Theor Appl Genet (2007) 114:1345–1355
DOI 10.1007/s00122-007-0521-5

ORIGINAL PAPER

Prediction of single-cross hybrid performance in maize using haplotype blocks associated with QTL for grain yield

Tobias A. Schrag · Hans Peter Maurer · Albrecht E. Melchinger · Hans-Peter Piepho · Johan Peleman · Matthias Frisch

Received: 26 October 2006 / Accepted: 4 February 2007 / Published online: 24 February 2007

Abstract Marker-based prediction of hybrid performance facilitates the identification of untested single-cross hybrids with superior yield performance. Our objectives were to (1) determine the haplotype block structure of experimental germplasm from a hybrid maize breeding program, (2) develop models for hybrid performance prediction based on haplotype blocks, and (3) compare hybrid performance prediction based on haplotype blocks with other approaches, based on single AFLP markers or general combining ability (GCA), under a validation scenario relevant for practical breeding. In total, 270 hybrids were evaluated for grain yield in four Dent × Flint factorial mating experiments. Their parental inbred lines were genotyped with 20 AFLP primer–enzyme combinations. Adjacent marker loci were combined into haplotype blocks. Hybrid performance was predicted on basis of single marker loci and haplotype blocks. Prediction based on variable haplotype block length resulted in an improved prediction of hybrid performance compared with the use of single AFLP markers. Estimates of prediction efficiency (R^2) ranged from 0.305 to 0.889 for marker-based prediction and from 0.465 to 0.898 for GCA-based prediction. For inter-group hybrids with predominance of general over specific combining ability, the hybrid prediction from GCA effects was efficient in identifying promising hybrids. Considering the advantage of haplotype block approaches over single marker approaches for the prediction of inter-group hybrids, we see a high potential to substantially improve the efficiency of hybrid breeding programs.

Communicated by H. C. Becker.

Tobias A. Schrag and Hans Peter Maurer contributed equally to this work.

T. A. Schrag · H. P. Maurer · A. E. Melchinger (✉) · M. Frisch
Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim,
70593 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de

H.-P. Piepho
Bioinformatics Unit of the Institute for Crop Production and Grassland Research, University of Hohenheim,
70599 Stuttgart, Germany

J. Peleman
Keygene, P.O. Box 216,
6700 AE Wageningen, The Netherlands

Introduction

Prediction methods for single-cross performance have the potential to substantially improve the efficiency of maize (*Zea mays* L.) hybrid breeding programs. Several hundred single-cross combinations could potentially be generated by breeders each year. However, due to expensive and time-consuming field trials, only a small fraction of all possible single crosses can be tested. Performance prediction of single-cross hybrids utilises available data at no costs of additional trials and facilitates the identification of untested single-cross hybrids with superior yield performance.

General combining ability (GCA) estimates of the parental lines provide an established and simple approach to predict hybrid performance (Cockerham 1967; Melchinger et al. 1987). Prediction based on GCA alone ignores specific combining ability (SCA), which is related to heterosis and constitutes an important component of hybrid performance (Gardner and Eberhart 1966). Therefore,

marker-based approaches were developed with the aim of improving the prediction of inter-group hybrids. However, estimates of genetic distances between the parental lines using unselected DNA markers for prediction of inter-group hybrids were not promising (Melchinger 1999). This was explained with quantitative genetics theory by Charcosset and Essioux (1994). Extending the best linear unbiased prediction (BLUP) approach (Bernardo 1994, 1996) with marker data in addition to the trait data (TM-BLUP) resulted only in marginal improvements for predicting single-cross performance (Bernardo 1998, 1999). Vuylsteke et al. (2000) presented a linear regression approach to predict hybrid performance and SCA for grain yield using marker-based estimates of genotypic value for inter-group hybrids. Recently, this approach was enhanced and validated for the prediction of grain yield and grain dry matter content in four factorial experiments (Schrag et al. 2006).

In the latest two publications, the problem of an increased rate of false-positives due to multiple testing was addressed by a stringent threshold for the comparison-wise error, but not by controlling the experiment-wise error. Furthermore, this approach does not account for correlation of marker information, which can be the result of (1) close linkage between marker loci, particularly with high marker densities, (2) closely related individuals, as occur in breeding programs, and (3) sampling a limited number of genotypes. As a consequence, the effect of a quantitative trait locus (QTL) linked to a series of correlated markers can be inflated and, thereby, increase the prediction error. Second, ignoring the correlation of markers for example with the Bonferroni method results in an overly stringent adjustment for multiple testing and thereby reduces the power of detecting QTL. These problems can be addressed by combining highly correlated adjacent markers into haplotype blocks.

Finding block boundaries by specifying a fixed block length (Jansen et al. 2003) is a straightforward approach. However, it ignores the correlation structure of the actual marker data. Data-driven strategies account for the given marker data and determine haplotype block boundaries by (1) optimising for minimum linkage disequilibrium (LD) between blocks and maximum LD within blocks (Gabriel et al. 2002), (2) assessing the haplotype diversity within blocks (Patil et al. 2001; Zhang et al. 2002), or (3) simultaneously using information about LD decay between blocks and diversity of haplotypes within blocks (Anderson and Novembre 2003). Such data-driven approaches were developed with the aim of assisting association-based methods for mapping of disease genes by the use of single nucleotide polymorphism (SNP) data from the human genome. These strategies optimise the block boundaries in order to explain a high proportion of the haplotype diversity with a low number of SNPs, called ''haplotype tagging SNPs''. However, for marker-assisted prediction of hybrid performance, the aim is to reduce the number of estimated parameters while utilising the total haplotype diversity described by all markers. Such criteria to find haplotype block boundaries and predict the hybrid performance on the basis of haplotype blocks have not been investigated hitherto.

Our objectives were to (1) determine the haplotype block structure of experimental germplasm from a hybrid maize breeding program, (2) develop models for hybrid performance prediction based on haplotype blocks, and (3) compare hybrid performance prediction based on haplotype blocks with other approaches, based on single amplified fragment length polymorphism (AFLP) markers or GCA, under a validation scenario relevant for practical breeding.

