Theor Appl Genet (2003) 106:559–567
DOI 10.1007/s00122-002-1144-5

D. R. Jordan · Y. Tao · I. D. Godwin · R. G. Henzell
M. Cooper · C. L. McIntyre

Prediction of hybrid performance in grain sorghum using RFLP markers

Received: 18 July 2001 / Accepted: 13 May 2002 / Published online: 19 December 2002

Abstract Heterosis is an important component of hybrid yield performance. Identifying high yielding hybrids is expensive and involves testing large numbers of hybrid combinations in multi-environment trials. Molecular marker diversity has been proposed as a more efficient method of selecting superior combinations. The aim of this study was to investigate the value of molecular marker-based distance information to identify high yielding grain sorghum hybrids in Australia. Data from 48 trials were used to produce hybrid performance-estimates for four traits (yield, height, maturity and stay green) for 162 hybrid combinations derived from 70 inbred parent lines. Each line was screened with 113 mapped RFLP markers. The Rogers distances between the parents of each hybrid were calculated from the marker information on a genome basis and individually for each of the ten linkage groups of sorghum. Some of the inbred parents were related so the hybrids were classified into 75 groups with each group containing individual hybrids that showed similar patterns of Rogers distances across linkage groups. Correlations between hybrid-group performance and hybrid-group Rogers distances were calculated. A significant correlation was observed between whole genome-based Rogers distance and yield (r = 0.42). This association is too weak to be of value for identifying superior hybrid combinations. One reason for the generally poor association between parental genetic diversity and yield may be that important QTLs influencing heterosis are located in particular chromosome regions and not distributed evenly over the genome. Variation in the sign and magnitude of correlations between Rogers distance and hybrid-group performance for particular linkage groups observed in this study support this hypothesis. The concept of using diversity on individual linkage groups to predict performance was explored. Using data from just two linkage groups 38% of the variation in hybrid performance for grain yield could be explained. A model combining phenotypic trait data and parental diversity on particular linkage groups explained 71% of the variation in grain yield and has potential for use in the selection of heterotic hybrids.

Keywords Sorghum · Genetic diversity · RFLP · QTL · Heterosis · Stay-green height yield

Introduction

In Australia, hybrid grain sorghum (*Sorghum bicolor* L. Moench) is grown predominantly as a rain fed crop in a wide range of environments from northern New South Wales to Central Queensland (approximately between latitudes 22° to 32°S). Australian sorghum breeding programs use multi-environment trials (METs) in an attempt to sample a range of the possible types of environments and select hybrids that show broad adaptation.

Heterosis is an important component of the performance of sorghum hybrids. Conventional selection of superior hybrid combinations involves testing of large numbers of hybrid combinations in METs, which is expensive. Interest has been shown in methods to reduce the cost and increase the speed with which superior hybrids are identified. In other cereals where heterosis is important, attempts have been made to use genetic diversity measures based on molecular or allozyme markers to identify superior hybrid combinations (Everson and Schaller 1955; Smith et al. 1990; Zhang et al. 1994, 1995;

Communicated by Langridge

D.R. Jordan (✉) · R.G. Henzell
Department of Primary Industries, Hermitage Research Station,
Warwick, Queensland 4370, Australia
e-mail: David.R.Jordan@dpi.qld.gov.au

Y. Tao · C.L. McIntyre
CSIRO Plant Industry, St Lucia Queensland 4067, Australia

I.D. Godwin · M. Cooper
School of Land and Food Sciences, The University of Queensland,
Brisbane, Queensland 4072, Australia

Currently:
M. Cooper, Pioneer Hi-Bred International,
Incorporated PO Box 1004, Johnston, Iowa 50131, USA

