the "dent" and the "flint" populations, which were separated for more than 500 years (Stich et al. 2007). During this time, linkage between markers and QTL will have dissipated and possibly changed in sign (Charcosset and Essioux 1994) and QTL allele frequencies can have drifted into different directions. Hence, it might be necessary to model the marker effects as specific to a population.

Simulation studies have proven to be a powerful tool in comparisons of biometric models and methods for genomic prediction, because the true genotypic values, and the effects, positions and allele frequencies of the underlying QTL, as well as the LD between QTL and markers are known. This allows to investigate the effects of marker density and genetic architecture of traits as well as other factors relevant for genomic prediction. However, important idiosyncrasies of real-world data sets are hard to mirror in simulations. This often hampers the extrapolation of the results to the real world. Following the example of Zhong et al. (2009), we therefore simulated the QTL onto the observed marker profiles of existing genotypes from an actual maize breeding program.

Our objectives were to compare (i) biometric models that differ in inclusion of dominance and population-specific marker effects with regard to their performance for genomic predictions of hybrids, (ii) the utility of the marker effect estimation methods GBLUP and BayesB, (iii) the prediction accuracy for hybrids for which two, one or no parent(s) had been evaluated in other hybrid combinations, and (iv) the prediction accuracy on different levels of marker density and under convergent and divergent parental populations. For this, we used genomic data from an ongoing maize breeding program and simulated phenotypes.

Methods

Models

Model U_1 considers only additive effects, modeled as unspecific to a population,

$$\boldsymbol{y} = \mathbf{1}\mu + (\boldsymbol{Z}_d + \boldsymbol{Z}_f)\boldsymbol{u} + \boldsymbol{e}, \quad (1)$$

where μ is the intercept (and $\mathbf{1}$ a column vector of 1s), $\boldsymbol{y}$ is a $N \times 1$ ($N \equiv$ no. observations) vector of phenotypic hybrid entry means, adjusted for all other non-genetic effects, pertaining to the environment or field design. The vector of random effects of bi-allelic markers $\boldsymbol{u}$ is related to $\boldsymbol{y}$ through the known incidence matrix $(\boldsymbol{Z}_d + \boldsymbol{Z}_f)$ which has dimensions $N \times M$ ($M \equiv$ no. markers). The elements of the component matrices $\boldsymbol{Z}_d$ and $\boldsymbol{Z}_f$ code the presence ($z_{ij} = 1/2$) and absence ($z_{ij} = -1/2$) of the target allele in the gametes of the parental dent and flint inbred lines, respectively, of the corresponding hybrid. Which of the two alleles was chosen as the target allele was arbitrary, but the target allele was identical in dent and flint lines. Note that while $\boldsymbol{Z}_d$ and $\boldsymbol{Z}_f$ are defined based on the genotypes of the parental gametes, $(\boldsymbol{Z}_d + \boldsymbol{Z}_f)$ reflects the genotype of the F_1 hybrid and is coded as 1 and -1 for the two homozygous genotypes and 0 for the heterozygous genotype. Finally, $\boldsymbol{e}$ is the residual vector.

Model U_2, extends U_1 by dominance effects,

$$\boldsymbol{y} = \mathbf{1}\mu + (\boldsymbol{Z}_d + \boldsymbol{Z}_f)\boldsymbol{u} + \boldsymbol{D}\boldsymbol{d} + \boldsymbol{e} \quad (2)$$

where matrix $\boldsymbol{D} = -2(\boldsymbol{Z}_d \circ \boldsymbol{Z}_f) + \frac{1}{2}\boldsymbol{J}$, with $\boldsymbol{J}$ being a $N \times M$ matrix containing only 1s and $\circ$ denoting element-wise multiplication. Thus, $\boldsymbol{D}$ codes heterozygous genotypes as 1 and homozygous genotypes as 0, in harmony with the F_∞ metric commonly used in textbooks on quantitative genetics (e.g., Falconer and Mackay 1996). The vector $\boldsymbol{d}$ contains the random dominance effects.

Model S_1 is again an additive model. However, this time we model the marker effects as specific to the population of origin,

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z}_d\boldsymbol{u}_d + \boldsymbol{Z}_f\boldsymbol{u}_f + \boldsymbol{e}. \quad (3)$$

Vectors $\boldsymbol{u}_d$ and $\boldsymbol{u}_f$ contain the random marker effects pertaining to the dent and flint population.

The most complex model S_2 extends model S_1 by dominance effects between the marker alleles of the two populations,

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z}_d\boldsymbol{u}_d + \boldsymbol{Z}_f\boldsymbol{u}_f + \boldsymbol{D}\boldsymbol{d}_{df} + \boldsymbol{e}. \quad (4)$$

While $\boldsymbol{D}$ from model S_2 is identical to $\boldsymbol{D}$ from model $U_2, \boldsymbol{d}_{df} \neq \boldsymbol{d}$ in general, because of the different formulation of additive effects.

Estimation of marker effects

GBLUP

GBLUP marker effects were estimated by solving the mixed model equations corresponding to models U_1, U_2, S_1 and S_2. The shrinkage factors of marker effects were computed from

GCA and SCA variance components (see below for details on the estimation of variance components). For example, the shrinkage factor for $\boldsymbol{u}_d$ was $\sigma_e^2/(\sigma^2_{\mathrm{GCA}^d}/M)$, where $\sigma^2_{\mathrm{GCA}^d}$ is the variance component of GCA effects pertaining to gametes from the dent pool and σ_e^2 is the residual variance component. For models U_1 and U_2, the genetic variance component for the shrinkage factor of $\boldsymbol{u}$ was pooled. In case of computationally singular coefficient matrices, we used the function "make.positive.definite" from R package "corpcor" (Schaefer et al. 2012) to ensure invertability.

BayesB

For sake of brevity, we will give details only for model U_1. The extension to the other models follows straightforwardly by specifying analogous but independent prior distributions for the added corresponding parameters.

In analogy to Meuwissen et al. (2001), we specified $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{G}_u)$, where $\boldsymbol{G}_u$ is a diagonal matrix of dimension $M \times M$, with $diag(\boldsymbol{G}_u) = [\sigma^2_{u_1}, \sigma^2_{u_2}, \ldots, \sigma^2_{u_M}]$, as our prior for $\boldsymbol{u}$. Our prior for the marker specific variance of effects, $\sigma^2_{u_i}$, was

$$p(\sigma^2_{u_i}|v_u, S^2_u)\begin{cases} = 0 & \text{with probability } \pi_u \\ = \chi^{-2}(v_u, S^2_u) & \text{with probability } (1-\pi_u), \end{cases} \quad (5)$$

where v_u, S^2_u and π_u are hyperparameters. The subscript "u" indicates that the hyperparameters are specific to additive marker effects $\boldsymbol{u}$ from model U_1. For the sake of simplified notation, we will later on drop the subscript when discussing the role of these parameters in general and indicate textually when we refer to a specific case. Following the suggestions of Yang and Tempelman (2012), we modeled the hyperparameters v_u, S^2_u and π_u as uncertain by assigning hyperprior distributions to them. For v_u, we chose the following prior:

$$p(v_u)\begin{cases} \propto (v_u+1)^{-2} & \text{if } 0 < v_u < 100 \\ = 0 & \text{else.} \end{cases} \quad (6)$$

We observed better convergence by setting an upper bound to v_u. Furthermore, we specified $p(S^2_u) = Gamma(\alpha_S = 0.1, \beta_S = 0.1)$. We used a mildly informative prior for π_u, namely $p(\pi_u) = Beta(\alpha_\pi = 3, \beta_\pi = 3)$, the probability peak of which is around $\pi_u = 0.5$, but still gives substantial probability to $0.1 > \pi_u > 0.9$.