Materials and methods

Phenotypic data

The experimental design and biometrical analysis of the phenotypic data was described in detail by Schrag et al. (2006). Briefly, we analysed four Dent × Flint factorial mating experiments (14 × 7, 11 × 4, 14 × 6, 11 × 4), further referred to as Exps. 1 to 4. The matings were produced from 52 maize elite inbred lines developed within the breeding program of the University of Hohenheim, Germany. Eight Dent lines and six Flint lines were included in multiple factorials. Each factorial was evaluated in field trials at four to six locations in Germany under diverse agroecological conditions. Hybrid performance of the crosses was recorded for grain yield in Mg ha^{-1} adjusted to 155 g kg^{-1} grain moisture. Adjusted entry means and effective error mean squares (Cochran and Cox 1957) of each trial were used to calculate the variance components as well as GCA and SCA effects of the factorial mating designs (Comstock and Robinson 1952) for each experiment in the combined analysis of variance across locations. For the GCA prediction approach, hybrid performance of untested hybrids was predicted using the GCA estimates, as described in Eqs. 2 and 3 by Schrag et al. (2006).

Molecular data

The inbred lines were assayed for AFLP markers based on published protocols (Vos et al. 1995). Genotyping was conducted with 20 AFLP primer–enzyme combinations, described by Schrag et al. (2006). It was assumed that each AFLP band position corresponds to a locus and presence or absence of the band was scored as two alleles of that locus.

Positions of mapped AFLP bands were obtained from an integrated AFLP map (Peleman et al. 2000; Vuylsteke et al. 1999). Out of 910 mapped markers, subsets were specifically selected for the parental inbred lines of each experiment. Markers were only included if they (1) were polymorphic and (2) exhibited less than 30% missing observations in the specific set of inbred lines. For each pair of linked markers, Fisher's exact test was calculated to test genotypic LD (Zaykin et al. 1995) at a significance level of $\alpha = 0.05$. For the underlying Monte Carlo method, 17,000 replications were used (Guo and Thompson 1992).

Analysis of haplotype block structure

To consider a group of marker loci as a unit, adjacent marker loci were combined and regarded as haplotype blocks. The first and the last marker locus of a haplotype block constituted the haplotype block boundaries. Within each haplotype block, the observed sets of marker alleles were regarded as haplotype alleles (Fig. 1). For haplotype blocks comprising n bi-allelic markers, the possible number of haplotype alleles is 2^n. Therefore, haplotype blocks were regarded as multi-allelic markers. Haplotype alleles, which included missing data at underlying marker loci, were defined as missing. Thus, for blocks of length two or longer, missing marker observations caused the information loss of the remaining marker observations in the considered haplotype allele.

Separately for each experiment, the haplotype block structure was analysed with three different methods. For the HB1 approach, haplotype block boundaries were determined by specifying a fixed block length of four adjacent mapped markers along the chromosome (Jansen et al. 2003). This straightforward approach was included to allow comparisons with more elaborate, data-driven approaches HB2 and HB3, which determined the block boundaries by performing the following steps for each chromosome: (1) generating all potential blocks of adjacent markers with block lengths between one and a defined maximum block length, (2) determining the number of haplotype alleles within each potential block, (3) discarding all potential blocks which violated the restrictions specifically defined for HB2 and HB3 (the restrictions will be described in the following paragraph), (4) determining an optimum haplotype block solution with the lowest chromosome-wise haplotype allele number using Dijkstra's shortest path algorithm (Dijkstra 1959), and (5) assigning the haplotype alleles.

For the HB2 approach, the maximum block length was four markers to allow comparison with the HB1 approach. Furthermore, potential blocks were discarded if at least one marker included in the potential block had missing observations. With this restriction, markers affected by missing observations were considered as blocks with the length of one. For the HB3 approach, maximum block length was 15 markers, and solutions were retained only if strong LD was observed within the investigated haplotype block. This was achieved by discarding potential blocks, where the median of Fisher's exact P-value across all pair-wise comparisons between markers in the investigated haplotype block was higher than $\alpha = 0.05$.

Prediction based on total effects of associated markers

For each experiment, markers associated significantly with hybrid performance were identified separately. For each marker the genotypic class of a hybrid was determined by the marker of the homozygous parental inbreds. Modifying the approach of Vuylsteke et al. (2000), for each marker the genotypic effects were estimated and tested across all hybrids in the estimation set with the following model:

$$y_{ck} = \mu + \tau_c + e_{ck} \quad (1)$$

where y_{ck} = mean performance of the k-th hybrid of genotypic class c; μ = grand mean; τ_c = effect of genotypic class c with zero-sum constraint $\Sigma\tau_c = 0$; and $e_{ck} \sim N(0; \sigma^2_e)$ residual error of y_{ck}. The effect of marker genotypic class on hybrid performance was tested with an F test of H_0: $\tau_1 = \tau_2 = \cdots = \tau_C$ at a false discovery rate of 5% (Benjamini and Hochberg 1995). Across all markers, which were significantly associated with hybrid performance, the genotypic value for each hybrid was then estimated by the sum of its τ_c. These genotypic value estimates were considered as the total effects of associated markers (TEAM) and were used as predictor for hybrid performance in a simple linear regression:

$$y_{ij} = a + \mathrm{TEAM}_{ij} \cdot b \quad (2)$$

where y_{ij} = mean performance of the hybrid between parental inbreds i and j; TEAM_{ij} = total effects of associated markers for hybrid ij. The TEAM values of the test set hybrids were determined using the τ_c estimates obtained from the estimation set. In cases where for a given marker (1) the τ_c of the genotypic class c could not be estimated due to the lack of observations for genotypic class c in the estimation set or (2) the genotypic class of the test set hybrid was unknown due to missing marker data, the average of the τ_c estimates weighted by the number of observations k_c in the estimation set was used as a substitute. Hybrid performance of the hybrids in the test set was then predicted by using their TEAM values in the simple linear regression (Eq. 2), for which parameters a and b were obtained in the estimation set. The TEAM procedure was applied in the same manner to (1) single