Martin et al. 1995; Lanza et al. 1997). The results of these studies have been variable but, in general, parental diversity as measured by molecular markers showed higher correlations with performance when the parents were related, as opposed to crosses between unrelated parents (e.g. crosses within heterotic groups of maize as opposed to crosses between lines from different heterotic groups). These results are consistent with the theoretical predictions made by Bernardo (1992), Charcosset and Essioux (1994) and Charcosset et al. (1991). These authors outlined a set of conditions that they considered important if molecular marker diversity was to usefully predict heterosis. These conditions included: (1) linkage of a large proportion of markers to QTLs controlling the trait in question, (2) gametic phase linkage disequilibrium between marker alleles and QTL alleles, (3) strong dominance effects, and (4) relatively high trait heritability. The lack of success in predicting hybrid performance using genetic distance measures based on genetic markers distributed across the genome have been attributed to the failure of experimental populations to satisfy conditions (1) (Zhang et al. 1994, 1995) and (2) (Frei et al. 1986; Godshalk et al. 1990; Melchinger et al. 1990a, b).

Sorghum hybrids are produced using a cytoplasmic male sterility system based on the interaction between kafir nuclear genes and milo cytoplasm, developed by Stephens and Holland (1954). Hybrid cultivars are developed from crosses between male-sterile (A/B) lines in milo cytoplasm and restorer (R) lines. Heterotic groups are not as clearly defined in sorghum as in maize. Studies using molecular markers cluster A/B and R lines separately (Ahnert et al. 1996); however, the majority of RFLP patterns were common to both groups suggesting that the A/B and R line groups have not diverged to an extreme degree (Vierling et al. 1994; Ahnert et al. 1996). Genetic diversity in elite sorghum lines in the USA is relatively narrow (Duncan et al. 1991; Poehlman 1987; Ahnert et al. 1996) and the genetic diversity of sorghum in Australia may be considerably narrower due to selection for genes for midge resistance through selection, linkage drag and genetic drift (Jordan et al. 1998). Many Australian sorghum inbreds share recent common ancestry as a result of selection for resistance to sorghum midge (Henzell et al. 1986), leaf disease resistance (Henzell et al. 1982) and Stay green (non-senescence) (Henzell, personal communication). Shared ancestry particularly in regions linked to important traits is likely to result in gametic phase linkage disequilibrium in both male and female breeding populations. The pedigree breeding method which is used exclusively by Australian sorghum breeders is likely to maintain linkage disequilibrium due to limited opportunities for crossing over. Hence, the levels of gametic phase linkage disequilibrium between QTL and marker alleles may be quite high, fulfilling one of the major criteria necessary for molecular marker diversity to predict F1 performance.

The paper will dissect the relationship between the performance of sorghum hybrids and parental genetic diversity as measured by RFLP markers with a view to using the relationship to predict hybrid performance.

Materials and methods

Genetic materials

Seventy sorghum inbred parent lines consisting of 37 male parents (R lines) and 33 female parents (A/B lines) predominantly from the Queensland Department of Primary Industry (QDPI) sorghum breeding program were used in this study. The lines included both publicly released and experimental germplasm tested during a 14-year period. Performance data on 137 F1 hybrids generated from these parents was used for this study. The parent lines were a sample of the material tested in the program during this period. The sample was not random in that each of the lines had been tested in various hybrid combinations and selected over a number of years. Almost all of the lines had been used for crossing to generate new breeding material. All parent lines were tested in hybrid combination in more than one trial and in more than 1 year.

RFLP marker data

Eighty four RFLP markers were chosen to give relatively even coverage of the ten linkage groups identified by Tao et al. (2000). The 70 parent lines were then scored with the 113 RFLP markers. DNA extraction, restriction enzyme digestion, Southern blotting, DNA probe preparation and labelling, hybridisation and autoradiography were carried out as described by Tao et al. (1996). One of the following five restriction enzymes, *Eco*RI, *Eco*RV, *Hin*dIII, *Dra*I and *Xba*I, was used with each of the RFLP probes. The probe/enzyme combination was chosen based on the enzymes found to be polymorphic in the mapping cross of Tao et al. (2000). The 84 RFLP markers included probes derived from six cereal species (Tao et al. 2000).