Finally, our prior for the residual variance was $p(\sigma^2_e) = \chi^{-2}(v_e = -1, S^2_e = 0)$.

To fit the models, we ran the Gibbs-sampler for 100,000 iterations. The first 25,000 were discarded as burn-in and only samples from every 30th post burn-in iteration were stored. These parameters were chosen to ensure an effective sample size of > 100 for the hyperparameters v, S^2 and π. The effective sample size was estimated with the function "effectiveSize" from the "coda" R package (Plummer et al. 2010), the functionality of which was also used to monitor convergence in general.

The Gibbs sampling strategy and fully conditional distributions (FCD) of all the parameters are described in detail in Yang and Tempelman (2012). We used the independence Metropolis–Hastings (MH) algorithm to sample from the FCD of $\sigma^2_{u_i}$ as described by Yang and Tempelman (2012), with 5 MH steps. To sample from the FCD of v_u we followed the recommendations of Kizilkaya et al. (2003) and employed the random walk MH algorithm. The variance of the Normal candidate distribution was tuned during the burn-in to achieve an acceptance probability of ≈ 0.45, as suggested by Müller (1991). The MH sampler was run for 100 steps during burn-in and for 10 steps post burn-in. We used the posterior means of the marker effects as point estimates to predict the genotypic values.

Genome

Parents and hybrids

We based our simulations on the single-nucleotide polymorphism (SNP) marker genotypes of 100 dent and 100 flint inbred lines from the maize breeding program of the University of Hohenheim, genotyped with the Illumina SNP chip MaizeSNP50 (Ganal et al. 2011). Ignoring residual heterozygosity and mutational events, the phased marker genotypes of the hybrids can be inferred from the genotypes of their parental inbred lines. Thus, we created in silico all possible 10,000 hybrids of the complete factorial of the 100 dent $\times$ 100 flint crosses.

Consequently, in terms of the allele frequency distribution, LD pattern, and population substructure of the two populations of parent lines and their hybrid population, our simulations represent the situation encountered in an actual breeding program.

Marker data

We removed all markers with more than 5 % missing values, where we treated heterozygous marker genotypes as "missing" as well. Remaining missing marker genotypes were imputed using version 3.3.1 of "BEAGLE" (Browning and Browning 2009). Here, we assumed known haplotype phases, because the lines were regarded to be fully homozygous. A total of 39,627 markers were subsequently available for further analysis.

We investigated the LD structure of the inbred line populations by fitting second-order natural smoothing

splines (with < 80 effective degrees of freedom) onto the scatterplot of LD (r^2) versus physical distance (Δ) in mega base pairs (Mbp) between markers on the same chromosome. This was done separately within the set of dent lines, flint lines and across both sets. For the within set LD, all markers with a minor allele frequency (MAF) > 0.05 for this set were considered; for the LD across sets, all markers with MAF > 0.05 within both sets were considered.

To investigate the persistence of linkage phases across the two inbred line populations, we first binned all marker pairs according to Δ in 100 discrete bins of ≈ 0.035 Mbp width. Then we computed the proportion of pairs within each bin that had the same linkage phase (determined by the sign of the r statistic) within the dent and flint population. We also used second-order natural smoothing splines (with 5 effective degrees of freedom) on the scatterplot of this proportion versus the center values of the bins.

To visualize the genetic differences between the set of dent lines and flint lines as well as the population substructure within these populations, we generated a neighbor joining tree based on the Modified Rogers' distance (MRD) between the marker profiles of the lines.

The LD statistics, the neighbor joining tree as well as the statistics concerning allele frequencies were based on the full set of 39,627 SNP markers of all lines.

For the simulations, we used only those markers from the 39,627 remaining ones that segregated in each inbred line population with MAF > 0.05. We then reduced the number of markers to $\approx$ 10 Mbp^{-1}, 1 Mbp^{-1}, and 0.3 Mbp^{-1}. The resulting number of available markers was $\approx$ 5000 (10 Mbp^{-1} density), $\approx$ 1450 (1 Mbp^{-1}) and $\approx$ 580 (0.3 Mbp^{-1}), respectively. Small random frame shifts were permitted when spacing the markers to allow the set of markers used for analysis to change from one replication of the simulations to another.

Traits

From the set T of SNP loci not used as markers, we identified the subset T_d where the allele frequency (of the same marker allele) in the population of dent lines, p_d, and flint lines, p_f, had $|p_d - p_f| > 0.6$ ("divergent" parental populations). Conversely, the subset $T_c \subset T$ consisted of those markers with $|p_d - p_f| < 0.05$ ("convergent" parental populations). Henceforth, these two will also be referred to as "divergent" and "convergent" inter-population structure scenarios. We then randomly sampled 300 SNP markers to be assigned as QTL, from either T_d or T_c, depending on the scenario. The restrictions thereby were that the MAF had to be > 0.025 in both sets of lines and the physical distance Δ between neighboring QTL had to be > 0.5 Mbp.

A random subset of 250 of these QTL were assigned additive (a) and dominance (d) effects, defined according to Falconer and Mackay (1996). The additive effects were drawn from a reflected Gamma distribution with parameter scale = 1.66 and shape = 0.4, as was often used in the literature, e.g., by Meuwissen et al. (2001). The dominance effects of these loci were obtained as the product between the absolute additive effect and the degree of dominance. The degrees of dominance were drawn from a Normal distribution with mean 1.0, a value based on experimental estimates for grain yield in maize (Gardner and Lonnquist 1959; Gardner 1963), and variance 0.75. To the remaining 50 QTL, we assigned pure dominance effects, drawn from a Normal distribution with mean and variance equal to the observed mean and variance of the dominance effects of the 250 other QTL.

After assigning the QTL, we computed the genotypic values of the 10,000 hybrids by summing the additive and dominance effects across all QTL. The genotypic values were then scaled to unit variance and centered to zero mean.

For computing the GBLUP shrinkage factors, as well as for interpreting our results in comparison with experimental data from maize, we estimated variance components, pertaining to general (GCA) and specific (SCA) combining ability effects, from a simulated completely randomized design with two replications, including all 10,000 hybrids. For this, we added a normally distributed noise variable to the genotypic values of the hybrids to arrive at a broad sense heritability of $h^2 = 0.75$ on an entry mean basis, a typically observed value for grain yield and grain moisture in the maize breeding program of the University of Hohenheim (Schrag et al. 2006). We then fitted the following model:

$$y_{ijk} = \mu + \mathrm{GCA}_i^d + \mathrm{GCA}_j^f + \mathrm{SCA}_{i\times j} + e_{ijk}, \qquad (7)$$

where μ is the intercept, y_{ijk} is the phenotypic value of a hybrid between dent line i and flint line j in the kth replication, GCA_i^d is the GCA effect of the ith dent line, GCA_j^f is the GCA effect of the jth flint line, $\mathrm{SCA}_{i\times j}$ the SCA effect of the hybrid between dent line i and flint line j and e_{ijk} is the residual in the kth replication of hybrid y_{ij}. For simplicity's sake, we regarded the inbred lines as unrelated. We used model (7) to estimate the variance components pertaining to GCA effects of the dent ($\sigma^2_{\mathrm{GCA}^d}$) and flint ($\sigma^2_{\mathrm{GCA}^f}$) lines and the SCA ($\sigma^2_{\mathrm{SCA}}$) effects.