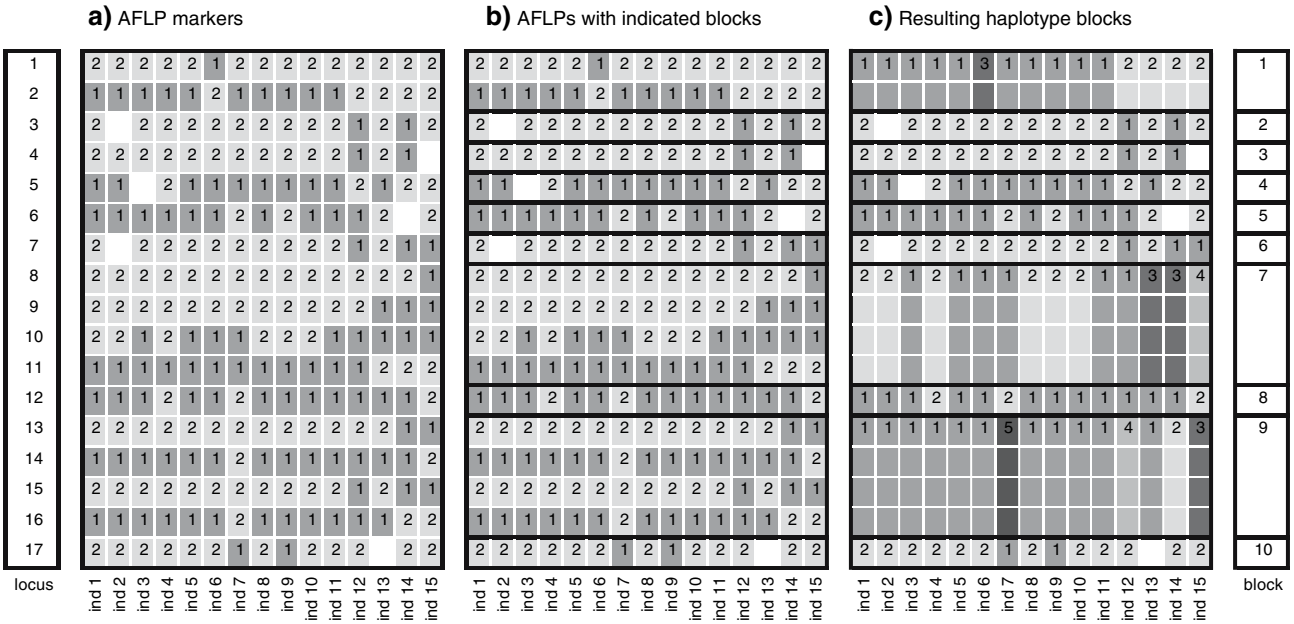

Fig. 1 Conversion of AFLP allele data into haplotype alleles, exemplarily showing the HB2 approach applied to a segment of 17 loci on chromosome 1 with marker data from Exp. 4. Missing AFLP observations were indicated as blank cells, observed alleles were colored in *grey*. On basis of the underlying AFLP data (**a**), block boundaries were identified by the haplotype block algorithm (**b**). Within resulting haplotype blocks, the observed sets of marker alleles were regarded as haplotype alleles (**c**)

AFLP markers (SM) as well as (2) haplotype blocks HB1 to HB3, which were regarded as multi-allelic markers.

Prediction based on multiple linear regression

In a second approach, prediction of hybrids based on estimation of QTL effects was regarded as a multiple linear regression (MLR) problem:

$$y_{ij} = \mu + \sum_{m,c} a^{m,c} \cdot x_{ij}^{m,c} + e_{ij} \qquad (3)$$

where y_{ij} = mean performance of the hybrid between parental inbreds i and j; μ = intercept; $a^{m,c}$ = effect of genotypic class c at marker locus m; $x_{ij}^{m,c}$ = indicator variable [0,1] for the genotype c at marker locus m of hybrid ij; and $e_{ij} \sim N(0; \sigma^2_e)$ residual error of y_{ij}. The summation is over genotypic classes c of markers m affecting the trait. Markers were added to the reduced model using forward selection. At each step, the most significant marker was added to the model until no other marker had an F value of 4.0 or higher. Additionally, a forward selection procedure was performed with an alpha-to-enter of 0.05 and 0.20, respectively, divided by the appropriate number of markers in each experiment. Further, we employed the Schwarz Bayesian criterion for a genome-wide forward selection procedure and a two-step forward selection procedure. For the latter approach, variables were forward selected separately for each chromosome in the first step, which then provided the shortlist for the genome-wide forward selection in the second step. In those cases where for a given marker m the effect $a^{m,c}$ of the genotypic class c could not be estimated due to the lack of observations for c in the estimation set, the average of all estimated $a^{m,c}$ for marker m, weighted by the number of observations in the estimation set, was used as a substitute effect in the prediction of test set hybrids. The MLR procedure was applied in the same manner to SM as well as HB1 to HB3. Model selection and prediction was performed with software R using routines from its 'stats' package (R Development Core Team 2004).

Evaluating the efficiency of prediction models for hybrid performance

Cross-validation was performed with 100 randomised replications per experiment to evaluate the efficiency of the GCA- and marker-based prediction models. For each cross-validation run, the entire factorial data set of an experiment was divided into an estimation set and test set. In order to mimic the situation in plant breeding, five Dent lines and three Flint lines were used as testers. Therefore, in each cross-validation run, five Dent and three Flint testers were chosen at random (Fig. 2, exemplarily for Exp. 1). The chosen testers were crossed with all lines of the respective opposite heterotic group, forming the estimation set. All remaining crosses formed the test set. The estimation set

	D01	D02	D03	D04	**D05**	D06	**D07**	D08	D09	D10	**D11**	D12	**D13**	D14
F01	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES
F02	ES	TS	TS	TS	ES	TS	ES	TS	TS	TS	ES	TS	ES	TS
F03	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES
F04	ES	TS	TS	TS	ES	TS	ES	TS	TS	TS	ES	TS	ES	TS
F05	ES	TS	TS	TS	ES	TS	ES	TS	TS	TS	ES	TS	ES	TS
F06	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES	ES
F07	ES	TS	TS	TS	ES	TS	ES	TS	TS	TS	ES	TS	ES	TS

Fig. 2 Randomised subdivision of the Exp. 1 data set used for cross validation (exemplarily showing one possible randomisation). In each of the 100 cross-validation runs, five of the 14 Dent lines (D01–D14) and three of the seven Flint lines (F01–F07) were randomly chosen as testers. The chosen testers were crossed with all lines of the respective opposite heterotic group, forming the estimation set (ES). All remaining crosses formed the test set (TS)

was used for QTL detection and parameter estimation in each prediction approach. In the test set, predictions derived from the estimation set were tested for their validity by determining the squared correlation coefficient (R^2) and the square root of mean square deviation (RMSD) between observed and predicted hybrid performance values for each prediction approach.

Results

Biometrical analysis of field data

The estimates of variance components for Exps. 1–4 were given by Schrag et al. (2006), and will be presented briefly (Table 1). The SCA variance component for grain yield for Exp. 1 (0.170) was remarkably higher than for the remaining experiments (0.029–0.072). When averaging GCA variance over Flint and Dent, the ratio of SCA to GCA variance ranged from 0.19 to 1.12, with highest values for Exp. 1.

Analysis of molecular data

The marker subsets, which were specific for Exps. 1–4, comprised 720, 674, 638, and 583 mapped AFLP markers, respectively (Table 1). The HB3 approach resulted in the highest number of blocks (489–584), followed by HB2 (382–474) and HB1 (150–184). Accordingly, the average block lengths were shortest for HB3 with 1.16–1.29 markers per block, followed by HB2 with 1.36–1.62 markers per block. For HB1, the average block lengths were longest with 3.89–3.92 markers per block, which differed from 4.0 markers per block, as the number of AFLP markers per chromosome was not always a multiple of four. The average number of alleles per marker or haplotype block was 2.0 for SM in all experiments, followed by HB3 (2.12–2.26), HB2 (2.43–2.79), and HB1 (5.32–6.22). The average number of genotypic classes per marker or haplotype block was 3.0 for SM in all experiments, followed by HB3 (3.40–4.03), HB2 (4.63–6.14), and HB1 (17.52–23.46). Ranking the approaches for block length, number of alleles, and number of genotypic classes resulted in the same order across all experiments: SM, HB3, HB2, and HB1. In the SM subsets, between 2.0 and 3.8% of the AFLP observations were missing. For HB1, between 7.4 and 13.4% of the underlying AFLP marker observations were missing, whereas for HB2 the percentages were identical to SM. Consequently, between 5.4 and 9.6% of the AFLP markers were additionally discarded in HB1 compared with HB2. The percentage of linked marker pairs in LD for the populations of parental inbreds in Exps. 1–4 was 9.0, 5.9, 8.2, and 6.5%, respectively.

Selected markers or haplotype blocks for prediction

For the TEAM-based approaches, the average numbers of selected markers or haplotype blocks (Table 2) in each experiment were highest for SM (126.4–219.9), followed by HB2 (99.5–195.1), HB3 (91.7–182.8), and HB1 (58.3–116.0). For the MLR-based approaches, the average numbers of selected markers or haplotype blocks were distinctly lower compared with the TEAM-based approaches, with highest values for SM (6.9–9.0) and HB3 (7.0–8.6), followed by HB2 (5.4–7.9), and again lowest values for HB1 (1.1–2.2). Ranking the SM and HB1–HB3 approaches for the number of selected markers resulted in a similar order as for block length, number of alleles, and number of genotypic classes among both the TEAM-based and the MLR-based approaches.

Efficiency of prediction models for hybrid performance

Across all approaches, values of R^2 (Fig. 3) were lowest for Exp. 1, intermediate for Exp. 3, and highest for Exps. 2 and 4. In Exp. 1, the median R^2 (Table 2) was highest for

Table 1 Experimental setup, field data results and marker data results for Exps. 1–4

	Experiment			
	1	2	3	4
Experimental setup				
Dent × Flint	14 × 7	11 × 4	14 × 6	11 × 4
Parental lines	21	15	20	15
Hybrids	98	44	84	44
Variance components for grain yield (Mg ha^{-1})				
SCA	0.170	0.072	0.066	0.029
GCA	0.151	0.383	0.155	0.149
Number of markers or haplotype blocks				
SM	720	674	638	583
HB1	184	172	164	150
HB2	444	474	470	382
HB3	584	583	495	489
Average block length (in markers)				
HB1	3.91	3.92	3.89	3.89
HB2	1.62	1.42	1.36	1.53
HB3	1.23	1.16	1.29	1.19
Average number of alleles per marker or haplotype block				
SM	2.00	2.00	2.00	2.00
HB1	6.22	5.65	6.11	5.32
HB2	2.79	2.52	2.43	2.62
HB3	2.25	2.12	2.26	2.16
Average number of genotypic classes per marker or haplotype block				
SM	3.00	3.00	3.00	3.00
HB1	23.46	19.66	22.95	17.52
HB2	6.14	5.08	4.63	5.40
HB3	3.97	3.40	4.03	3.54
Missing AFLP marker observations (%)				
SM	2.0	3.8	2.7	3.0
HB1	7.4	13.4	9.8	10.8
HB2	2.0	3.8	2.7	3.0
HB3	2.8	5.0	4.6	4.2