Prediction of hybrid performance

A random sample of 75 lines from each population was taken to be "evaluated" in hybrid combinations in silico, as illustrated by Fig. 1. Hybrids for which both the dent and

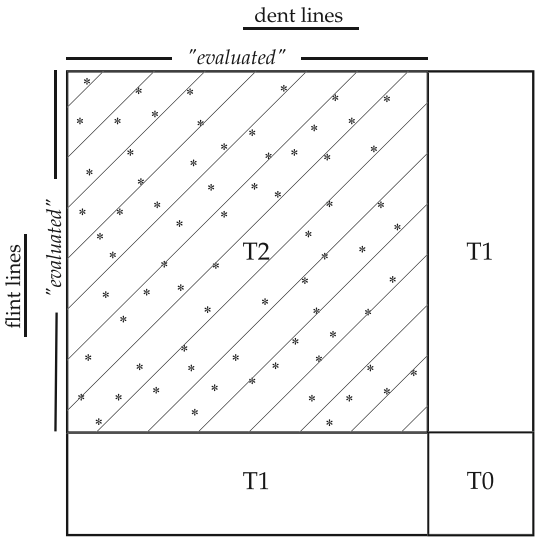

Fig. 1 Schematic visualization of the division of the complete factorial into *T2*, *T1*, and *T0* hybrids as well as the training set. The "evaluated" lines are a random sample of 75 dent and 75 flint lines from the whole set of 100 lines within each population. Hybrids, where both parents are "evaluated" belong to the *T2* group, those where only one parent is "evaluated", to the *T1* group, and those where no parent is "evaluated", to the *T0* group. A random sample of 800 hybrids from all 5625 T2 hybrids was used for training (indicated as *asterisks*). These were consequently excluded from the *T2* group

the flint parent were "evaluated" were assigned to the "T2" set. Hybrids having either the dent or the flint parent (but not both) as being "evaluated" were assigned to the "T1" set. Finally, hybrids where no parent was "evaluated" were assigned to the "T0" set. A random sample of $N = 800$ hybrids from the T2 group was used as the training population. The number $N = 800$ was chosen because it realistically reflects the number of hybrids that can be evaluated in resource intensive, multi-environment field trials in a medium-size breeding program. The phenotypes of the training hybrids were the entry means over the two replications of the hybrids created above for estimating variance components. These phenotypic values thus had $h^2 = 0.75$. The remaining $75 \times 75 - 800 = 4825$ T2 hybrids were used for validation, as were the $75 \times 25 \times 2 = 3750$ T1 hybrids and the $25 \times 25 = 625$ T0 hybrids. We evaluated the prediction accuracy separately for each validation set by computing the Pearson correlation between predicted genotypic values, based on the parental marker genotypes and estimates of the marker effects, and true genotypic values, based on the QTL genotypes and the simulated additive and dominance effects. We used the prediction accuracy as criterion to assess the model performance.

For each scenario of QTL-allele inter-population structure and marker density, we generated 50 data sets by repeating the whole simulation process (i.e., sampling of "evaluated lines", subset of markers, subset of QTL and their effects and environmental noise). All four models were fitted by both estimation methods to each data set, which therefore acted as a blocking factor and allowed to compare the observed prediction accuracies with paired *t* tests. We also computed the average accuracy for a data set by averaging over the values observed for the four models. We then used the standard deviation (SD) of these average accuracies as a measure of variability of the data sets, within a factor combination, i.e. the variability attributable to random sampling effects.

We conducted an analysis of variance to assess the influence of the factors: marker density, inter-population structure, validation group, estimation method, inclusion of dominance effects ("yes" or "no"), inclusion of population-specific effects ("yes" or "no"), and all two-way interactions between these on the prediction accuracies. The data set (i.e., the replications of the simulation) was included as a factor as well; it acted as a blocking factor for all factors related to model/method choice. We used *t* tests to evaluate the significance of differences between means of factor levels of interest to us. Thereby we used the data set as a blocking factor when appropriate. All reported differences were significant with $p < 0.05$, unless noted otherwise.

All computations were performed in the R statistical environment (R Development Core Team 2011). Model (7) was fitted with the "lme4" R package (Bates et al. 2011). The neighbor joining tree was generated with the package "ape" (Paradis et al. 2004). The natural splines were fitted with package "pspline" (Ramsey and Ripley 2010). BayesB was implemented as a C program integrated to R.

Results

Results related to observed genomic data

Allele frequency distribution

The average MAF was 0.185 within the dent lines and 0.135 within the flint lines. The proportion of markers with MAF < 0.05 (< 0.025) was 0.344 (0.255) within the dent lines and 0.441 (0.312) within the flint lines. The MAF within the dent lines was almost evenly spread across the whole value range (Fig. 2a), while they were more concentrated at lower values within the flint lines (Fig. 2b). The density of the distribution of absolute differences in allele frequencies ($|p_d - p_f|$, for markers segregating in both populations) had its peak at values close to zero and then declined with increasing values for the difference (Fig. 2c).

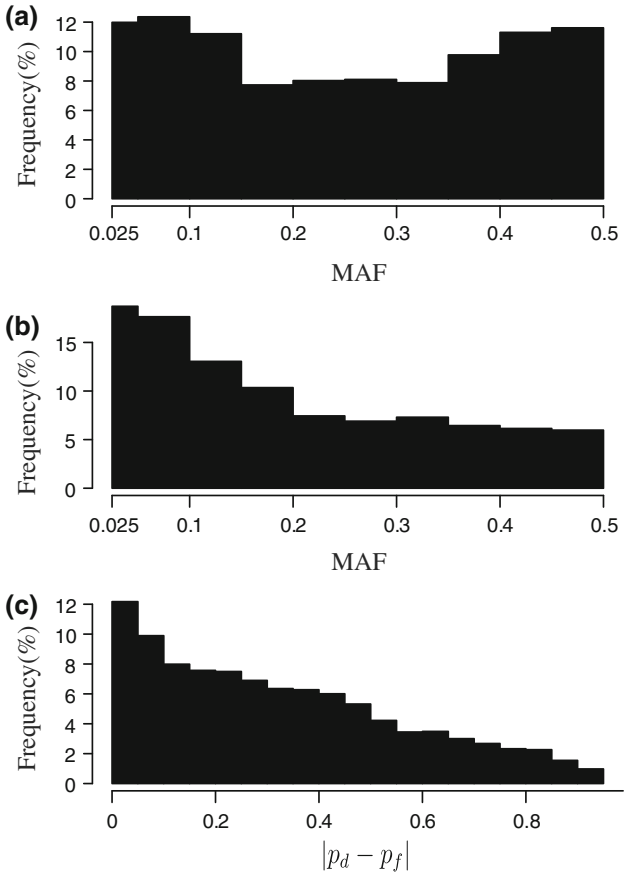

Fig. 2 Histogram of MAF within the set of dent (**a**) and flint (**b**) lines and histogram of $|p_d - p_f|$ (**c**), where p_d and p_f is the frequency of corresponding alleles within the set of dent and flint lines, respectively. Histograms (**a**) and (**b**) contain only markers with a MAF >0.025 in the corresponding set of lines, histogram (**c**) only markers with a MAF >0.025 within both line populations, from the whole set of 39,627 SNP markers

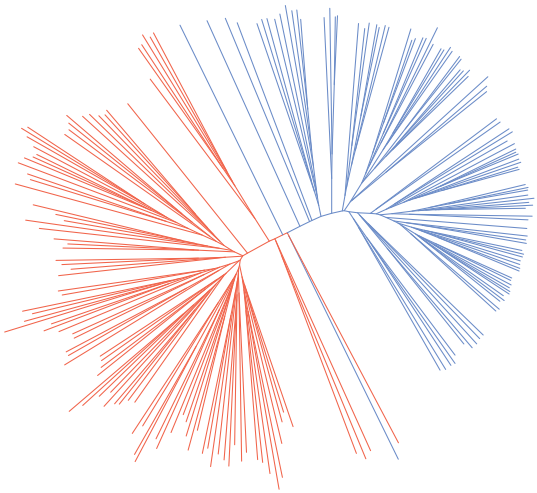

Fig. 3 Neighbor joining tree based on the Modified Rogers distance between the marker genotypes of the inbred lines. Lines from the dent population are indicated in *red*, lines from the flint population in *blue* (color figure online)

Population structure and LD

In a neighbor joining tree, the two sets of inbred lines formed two distinct groups, with only few intermediate genotypes (Fig. 3). The set of flint lines was more structured than the set of dent lines.