Experimental setup comprised the number of parental Dent and Flint lines, the total number of parental lines, and the number of hybrids. Variance components for specific combining ability (SCA) and general combining ability (GCA) were determined from field data and averaged over Dent and Flint. From cross-validation with 100 sampling rounds, the following averages were obtained for single markers (SM) and haplotype blocks (HB1-HB3): number of amplified fragment length polymorphism (AFLP) markers or haplotype blocks, block length, number of alleles, number of genotypic classes, and proportion of missing AFLP marker observations underlying the haplotype blocks

HB2-TEAM (0.466) and GCA (0.465), whereas for the MLR approaches, the median R^2 was clearly lower (0.305–0.370). In Exp. 2, GCA had the highest median R^2 (0.888) and a lower variance of R^2 values compared with the marker-based approaches. Among the marker-based procedures, the median R^2 was highest for HB3-TEAM (0.843) and HB2-MLR (0.843). For HB1 the median R^2 was clearly lowest, both for HB1-TEAM (0.755) and HB1-MLR (0.689). For Exp. 3, again the median R^2 for GCA was highest (0.754) and showed a low variance of R^2 values. Among the marker-based procedures, highest median R^2 were obtained with the MLR-based approaches HB3-MLR (0.702) and SM-MLR (0.689). However, the median R^2 of HB1-MLR was very low (0.363) compared with all other approaches in Exp. 3. Outliers with very low R^2 values were more pronounced for the MLR-approaches than for the TEAM-approaches. For Exp. 4, the differences between the approaches were clearly smaller compared with Exps. 1–3. The median R^2 for GCA was highest (0.898), followed by HB2-MLR (0.889) and HB1-MLR (0.888). Across all experiments, median R^2 of HB2 was generally higher than for HB1, and median R^2 of GCA was generally higher or equal compared with the marker-based prediction procedures. In general, the results for RMSD (Table 2) were very similar compared with those for R^2. However, in Exp. 4, the advantage of the HB1 and HB2 approaches in comparison with SM and HB3 was more pronounced for RMSD. In Exp. 2, outliers with high RMSD values for the MLR-based predictions (data not shown) were more distinct compared with the corresponding R^2 outliers.

Discussion

Haplotype blocks have been suggested as a means for association mapping (Anderson and Novembre 2003). We took this idea one step forward to improve the marker-based prediction of hybrids from germplasm of a commercial breeding program. The approach is linked to the idea that if parents were derived from few ancestors, the number of different haplotypes is expected to be smaller than the number of parents, thus enabling a large reduction in the number of estimated parameters (Jansen and Stam 1994). The correlation between alleles at different marker loci in a population is referred to as LD (Flint-Garcia et al. 2003). LD is a measure that highly depends on the genetic structure of the population or breeding pool from which the individuals were sampled. In contrast to studies, which were based on genetically diverse material of broad geographic origin and a large number of heterotic groups (Tenaillon et al. 2001), in our study the inbred lines originated from one commercial breeding program using the Dent/Flint heterotic pattern. The germplasm was derived from a limited number of ancestors, and relatedness, population stratification, and genetic drift can be regarded as the main forces for generating LD. Maurer et al. (2006) and Stich et al. (2006) analysed and thoroughly discussed the conditions in commercial breeding germplasm, result-

Table 2 Prediction results from cross-validation with 100 sampling rounds for Exps. 1–4

	Experiment			
	1	2	3	4
Average number of selected markers or haplotype blocks				
SM-TEAM	182.1	219.9	175.2	126.4
HB1-TEAM	87.6	116.0	80.2	58.3
HB2-TEAM	148.2	195.1	151.5	99.5
HB3-TEAM	148.0	182.8	143.5	91.7
SM-MLR	8.5	6.9	9.0	8.8
HB1-MLR	2.0	1.1	1.6	2.2
HB2-MLR	6.2	5.5	7.9	5.4
HB3-MLR	8.3	7.0	8.6	8.2
Median R^2				
SM-TEAM	0.418	0.829	0.636	0.857
HB1-TEAM	0.439	0.755	0.632	0.839
HB2-TEAM	0.466	0.822	0.640	0.856
HB3-TEAM	0.426	0.843	0.649	0.847
SM-MLR	0.346	0.837	0.689	0.869
HB1-MLR	0.305	0.689	0.363	0.888
HB2-MLR	0.322	0.843	0.631	0.889
HB3-MLR	0.370	0.833	0.702	0.861
GCA	0.465	0.888	0.754	0.898
Median RMSD				
SM-TEAM	0.484	0.375	0.338	0.270
HB1-TEAM	0.481	0.436	0.341	0.231
HB2-TEAM	0.468	0.398	0.337	0.232
HB3-TEAM	0.486	0.390	0.340	0.257
SM-MLR	0.549	0.385	0.309	0.212
HB1-MLR	0.555	0.455	0.440	0.162
HB2-MLR	0.540	0.379	0.344	0.175
HB3-MLR	0.565	0.389	0.290	0.226
GCA	0.465	0.301	0.276	0.184

Average number of selected markers or haplotype blocks, median squared correlation coefficient (R^2) and median square root of mean square deviation (RMSD) between observed and predicted hybrid performance were given. Predictions were based on the total effects of associated markers (TEAM) and on multiple linear regression (MLR), using single markers (SM) or haplotype blocks (HB1–HB3). Additionally, predictions were performed from estimates of general combining ability (GCA)

ing in high levels of LD and extended haploblocks. For example, in the latter study, 35 Flint and 37 Dent inbred maize lines exhibited extended AFLP haplotype blocks with less than 100 blocks across the whole genome. Consequently, the chosen set of AFLP markers is expected to adequately cover the limited number of haplotype blocks for the material investigated in our study.

Regardless of its origin, the correlation between markers affects the marker-based prediction of hybrids performance, as (1) for the TEAM approach, series of correlated markers are expected to overestimate the contributed effects of linked QTL, thereby increasing the prediction error, and (2) for the MLR approach, the number of parameters to be estimated for prediction is inflated. These issues were addressed by combining highly correlated adjacent markers into haplotype blocks.

In several publications from the field of human genomics, approaches were suggested to find haplotype block boundaries, using information about (1) LD decay between blocks and (2) diversity of haplotypes within blocks (Gabriel et al. 2002; Zhang et al. 2002). These strategies aim to identify a small number of so-called ''haplotype tagging SNPs'' (Johnson et al. 2001), with which the haplotype diversity can be explained sufficiently. This would allow the conduction of association-based mapping of disease genes in a more cost-effective way. For marker-based prediction of hybrid performance, however, the aim is to utilise the total haplotype diversity while reducing the number of parameters to be estimated for hybrid prediction. Therefore, we developed procedures to optimise the block boundaries with respect to a minimum overall haplotype allele number.

Procedures to find haplotype block boundaries

The straightforward HB1 approach considered the map order of markers. However, it ignored the actual allele data. Furthermore, in situations where the haplotype block of a given individual was affected by missing marker data, the resulting haplotype block allele could not be determined. This resulted in a loss of 5.4–9.6% of the AFLP marker information, as available marker data in affected block alleles could not be utilised. Therefore, we proposed a data-driven approach HB2, which (1) defined block boundaries in dependence on the chromosome-wise number of alleles and (2) restrictively defined block boundaries to minimise the information loss due to missing marker observations. The maximum block length for HB2 was limited to four markers to allow comparison with HB1.

The HB2 approach indirectly accounted for the correlation of markers, since combining correlated markers into a block reduces the allele number and would therefore be preferred by the algorithm. In contrast, the HB3 approach (1) allowed for a higher maximum block length of 15 markers, and (2) directly considered pair-wise LD, which was determined in the population of parental inbreds separately for each experiment. In cases of map regions with high marker density, this would facilitate the combination of a higher number of adjacent markers into one block, thus allowing a more accurate representation of the actual haplotype block structure existing in the germplasm of the parents.

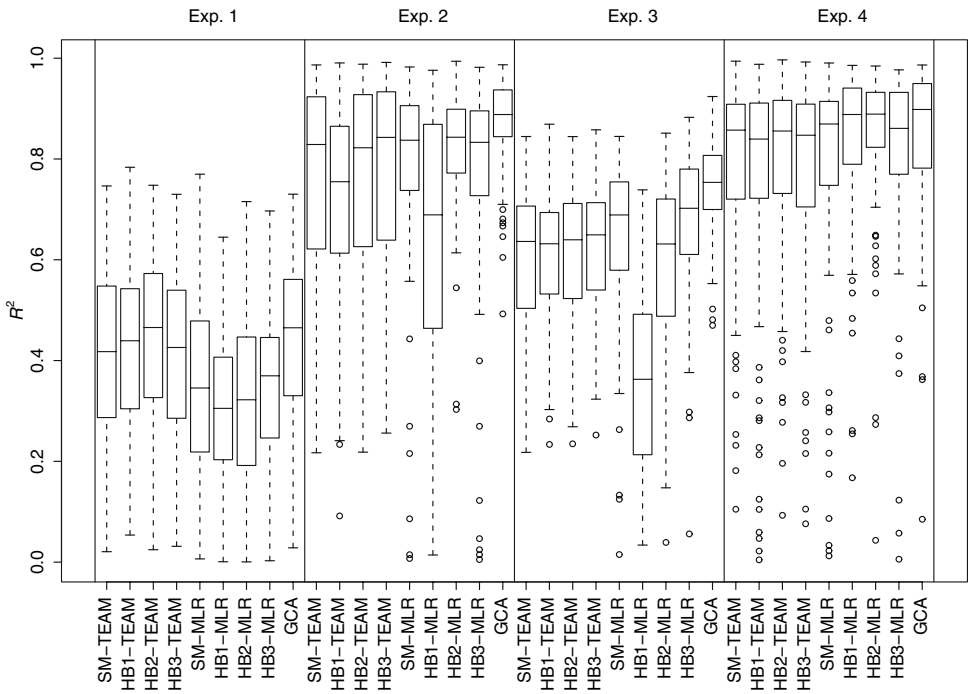

Fig. 3 Squared correlation coefficients (R^2) between observed and predicted hybrid performance for Exps. 1–4, obtained by cross-validation with 100 sampling rounds. Predictions were based on the total effects of associated markers (TEAM) and on multiple linear regression (MLR), using single markers (SM) or haplotype blocks (HB1–HB3). Additionally, predictions were performed from estimates of general combining ability (GCA). Boxplots were based on Tukey's five number summary

The marker order between genetic and physical maps can be inconsistent, which could impair the identification of haplotype blocks. However, misordering of loci within haplotype blocks does not affect any of the presented haploblock finding approaches. Misordering of loci across haplotype blocks could potentially disrupt the structure into shorter blocks. Even then the data-driven approaches HB2 and HB3 would robustly minimize the overall allele number for the given marker data. In addition, the underlying proprietary map (Peleman et al. 2000; Vuylsteke et al. 1999) was developed by integrating multiple mapping populations, resulting in a reduced probability of locus misorder between distant markers.

Prediction based on total effects of associated markers

The prediction approach presented by Vuylsteke et al. (2000) depends on bi-allelic marker data, thus cannot be applied to multi-allelic marker data such as SSRs or haplotype blocks. In their approach, the total contribution of selected markers (TCSM) for a given hybrid was obtained by summing up the additive and dominance effects across all selected markers. For the case of bi-allelic marker data, the genotypic value estimates TCSM and TEAM are perfectly correlated and therefore resulting in identical predictions of hybrids performance. However, with the modification of using effects τ_c of genotypic classes instead of additive and dominance effects, more than three genotypic classes can be accounted for. In this way, the TEAM-based prediction procedure can be extended from bi-allelic to multi-allelic marker data such as haplotype blocks and SSRs.