The second-order natural smoothing spline fits of pairwise LD (measured as r^2) versus physical distance Δ in Mbp between two markers showed very strong LD between markers for $\Delta < 0.25$ Mbp, within the set of dent lines ($r^2 > 0.30$), within the set of flint lines ($r^2 > 0.35$) and across the combined set of both dent and flint lines ($r^2 > 0.25$) (Fig. 4a). The decline in LD was rather steep up to $\Delta \approx 0.5$ Mbp and slowed down considerably for greater distances. Beyond $\Delta = 1.5$ Mbp, the LD remained almost constant at values of around $r^2 = 0.2$, for the whole range of Δ values considered. The LD across the sets of lines was generally lower than within the sets. For the whole range of Δ, LD within the set of flint lines was higher than within the set of dent lines (Fig. 4a).

The second-order natural smoothing spline fit for the proportion of marker pairs with the same linkage phase in both sets of lines versus Δ showed a trend similar to the curves for the LD. For marker pairs with $\Delta < 0.25$ Mbp, this proportion was greater than 0.7 (Fig. 4b). It then declined to a value of $\approx$ 0.575, where it remained almost constant.

Results related to simulations

Variance components

The ratio of SCA variance versus GCA variance, $\sigma^2_{\mathrm{SCA}}/(\sigma^2_{\mathrm{GCA}^d} + \sigma^2_{\mathrm{GCA}^f})$, was 0.069 for the "divergent" scenario and 0.243 for the "convergent" scenario (averaged over all data sets pertaining to the corresponding scenario).

Prediction accuracies

The prediction accuracies averaged over all 50 replications ranged from 0.65 to 0.95 (Table 1). The lowest value was observed for "convergent" parental populations in the T0 validation group for 0.3 Mbp^{-1} marker density and model U_1. The highest values were obtained for several cases of "divergent" parental populations in the T2 validation group.

Due to the high number of degrees of freedom, all factors had a significant influence as determined in the analysis of variance (supplemental Table 1). The only exceptions were the two-way interaction between method

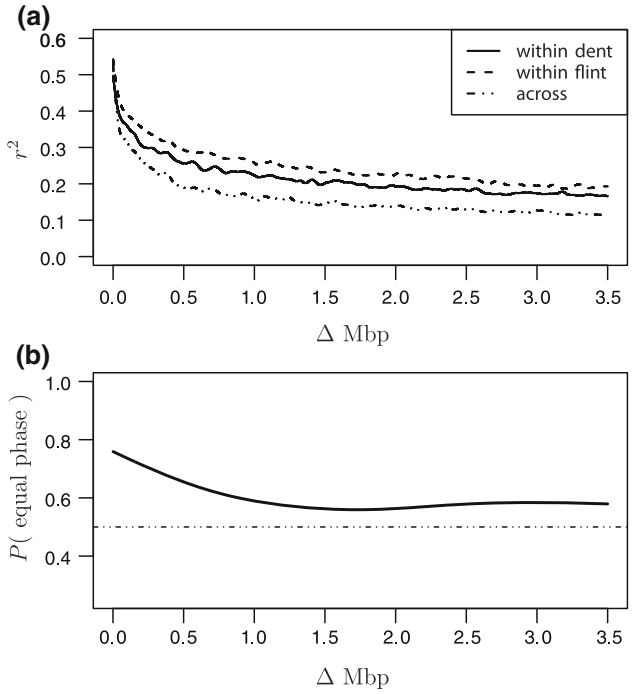

Fig. 4 (**a**) Second-order smoothing spline fits of LD (r^2) versus distance (Δ) in mega base pairs (Mbp) between markers on the same chromosome, within the set of dent lines (*full line*), flint lines (*dashed line*), and across both sets (*dotted-dashed line*). For the within set LD, all markers with a minor allele frequency (MAF) >0.05 for this set were considered; for the LD across sets, all markers with MAF >0.05 within both sets were considered. (**b**) Second-order smoothing spline fits of proportion of marker pairs with equal linkage phase (determined as equality of sign of *r* statistic) versus Δ between markers on the same chromosome. The *horizontal dotted-dashed line* indicates the value 0.5

and inclusion of population-specific effects and the interaction between inclusion of population-specific effects and inclusion of dominance effects.

The average observed prediction accuracies obtained with BayesB were with 0.006 significantly higher than those of GBLUP. While there was no significant difference for models that did not include dominance effects (U_1 and S_1), the difference between BayesB and GBLUP was 0.012 and significant for models that included dominance (U_2 and S_2). In certain situations, the differences could be substantial (Table 1). For example, with "convergent" parental populations and marker density 10 Mbp^{-1}, the difference between BayesB and GBLUP was 0.024 for models incorporating dominance. For this reason, the following presentation will focus on the results obtained with BayesB.

Irrespective of other factors, the prediction accuracies were higher by 0.088 under "divergent" than under the "convergent" parental populations. The mean prediction accuracy for the T0 validation group was by 0.073 lower than that of the T1 validation group and by 0.137 lower than that of the T2 group. The mean accuracy rose from 0.822 with 0.3 Mbp^{-1} density to 0.842 (1 Mbp^{-1}) to 0.850 (10 Mbp^{-1}).

Models incorporating dominance effects (U_2 and S_2) yielded a mean accuracy that was significantly higher by 0.026 than their only additive counterparts (U_1 and S_1). While always statistically significant, the differences in means increased with increasing marker density (for example, 0.021 under 0.3 Mbp^{-1} density versus 0.031 under 10 Mbp^{-1} density) and from T2 to T0 (0.023 for T2 vs. 0.028 for T0). They were further larger under the "convergent" scenario than under the "divergent" scenario (0.039 vs. 0.013).

The overall mean accuracy of models incorporating population-specific effects, S_1 and S_2, over models that did not, U_1 and U_2, was by 0.005 points significantly higher. The superiority of the specific models was generally smaller under the "convergent" scenario compared with the "divergent" scenario (0.002 vs. 0.008). At a marker density of 10 Mbp^{-1}, these models yielded even slightly inferior accuracies than their unspecific counterparts in some cases (Table 1). For the "divergent" scenario, the superiority of models S_1 and S_2 generally increased with decreasing marker density (from 0.004 at 10 Mbp^{-1} to 0.012 at 0.3 Mbp^{-1}) and from the T2 to T0 validation groups (0.001 for T2 vs. 0.015 for T0).

The SD of average (over models) prediction accuracies of data sets ranged from ≈0.1 to ≈0.01 (Table 1). It was generally higher under the "convergent" scenario than under the "divergent" scenario and decreased from T0 to T2. There was no obvious trend for the SD regarding the marker density.

Discussion

Observed genomic data

Allele frequency distribution

The more extreme distribution of MAF in flint compared with dent lines (Fig. 2) and the much higher proportion of MAF close to zero indicate that the flint population used was much more narrow in terms of allelic diversity than the dent population.

Similar to other maize breeding programs in Central Europe, the flint germplasm of the University of Hohenheim largely traces back to a relatively small number of first cycle lines extracted from European landraces (Fischer et al. 2008), the most prominent examples being the French lines F2 and F7 developed from Lacaune, the Spanish line EP1 developed from Lizagarotte, and the German line DK105 developed from Gelber Badischer. The first cycle lines were subsequently subject to intensive recycling

Table 1 Prediction accuracies for hybrids obtained with estimation method BayesB and GBLUP for various validation groups (T0, T1, T2), under the four models (U_1, U_2, S_1, S_2), and the "convergent" and "divergent" inter-population structure, for various marker densities (10, 1, and 0.3 Mbp^{-1}), averaged over 50 data sets

Marker density	Inter-population structure	Validation group	BayesB					GBLUP				
			Unspecific		Specific		SD	Unspecific		Specific		SD
			U_1	U_2	S_1	S_2		U_1	U_2	S_1	S_2	
10 Mbp^{-1}	convergent	T0	0.715^{a}	0.764^{b}	0.704^{c}	0.760^{b}	0.089	0.708^{a}	0.734^{b}	0.710^{a}	0.738^{c}	0.093
		T1	0.798^{a}	0.842^{b}	0.791^{c}	0.840^{b}	0.050	0.794^{a}	0.817^{b}	0.793^{a}	0.818^{b}	0.057
		T2	0.861^{a}	0.907^{b}	0.860^{c}	0.907^{b}	0.036	0.862^{a}	0.884^{c}	0.859^{b}	0.883^{c}	0.042
	divergent	T0	0.816^{a}	0.837^{b}	0.831^{b}	0.838^{b}	0.050	0.817^{a}	0.821^{b}	0.828^{c}	0.830^{c}	0.056
		T1	0.879^{a}	0.895^{b}	0.885^{c}	0.896^{b}	0.022	0.879^{a}	0.882^{b}	0.884^{bc}	0.904^{c}	0.026
		T2	0.935^{a}	0.947^{b}	0.936^{c}	0.948^{b}	0.007	0.935^{a}	0.938^{a}	0.934^{b}	0.937^{ab}	0.009
1 Mbp^{-1}	convergent	T0	0.703^{a}	0.742^{b}	0.713^{c}	0.746^{b}	0.078	0.704^{a}	0.722^{b}	0.709^{a}	0.727^{b}	0.087
		T1	0.784^{a}	0.82^{b}	0.789^{c}	0.822^{b}	0.046	0.785^{a}	0.801^{b}	0.785^{a}	0.803^{b}	0.051
		T2	0.855^{a}	0.890^{b}	0.855^{a}	0.890^{b}	0.036	0.855^{a}	0.870^{b}	0.852^{c}	0.869^{d}	0.040
	divergent	T0	0.805^{a}	0.829^{b}	0.822^{b}	0.836^{c}	0.057	0.819^{a}	0.823^{b}	0.833^{c}	0.838^{d}	0.050
		T1	0.873^{a}	0.890^{b}	0.881^{c}	0.894^{d}	0.025	0.878^{a}	0.881^{b}	0.884^{b}	0.888^{c}	0.028
		T2	0.934^{a}	0.945^{b}	0.936^{c}	0.946^{d}	0.008	0.934^{a}	0.936^{ab}	0.933^{b}	0.936^{ab}	0.012
0.3 Mbp^{-1}	convergent	T0	0.651^{a}	0.688^{b}	0.666^{c}	0.700^{d}	0.102	0.655^{a}	0.679^{b}	0.676^{b}	0.697^{d}	0.102
		T1	0.753^{a}	0.785^{b}	0.760^{c}	0.791^{d}	0.062	0.755^{a}	0.778^{b}	0.762^{c}	0.785^{d}	0.061
		T2	0.840^{a}	0.870^{b}	0.841^{c}	0.871^{d}	0.043	0.839^{a}	0.862^{b}	0.835^{c}	0.860^{d}	0.046
	divergent	T0	0.796^{a}	0.813^{b}	0.826^{c}	0.831^{c}	0.061	0.799^{a}	0.806^{b}	0.828^{c}	0.833^{d}	0.061
		T1	0.865^{a}	0.877^{b}	0.879^{b}	0.886^{c}	0.030	0.866^{a}	0.871^{b}	0.878^{c}	0.883^{d}	0.030
		T2	0.933^{a}	0.941^{b}	0.935^{c}	0.943^{d}	0.010	0.932^{a}	0.936^{b}	0.931^{c}	0.936^{b}	0.011

The standard deviations (SD) of 50 average prediction accuracies (averaged over models) are shown in the last column

Values followed by identical letters within a row and estimation method are not statistically different in paired t tests for $p < 0.05$

breeding with occasional introgression of exotic flint germplasm from tropical or subtropical CIMMYT germplasm or introgression of modern Lancaster lines from North America, but to a large extent, the flint germplasm pool was kept close. By comparison, the dent material traces back to numerous landraces and sources from North America used at the beginning of the era of hybrid breeding in Europe in the 1950s (Fischer et al. 2008). The main source was germplasm with Reid Yellow Dent background, but there was from the beginning a steady flow of germplasm from North American breeding programs that served as a continuous source for broadening the dent heterotic pool. Moreover, it was not uncommon to extract new dent inbreds by selfing late-maturing pure dent hybrids cultivated in Southern Europe. Thus, the history of the hybrid breeding program of the University of Hohenheim can well explain the observed differences in allele frequency distributions between the dent and flint heterotic pools.

However, we acknowledge that these results might be influenced by the reported ascertainment bias in the Illumina SNP chip MaizeSNP50 (Ganal et al. 2011), which can lead to an underestimation of diversity in germplasm not well represented by the SNP discovery and selection process. Assessment of an ascertainment bias could be achieved by comparison of the SNP marker data with other marker systems such as Simple Sequence Repeats (SSR), as performed by Van Inghelandt et al. (2010).

Surprisingly, the differences in allele frequencies $|p_d - p_f|$ between the dent and flint populations were in most cases not extreme (Fig. 2c). This was unexpected given the long history of separate evolution of the dent and flint germplasm and the relatively small effective population size practiced in each heterotic pool that would be expected to result in rather diverse allele frequencies based on random genetic drift alone. Nevertheless, it must be kept in mind that markers monomorphic in one or both of the pools were disregarded in our study by excluding SNP with MAF < 0.025 for the analysis. Despite the rather small differences, the two populations were found to be genetically clearly distinct (Fig. 3).

LD

We observed high levels of LD within the two populations compared with observations from other studies in maize. For example, Riedelsheimer et al. (2012) found that r^2 declined below 0.1 already after $\approx$ 0.5 Mbp in a set of diverse dent inbred lines. In a situation more comparable to

ours, Van Inghelandt et al. (2011) found a similar decline within commercial flint and dent germplasm. In our material, the LD within populations at 0.5 Mbp was considerably higher than observed in these studies. Furthermore, in contrast to the studies mentioned above, we observed strong long range LD, with $r^2 > 0.15$ even after 3 Mbp (Fig. 4a). However, in concordance with our results, Albrecht et al. (2011) found strong long-range LD within a commercial population of dent lines, with a considerable proportion of marker pairs showing $r^2 > 0.1$ even after 50 Mbp.