Prediction based on multiple linear regression

By defining haplotype blocks as blocks of adjacent markers, only the correlation between tightly linked markers is considered. To take genome-wide correlation of markers into account, an additional approach was examined. Sequential methods for multiple linear regression such as forward selection are appropriate approaches in situations where multicollinearity among the variables exists and the number of independent variables (i.e. markers) is large compared with the number of observations (i.e. hybrids). In a step-by-step procedure, variables are included into the model only if they significantly increase the variation explained by the enhanced model. With such an approach, markers are disregarded if correlated with already included markers. An F-to-enter forward selection approach was applied, but this does not ascertain the nominal alpha level (Piepho and Gauch 2001). This problem of compromised type-I error rates was addressed by additionally performing a forward selection procedure where the desired alpha level was divided by the number of variables (results not shown).

Model selection criteria such as the Schwarz Bayesian criterion provide means of comparing the goodness of fit of competing models while taking into account the principle of parsimony, and were recommended by Piepho and Gauch (2001) based on a simulation study of QTL mapping scenarios. The Schwarz Bayesian criterion was considered

in (1) a genome-wide forward selection procedure and (2) a two-step forward selection procedure (results not shown). However, when examining the prediction efficiencies of these various multiple linear regression approaches across all experiments, the prediction efficiencies of the *F*-to-enter procedure were highest or among the highest, and were therefore chosen as a reference to compare the multiple linear regression approaches with the TEAM-based prediction approaches.

Validation of predictions

In cross-validation, samples are repeatedly assigned at random to the estimation set or test set. In practical breeding programs for inter-group hybrids, the inbred lines from one heterotic group are crossed with a limited number of tester lines from the opposite group. With a leave-one-out validation, in each validation run only one hybrid was removed from the entire data set for parameter estimation, as published by Vuylsteke et al. (2000) and Schrag et al. (2006). Thereby, the number of parental tester lines was assumed overly high compared with the practical breeding situation, where a distinctly smaller proportion of hybrids in an experiment is available. Thus, for the sampling procedure in this work, parental lines instead of hybrid crosses were chosen as the sampling unit, allowing the set up of a validation procedure as close as possible to the practical situation in breeding programs.

Advantages of data-driven haploblock approaches for prediction

The prediction efficiency of the HB2-based approach was at least similar but in most cases clearly higher compared with the HB1-based approach. These results appear to be contrary to the lower number of haplotype blocks and haplotype alleles obtained for HB1, which (1) for the TEAM-based approaches reduces an overly stringent adjustment for multiple testing, thus increasing the power for QTL detection, and (2) for the MLR-based approaches suggests a reduction in the number of estimated parameters. However, the HB1 haplotype blocks exhibited very high numbers of alleles per locus, caused by using fixed block lengths and ignoring the correlation structure of the actual marker data. Consequently, the numbers of genotypic classes per locus (Table 1) and also the total numbers of genotypic classes were extremely high for HB1 compared with the SM, HB2, and HB3 approaches.

For such high numbers of genotypic classes (i.e. variables), as occurred with HB1, the numbers of observations in the factorial experiments were rather low, so that genotypic effects were often estimated with very few observations or not estimable at all. As a consequence for the TEAM-based approach, the genotypic values of the hybrids were poorly estimated. And for the MLR-based approach, only very few loci could be added to the model due to the high number of effect estimates per locus. This can be observed in Exps. 2–3, where only 1.1–1.6 blocks were selected on average by HB1-MLR, in contrast to 6.9–9.0 blocks for SM-MLR, and 5.5–7.9 blocks for HB2-MLR. Consequently, the prediction efficiencies of HB1-MLR were extremely poor for these experiments, compared with the remaining MLR approaches. Additionally, the use of fixed block lengths increases the risk of combining loci with effects differing in size and direction into one block. This may further reduce the power to detect significant haplotype blocks and thereby increase the prediction error.

The issues related to the use of fixed block lengths were addressed by the HB2 approach, which resulted in distinctly better predictions of hybrid performance. The HB2 approach indirectly accounted for the correlation of markers by optimising the block boundaries with respect to a minimum overall allele number. Furthermore, the approach avoided the propagation of marker information loss due to haplotype blocks. When assigning unique alleles to haplotypes, those haplotypes, which included missing data at underlying marker loci were defined as missing. Therefore, between 5.4 and 9.6% of the AFLP markers were additionally discarded in HB1 compared with HB2. This marker information loss is even more detrimental for the MLR-based procedures, since loci affected by missing observations were excluded from the analyses. As a result, the number of available HB1 haplotype blocks was strongly reduced, being an additional cause for the poor prediction efficiencies of HB1-MLR in Exps. 2–3. Summarizing, the use of haplotype blocks to improve hybrid prediction is only effective if the actual marker data of the involved genotypes is regarded in the process of finding block boundaries. Furthermore, a larger number of observed hybrids is needed to sufficiently estimate the marker effects as the basis for hybrid prediction.

Comparing haplotype blocks with single markers for prediction

A central question of this study was whether accounting for correlation between markers by means of haplotype blocks could improve the marker-based prediction of hybrid performance. Among the TEAM-based approaches, prediction efficiencies for HB2 were similar or better compared with the SM approach. Also, the HB3 approach was predicting the hybrid performance comparable with or better than the SM approach in most cases, whereas the HB1 approach resulted in poor prediction efficiencies. We conclude that the TEAM approach benefits from less stringent adjustment

for multiple testing and from avoidance of inflated QTL effects by means of haplotype blocks, but only if the actual marker data is considered for finding the block boundaries.

Among the MLR-based approaches, the prediction efficiencies obtained with the HB3 blocks were higher (Exps. 1 and 3), similar (Exp. 2), or only slightly lower (Exp. 4) compared with the SM approach. For the HB2-based procedures, the prediction efficiencies were ambiguous compared with SM, and for the HB1-based approach the prediction efficiencies were clearly lower. Thus, the most promising approach for MLR-based prediction was HB3, which directly considered LD and allowed for block lengths of up to 15 markers. Summarizing, in comparison with the SM approach, the prediction efficiencies of HB1 were often inferior. The HB2 blocks for the TEAM approach, as well as the HB3 blocks for the MLR approach were advantageous for hybrid prediction.