The consistently higher LD for the flint population in comparison with the dent population (Fig. 4) can also be attributed to the breeding history of each pool recapitulated above and corroborates that the flint heterotic group was kept more secluded than the dent heterotic group. Furthermore, the current flint germplasm might still exhibit traces of the population substructure created at the outset by using lines from distinct landraces and/or by inclusion of Lancaster germplasm. This residual population substructure might then have caused admixture LD. Remains of residual population substructure might be visible in the neighbor joining tree in Fig. 3, which shows a somewhat more defined structuring within the flint lines compared to within the dent lines.

A direct result of the narrow base of the European flint, and, to a lesser extent, the European dent population, the heavy reliance on recycling breeding, and the low effective population sizes within breeding materials is that many of the inbred lines within a population are in fact closely related. Therefore, they share large genome portions identical by descent. This constitutes another source of the large LD observed within populations.

We also observed strong LD across the two populations (Fig. 4a). While always lower than the LD within populations, it was still considerably higher than the LD levels observed by Riedelsheimer et al. (2012) and Van Inghelandt et al. (2011). LD across populations is a function of two factors: LD within each of the populations and admixture LD arising from differences in allele frequencies (Charcosset and Essioux 1994). The strong LD across populations observed in our study will largely be an effect of the latter cause, given that the differences in allele frequencies observed are distinctly different than zero for many markers (Fig. 2c). LD within populations can only contribute to LD across populations when the linkage phases are identical in both populations. Given the long separation of the two populations and the random nature of events that create LD, one would expect about 50 % of marker pairs with identical linkage phase in both populations. Consequently, the higher proportion of marker pairs with identical sign of the LD in dent and flint lines was surprising (Fig. 4b). Thus, not only population admixture, but also LD within populations contributed to the LD across populations.

Simulations

Variance components

The almost four times higher ratio $\sigma^2_{SCA}/(\sigma^2_{GCA^d} + \sigma^2_{GCA^f})$ under the "convergent" scenario compared with the "divergent" scenario is in harmony with the theoretical results of Reif et al. (2007). They found that the importance of σ^2_{SCA} compared to σ^2_{GCA} declines with increasing divergence of the two parent populations used in hybrid breeding. This is mirrored in our simulations where the subset of possible QTL for the "divergent" scenario (T_d) corresponds to markers with distinctly different allele frequencies in both populations and the subset for the "convergent" scenario (T_c) corresponds to markers with very similar allele frequencies. Overall, the ratios of variance components obtained from our simulation matched, for both scenarios, the ratios observed for grain yield and grain moisture in the maize breeding program of the University of Hohenheim very well (Schrag et al. 2006).

Prediction accuracies

Marker density and QTL scenario

The prediction accuracies increased with increasing marker density, as expected (Table 1). However, the differences between the accuracies observed for a marker density of 10 Mbp^{-1} and the 30 times lower density of 0.3 Mbp^{-1} were only moderate. This is most likely attributable to the strong long-range LD observed in both parent populations and across populations (Fig. 4) and suggests that useful screens of genotypes could be conducted with low-density chips already, for the type of material investigated here. Such chips may be produced with much lower costs than the present high-density chips. This would enable screening the huge set of new doubled haploid (DH) lines (about 10,000 per heterotic group) generated anew in each breeding cycle, which is currently too costly even with the relatively low costs of the high density chip employed in our study.

The prediction accuracies observed for the "convergent" and "divergent" parental populations can reflect different stages in a hybrid breeding program. While the "convergent" scenario corresponds to the very beginning, where the heterotic groups have still similar allele frequencies, the "divergent" scenario should reflect the situation at more advanced stages, where the allele frequencies have more diverged as result of reciprocal recurrent selection (Labate et al. 1999). One explanation for the

consistently lower prediction accuracies for the "convergent" scenario is the greater importance of σ^2_{SCA}. In the absence of epistasis, as assumed in our study, SCA is exclusively attributable to dominance effects (Reif et al. 2007). Dominance effects, representing higher order effects, are more difficult to estimate than additive effects and thus reduce prediction accuracy. The greater importance of σ^2_{SCA} might also reduce the impact of identical copies of gametes in the training set and T2 and T1 validation groups, because in this case the specific combination of the gametes in a hybrid will become more important.

Validation groups

As expected, the largest difference in the prediction accuracies were observed between the three validation groups T2, T1, and T0, irrespective of the setting for the other factors. All gametes produced by a fully homozygous line are identical. Therefore, all hybrids with the same parental inbred line on the dent or flint side, share identical copies of the parental gametes. Technically, they are half-sibs from a fully homozygous parent. Hybrids from the T2 group have on average about ten half-sibs with a common dent and a common flint parent in the training population. In other words, both of their gametes are represented with multiple identical copies in the training population. The genotypic values of T2 hybrids can therefore be predicted with very high accuracy; in some cases, the accuracy reached almost 0.95, i.e., $\approx$90 % of the genetic variance could be explained (Table 1). The above mentioned GCA effects can be seen as main effects of the dent and flint gametes. Given the preponderance of the GCA variance over the SCA variance, the fact that the gametic copies are found in different combinations in the training set than in the remaining T2 set, has little impact because accuracy will largely depend on the prediction of additive effects. In fact, we would expect similarly high accuracies when using the estimated GCA effects from a model such as (7). Hybrids from the T1 group have only one gamete (either from the dent or the flint side) represented by identical copies in the training set. Consequently, their prediction accuracy was intermediate between the T2 group and the T0 group, which is not represented by identical copies of gametes in the training set at all.

Estimation methods GBLUP and BayesB

The same general trends for the prediction accuracy in terms of relative model performance as well as in terms of factors such as marker density were observed for both estimation methods, GBLUP and BayesB. Thus, the choice of estimation method was clearly not critical for obtaining our results.

Overall, both methods yielded high prediction accuracies, with practically no difference under the additive models U_1 and S_1 (Table 1). However, we observed that BayesB could outperform GBLUP when the best models U_2 and S_2 were used, i.e., when dominance effects were included. When focusing only on these models, and on the most relevant and interesting scenarios (T0 with 10 Mbp^{-1} density), the differences were with 0.026 ("convergent" inter-population structure) and 0.012 ("divergent" inter-population structure) sizable and significant. This leads to the conclusion that BayesB succeeded better in estimating dominance effects than GBLUP.

As follows from the formulation of GBLUP, additive ($\boldsymbol{u}, \boldsymbol{u}_d$ and $\boldsymbol{u}_f$) as well as dominance ($\boldsymbol{d}, \boldsymbol{d}_{df}$) effects were evenly shrunken towards zero with no magnitude differences in effect sizes. Interestingly, while BayesB followed this pattern for the additive effects, whose estimates were very similar to those of GBLUP, it differed considerably for dominance effects. With BayesB, the dominance effects of most markers were shrunken extremely towards zero. In some cases, only few markers with sizable dominance effects remained per chromosome. Indeed, for dominance effects, BayesB tended towards solutions typical for Bayesian adaptive shrinkage methods (Li and Sillanpää 2012; Xu 2003). Supplemental Figure 1 depicts the marker effect estimates obtained by GBLUP and BayesB for a representative example.

These observations were also mirrored on the level of the hyperparameters v, π and S^2, which appear in Eq. (5). The posterior estimate of the parameter π gives the probability of marker effect variances equal to exactly zero, which leads to marker effects of exactly zero as well. The posterior estimate of v quantifies the dependence of the individual marker effect variances on the typical value, which is given by the posterior estimate of S^2. The higher v, the higher the dependence and the smaller the deviations from S^2.