Marker-based prediction versus GCA-based prediction

The largest experiment (Exp. 1), which also showed the highest SCA variance and SCA:GCA ratio, yielded the lowest prediction efficiencies of all experiments and the largest variance of R^2 for the GCA-based approach. In our study, Exp.1 is the only experiment where a marker-based prediction approach (HB2-TEAM) achieved a prediction efficiency, which was equal to that obtained with the GCA-based prediction. The smallest experiments (Exps. 2 and 4) exhibited the lowest SCA:GCA ratios and the highest prediction efficiencies. In Exp. 4, MLR-based approaches obtained prediction efficiencies close to those obtained with the GCA-based approach, in contrast to Exp. 1, where a TEAM-based approach achieved prediction efficiencies equal to the GCA-based approach. Summarizing, in our experiments from a breeding program for inter-group hybrids with predominance of GCA over SCA variances, the hybrid prediction from GCA effects was efficient in identifying promising hybrids and resulted in equal or higher prediction efficiencies compared with all investigated marker-based approaches.

Data from field experiments of hybrids and genotyping of their parental inbreds provided the basis for assessing the presented approaches for prediction of untested hybrids. This has the clear advantage of being a real-life example. However, even with that it is difficult to examine the approaches under defined conditions while assessing the influence of specified scenarios, such as the effect of experiment size or relevance of SCA. Therefore, simulation studies could be a valuable tool for undertaking further extensive evaluation of the presented approaches. In the current work, the issue of missing marker data was addressed by (1) discarding markers that were strongly affected by missing observations, and (2) using the average of the τ_c estimates as a substitute for missing observations in the TEAM-based prediction procedures. However, in the MLR approach, markers could only be considered if completely unaffected by missing observations. Procedures to estimate haplotypes that are affected by missing observations could be combined with either of the marker-based prediction approaches and thereby further improve their value for performance prediction of untested hybrids. Thus, considering the advantage of data-driven haplotype block approaches over single marker approaches for prediction of inter-group hybrids, we see a high potential to substantially improve the efficiency of hybrid breeding programs.

Acknowledgments This project was supported by the German Research Foundation DFG in the framework program ''Heterosis in Plants'' (research grants FR 1615/3-1 and 3-3). We thank Dr. J. Muminovic for helpful comments on this manuscript. The staff at the Plant Breeding Research Station at Eckartsweier and Hohenheim is gratefully acknowledged for conducting the field experiments.

References

Anderson EC, Novembre J (2003) Finding haplotype block boundaries by using the minimum-description-length principle. Am J Hum Genet 73:336–354

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 57:289–300

Bernardo R (1994) Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci 34:20–25

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

Bernardo R (1998) Predicting the performance of untested single crosses: trait and marker data. In: Lamkey KR, Staub JE (eds) Concepts and breeding of heterosis in crop plants. Crop Science Society of America, Madison, pp 117–127

Bernardo R (1999) Marker-assisted best linear unbiased prediction of single-cross performance. Crop Sci 39:1277–1282

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

Cochran WG, Cox GM (1957) Experimental designs. Wiley, New York

Cockerham CC (1967) Prediction of double crosses from single crosses. Der Züchter 37:160–169

Comstock RE, Robinson HF (1952) Estimation of average dominance of genes. In: Gowen JW (ed) Heterosis. ISU Press, Ames, pp 494–516

Dijkstra EW (1959) A note on two problems in connexion with graphs. Numer Math 1:269–271

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225–2229

Gardner CO, Eberhart SA (1966) Analysis and interpretation of the variety cross diallel and related populations. Biometrics 22:439–452

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48:361–372

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447–1455

Jansen RC, Jannink JL, Beavis WD (2003) Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. Crop Sci 43:829–834

Johnson GCL, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RCJ, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SCL, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. Nat Genetics 29:233–237

Maurer HP, Knaak C, Melchinger AE, Ouzunova M, Frisch M (2006) Linkage disequilibrium between SSR markers in six pools of elite lines of an European breeding program for hybrid maize. Maydica 51:269–279

Melchinger AE (1999) Genetic diversity and heterosis. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. ASA-CSSA, Madison, pp 99–118

Melchinger AE, Geiger HH, Seitz G, Schmidt GA (1987) Optimum prediction of three-way crosses from single crosses in forage maize (*Zea mays* L.). Theor Appl Genet 74:339–345

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BTN, Norris MC, Sheehan JB, Shen NP, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SPA, Cox DR (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719–1723

Peleman J, van Wijk R, Van Oeveren J, Van Schaik R (2000) Linkage map integration: an integrated genetic map of *Zea mays* L. In: Proceedings of plant and animal genome conference VIII, San Diego, US, 9–12 Jan 2000. Poster P472

Piepho HP, Gauch HG (2001) Marker pair selection for mapping quantitative trait loci. Genetics 157:433–444

R Development Core Team (2004) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Schrag TA, Melchinger AE, Sorensen 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, Maurer HP, Melchinger AE, Frisch M, Heckenberger M, van der Voort JR, Peleman J, Sorensen AP, Reif JC (2006) Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. Mol Breed 17:217–226

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp *mays* L.). Proc Natl Acad Sci USA 98:9161–9166

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP—A new technique for DNA-fingerprinting. Nucleic Acids Res 23:4407–4414

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lübberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M (1999) Two high-density AFLP (R) linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921–935

Vuylsteke M, Kuiper M, Stam P (2000) Chromosomal regions involved in hybrid performance and heterosis: their AFLP (R)-based identification and practical use in prediction models. Heredity 85:208–218

Zaykin D, Zhivotovsky L, Weir BS (1995) Exact tests for association between alleles at arbitrary numbers of loci. Genetica 96:169–178

Zhang K, Deng M, Chen T, Waterman MS, Sun F (2002) A dynamic programming algorithm for haplotype block partitioning. Proc Natl Acad Sci USA 99:7335–7339