For example, for "convergent" parental populations with marker density 10 Mbp^{-1} and model S_2, the posterior means of v, averaged over the 50 replications, were 4.95, 9.65 and 8.97 for dominance effects and additive effects pertaining to the dent and flint pool, respectively. Here, the differences between v for dominance and for additive effects were significant. Thus, the dependence of the marker effect variances on S^2 (posterior estimates always around 0.0004) was low for dominance effects, which allowed for some large and many tiny variances and therefore for adaptive shrinkage of marker effects. For variances of additive effects, the inverse Chi-square part of the prior mixture distribution in Eq. (5) had an almost 10 times lower variance and thus forced all individual marker effect variances to stay in the vicinity of the typical value

S^2, when they were not exactly zero. Similar, but less obvious trends were observed for posterior estimates of π. Under the same scenario as above, the posterior means of π, averaged over the 50 replications, were with 0.617 significantly higher for dominance effects, than the posterior means of π for additive effects, which were 0.566 and 0.556 for effects pertaining to dent and flint lines respectively. Thus, dominance effects had a prior probability to be shrunken to exactly zero that was approximately five percentage points higher than for additive effects.

Both π and ν can control model sparsity, π in a way similar to an indicator variable and ν by allowing for adaptive shrinkage. As observed by Pikkuhookana and Sillanpää (2009), compared to adaptive shrinkage, the influence of an indicator variable on overall sparsity is small when both sources are used simultaneously. This might explain why the differences for π were relatively less relevant than the differences for ν.

As argued by Yang and Tempelman (2012), specifying hyperprior distributions on the key hyperparameters allows them to be estimated from the data, which in turn allows BayesB to converge to the optimal solution, when a sufficient amount of data is used. For example, when π tends towards zero and ν becomes large, the solutions of BayesB would approximate those of GBLUP. Our results suggest that a GBLUP like model provided close to optimal solutions for additive effects but not necessarily for dominance effects. This property of BayesB, when formulated as suggested by Yang and Tempelman (2012), obviates the need to include GBLUP as a reference when a direct comparison of these methods is not a main objective.

Computationally, BayesB is considerably more demanding than GBLUP. However, as long as computations are feasible at all, computation time is a critical issue only when the computations have to be repeated many times, e.g., in elaborate simulation or cross-validation studies. When genomic prediction is just an intermediate step in a breeding program or academic study, the algorithm needs to be run only for a few times, maybe even only once. Then computation time will be less an issue when judging the relative preference of a method. Furthermore, the high computational demand for BayesB mainly stems from the high number of iterations used. We used these high numbers only as a safeguard to ensure convergence and high effective sample sizes on all parameters of interest. However, we observed that virtually identical results in terms of predicted genotypic values could be obtained when using only a fraction of the number of iterations (say 2,000 iterations with burn-in length of 1,000). Thus, when inference on individual parameters is of less interest, as is most often the case in genomic prediction studies, the computation time of BayesB could be dramatically reduced by lowering the number of iterations.

To summarize, BayesB yielded higher prediction accuracies than GBLUP under the best models and in the most relevant scenarios, and proved to be flexible enough to converge to an optimal solution for all types of effects. Based on our results, we can therefore recommend its use, despite its higher computational demands. For this reason, and because the general trends were similar for both methods, the following section on model comparison will focus on the results obtained with BayesB.

Model comparison

In order to esteem the observed differences between the models for a given combination of the other factors analyzed, we emphasize again that the data set acted as a blocking factor common to all models. The comparatively large fluctuations (measured as SD) in baseline accuracy between data sets (Table 1) therefore did not enter in the comparison of the differences between the models. In other words, even though the overall prediction accuracy was fluctuating across different runs, the in most cases significant differences between the models suggest that the relative superiority of the models vis-a-vis each other was stable.

Even though the importance of σ^2_{SCA} compared to that of $\sigma^2_{\mathrm{GCA}^d} + \sigma^2_{\mathrm{GCA}^f}$ was low, incorporating dominance effects into the model improved the prediction accuracies in all cases (Table 1). The greater improvement under the "convergent" scenario than under the "divergent" scenario can be explained by the increased importance of σ^2_{SCA} in the former case. The greater improvement with higher marker density (Table 1) suggests that estimation of dominance effects of markers profits relatively more from an increase in LD than the estimation of additive marker effects.

Modeling marker effects specific to a population (i.e. models S_1 and S_2) led generally to higher prediction accuracies than observed for models, where marker effects were population unspecific (i.e., models U_1 and U_2) (Table 1). However, the differences were only moderate and not in all cases statistically significant. The good performance of the unspecific models can be explained with the high LD across populations and the high congruency of linkage phases for loci pairs in close proximity (Fig. 4). Consequently, the largest differences between the two model classes were observed for a marker density of 0.3 Mbp^{-1}, a situation in which the LD across populations is at a considerably lower level than the LD within the populations, and the proportion of marker pairs with identical linkage phases in both populations is close to 50 %. For a marker density of 10 Mbp^{-1}, entailing a distance between markers and QTL $<$ 0.1 Mbp in the

simulations, both the LD across populations and the proportion of pairs with identical linkage phases were at a very high level. Moreover, models S_1 and S_2 require the estimation of a greater number of marker effects, which may reduce their estimation accuracy and thus acts as a penalty.

The larger differences between the four models under "divergent" parental populations could also be ascribed to LD. In our simulations, we determined the proportion of pairs of QTL and closest adjacent markers with identical linkage phase in both populations and found slightly lower values under the "divergent" scenario compared to what was found under the "convergent" scenario (data not shown). This could explain why it might be more beneficial to model marker effects as population specific under that scenario.

Our work has some parallels to a study conducted by Ibánez-Escriche et al. (2009) in the context of beef cattle breeding. They used BayesB for fitting models with breed-specific and unspecific additive marker effects to simulated crossbred data sets. Their ultimate goal was to use the marker effects for selection in the purebred parental populations. Their main results are in close agreement with ours, in that the differences between models in prediction accuracy were small; if the specific models had an advantage, then only under lower marker densities. However, there are rather fundamental differences between the situation encountered in beef cattle and hybrid maize breeding. In the latter case, the ultimate goal is not selection of parental genotypes but the identification of high-performing hybrids as such. Furthermore, the genetic makeup of the populations simulated by Ibánez-Escriche et al. (2009), and that of beef cattle breeding programs in general, differ fundamentally from the situation encountered in hybrid maize breeding and simulated by us. First, maize hybrids share identical copies of gametes with other hybrids and with their parents. Second, dominance is an important factor contributing to yield in maize hybrids. However, dominance was not considered as a factor by Ibánez-Escriche et al. (2009).

In summary, modeling marker effects as population specific was most beneficial under lower LD levels, as might be observed in diverse or exotic material (Riedelsheimer et al. 2012; Yan et al. 2009). In this study, the specific models had an clear advantage up to a marker density of 1 Mbp^{-1} (Table 1), i.e., up to an LD level between neighboring markers of $r^2 > 0.225$ within and $r^2 < 0.175$ across populations (Fig. 4a). Such levels of LD can easily be attained in the type of populations investigated here. However, reaching such levels of LD between neighboring markers would require SNP chips with much higher density or even whole sequence data for more diverse or exotic germplasm (Riedelsheimer et al. 2012; Yan et al. 2009).

Compared with other factors, such as the type of material to predict (T2, T1 or T0), the differences in prediction accuracy due to choice of model were rather small. Nevertheless, the differences were still sizable in most cases. Especially, the improvements achieved by incorporating dominance effects were large enough to be of practical importance. Furthermore, the choice of a more complex model is not associated with any additional costs. Therefore, also smaller gains in accuracy can be exploited for free.

Viewing the prediction accuracy as criterion for choice of the best model, we conclude that models incorporating dominance and population specific marker effects can provide better fits to the data than simpler models.

Implications for hybrid breeding

The square root of h^2 can be seen as the phenotypic equivalent of the prediction accuracy. It corresponds to the accuracy of predicting the genotypic value of hybrids from their phenotypes alone, without using any genomic or pedigree information. To put our results in perspective with the current approach in hybrid breeding, we therefore have to compare the observed prediction accuracies with $\sqrt{h^2} = 0.866$.

Focusing on model S_2 with effects estimated by BayesB, the prediction accuracy of T2 hybrids was always considerably higher than $h = 0.866$. This suggests that the genomic predicted genotypic values of T2 hybrids can be more accurate than prediction based on field trial data. Given the significantly higher accuracies when incorporating dominance effects, we further note that genomic prediction of T2 hybrids is also more accurate than GCA based prediction, which cannot accommodate for SCA. This is irrespective of whether the GCA values of the parental lines were obtained from field data or are themselves genomic predictions (Albrecht et al. 2011; Riedelsheimer et al. 2012).

Because 3/4 of the lines in our simulation were "evaluated", the T2 group was by far the largest of the three groups. At present, large breeding companies generate up to 10,000 DH lines per year and heterotic pool (Melchinger, personal communication). With these dimensions, the proportion of "evaluated" lines will be much smaller and the T1 and T0 groups will dominate the factorials.

The prediction accuracies observed for the T1 group with model S_2 were under the "divergent" scenario always higher than $h = 0.866$ and under the "convergent" scenario close to $h = 0.866$ for high LD levels (Table 1). Thus, promising T1 hybrids too could be identified with high accuracy using genomic prediction alone. Even for T0 hybrids, the prediction accuracies were not dramatically lower than $h = 0.866$ at high LD levels (Table 1).

It would be interesting to compare our results with predictions of hybrid performance based on pedigree information, as proposed by Bernardo (1996). However, pedigree information was not available for all our lines and not complete or reasonably deep for the remainder.

It need to be noted that practitioners are unlikely to solely rely on genomic predictions of hybrid performance. Therefore, implementing genomic prediction holds most promise as an initial stage in a multi-stage selection scheme. Accordingly, the number of available parental lines could be reduced to the parents of the most promising hybrids, based on the genomic prediction of their hybrid performance. This could be coupled with genomic prediction of line per se performance for resistance traits or traits related to the economics of seed production. In a second stage, promising experimental hybrids from factorial crosses of the remaining lines could then be evaluated in extensive field trials. As a side product, these field-evaluated experimental hybrids could be used to update and extend the training set. Further research is warranted to confirm our simulation results with experimental phenotypic data.

Acknowledgments This research was funded by the German Federal Ministry of Education and Research (BMBF) within the AgroClustEr Synbreed—Synergistic plant and animal breeding (FKZ: 0315528d).

References

Albrecht T, Wimmer V, Auinger HJ, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön CC (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350

Bates D, Maechler M, Bolker B (2011) lme4: linear mixed-effects models using S4 classes. http://CRAN.R-project.org/package=lme4, r package version 0.999375-39

Bernardo R (1996) Best linear unbiased prediction of maize single-cross performance. Crop Sci 36:50–56

Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 84:210–223

Charcosset A, Essioux L (1994) The effect of population structure on the relationship between heterosis and heterozygosity at marker loci. Theor Appl Genet 89:336–343

Clark S, Hickey JM, van der Werf JH (2011) Different models of genetic variation and their effect on genomic evaluation. Genet Sel Evol 43:18

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. 4th edn. Longmans Green, Harlow

Fischer S, Möhring J, Schön CC, Piepho HP, Klein D, Schipprack W, Utz HF, Melchinger AE, Reif JC (2008) Trends in genetic variance components during 30 years of hybrid maize breeding at the University of Hohenheim. Plant Breeding 127:446–451

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner EM, Hansen M, Joets J, Le Paslier MC, McMullen MD, Montalent P, Rose M, Schön CC, Sun Q, Walter H, Martin OC, Falque M (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6:e28,334

Gardner C (1963) Estimates of genetic parameters in cross-fertilizing plants and their implications in plant breeding. In: Statistical genetics and plant breeding. Comittee on Plant Breeding and Genetics of the Agricultural Board at the North Carolina State College Raleigh, NC, vol 982, pp 225–251

Gardner C, Lonnquist J (1959) Linkage and the degree of dominance of genes controlling quantitative characters in maize. Agron J 51:524–528

Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136:245–257

Hayes BJ, Pryce J, Chamberlain AJ, Bowman PJ, Goddard M (2010) Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genet 6:e1001,139

Ibánez-Escriche N, Fernando RL, Toosi A, Dekkers JCM (2009) Genomic selection of purebreds for crossbred performance. Genet Sel Evol 41:12

Kizilkaya K, Carnier P, Albera A, Bittante G, Tempelman R (2003) Cumulative t-link threshold models for the genetic analysis of calving ease scores. Genet Sel Evol 35:489–512

Labate J, Lamkey K, Lee M, Woodman W (1999) Temporal changes in allele frequencies in two reciprocally selected maize populations. Theor Appl Genet 99:1166–1178

Li Z, Sillanpää MJ (2012) Estimation of quantitative trait locus effects with epistasis by variational Bayes algorithms. Genetics 190:231–249

Maenhout S, De Baets B, Haesaert G (2010) Prediction of maize single-cross hybrid performance: support vector machine regression versus best linear prediction. Theor Appl Genet 120:415–427

Meuwissen TH, Goddard M (2010) Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185:623–631

Meuwissen TH, Hayes BJ, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Müller P (1991) A generic approach to posterior integration and Gibbs sampling. Technical report, Purdue University # 91-09

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290

Pikkuhookana P, Sillanpää MJ (2009) Correcting for relatedness in Bayesian models for genomic data association analysis. Heredity 103:223–237

Plummer M, Best N, Cowles K, Vines K (2010) coda: output analysis and diagnostics for MCMC. http://CRAN.R-project.org/package=coda, r package version 0.14-2

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/, ISBN 3-900051-07-0

Ramsey J, Ripley B (2010) pspline: penalized smoothing splines. http://CRAN.R-project.org/package=pspline, r package version 1.0-14

Reif JC, Gumpert FM, Fischer S, Melchinger AE (2007) Impact of interpopulation divergence on additive and dominance variance in hybrid populations. Genetics 176:1931–1934

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220

Schaefer J, Opgen-Rhein R, Zuber V, Silva APD, Strimmer K (2012) corpcor: efficient estimation of covariance and (partial) correlation. http://CRAN.R-project.org/package=corpcor, r package version 1.6.2

Schrag TA, Melchinger AE, Sørensen AP, Frisch M (2006) Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. Theor Appl Genet 113:1037–1047

Stich B, Melchinger AE, Piepho HP, Hamrit S, Schipprack W, Maurer HP, Reif JC (2007) Potential causes of linkage disequilibrium in a European maize breeding program investigated with computer simulations. Theor Appl Genet 115:529–536

Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–1299

Van Inghelandt D, Reif JC, Dhillon BS, Flament P, Melchinger AE (2011) Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. Theor Appl Genet 123:11–20

Xu S (2003) Estimating polygenic effects using markers of the entire genome. Genetics 163:789–801

Yan J, Shah T, Warburton M, Buckler E, McMullen M, Crouch J (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One 4:e8451

Yang W, Tempelman RJ (2012) A Bayesian antedependence model for whole genome prediction. Genetics 190:1491–1501

Zhong S, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a Barley case study. Genetics 182:355–364