Analysis of RFLP data

Genetic distances between the inbred parents of each hybrid were estimated using the Rogers distance equation (Rogers 1972):

$$RD_{ij} = \sum_{k=1}^{L} \left[0.5 \sum_{n=1}^{A_K} (p_{ikn} - p_{jkn})^2 \right]^{0.5} / L,$$

where RD_{ij} is the Rogers distance between inbreds i and j; L is the total number of RFLP loci compared between inbreds i and j; A_k is the number of RFLP alleles at the Kth locus and p_{ikn} and p_{jkn} are the frequencies of the nth RFLP allele at the kth RFLP locus in the inbreds i and j respectively. Since the parent lines used in this study were homozygous inbred lines, the Rogers distance value obtained corresponds to the proportion of heterozygous RFLP loci in each hybrid. Rogers distances were calculated between the parents of each hybrid for the whole set of markers and on an individual linkage group basis.

Clustering of hybrids

Some of the parents within the male and female parental line groups were closely related to each other (e.g. lines derived from the same or a closely related cross). In the breeding program particular male or female parents were used as testers so in many cases closely related lines were crossed to the same inbred tester. Therefore, in a number of instances particular hybrid combinations could have very similar patterns of heterozygosity across linkage groups. As a result, particular patterns of heterozygosity across linkage groups were likely to have more representatives within the sample of hybrids while other combinations would be under-represented. We were interested in the general relationship between heterozygosity and hybrid performance, so to avoid potential for bias the 162 hybrids were classified into groups based on the patterns of heterozygosity across their linkage groups. The average Rogers distance for each of the ten linkage groups from the 162 hybrids was used for cluster analysis. The average Rogers distance data for each linkage group were standardised and clustered using squared Euclidean distances as the dissimilarity measure and incremental sum of squares as the fusion strategy (Ward 1963; Burr 1968, 1970; Wishart 1969). Using the groups from this cluster analysis, average Rogers distances and performance for the traits were calculated for each group of hybrids. In addition, mean, minimum and maximum Rogers distance were calculated for each of the ten linkage groups based on the data from the hybrid groups.

Hybrid performance estimates

The QDPI conducts METs to test the yield performance of hybrids produced from crosses between inbred parent lines developed in the program. Data from 48 of these trials including 15 locations and 14 years (Henzell, unpublished data) were used to obtain estimates of hybrid performance for 162 F1 sorghum hybrids derived from the 70 parent lines. Each trial contained only a subset of the 162 hybrids (on average 20). The 162 hybrids consisted exclusively of hybrids derived from crosses between R lines (male parents) and A/B lines (female parents). The trials were all replicated using randomised complete block or alpha designs. Performance data were for grain yield, maturity, height and stay green (leaf nonsenescence). Grain yield was measured on a plot basis. Maturity was measured as the number of days from planting to 50% anthesis. Plant height was measured from the ground to the tip of the panicle and stay green was a visual rating of the proportion of green leaf at maturity.

Due to the unbalanced nature of the data set, the residual maximum-likelihood technique (REML) (Patterson and Thompson 1971) was used to calculate best linear unbiased estimates (BLUEs) of the performance of the hybrid main-effect over the 48 environments for each of the four traits. The BLUEs were calculated using the computer program ASREML that employs the average information algorithm (Gilmour et al. 1995). Data were only available for a subset of 148 hybrids for stay green and 99 hybrids for height.

The BLUEs of the hybrid effects for each trait were calculated for each hybrid using the following linear model:

$$y_{ijk} = m + h_i + e_j + (he)_{ij} + \varepsilon_{ijk},$$

where y_{ijk} is the phenotypic observation k for the trait measured for hybrid i in environment $j(i = 1,\ldots, n_h, j = 1,\ldots, n_e, k=1,\ldots, n_r$ where n_h, n_e and n_r are the number of hybrids, environments and replicates within an environment, respectively); m is a constant (the grand mean of all observations); h_i is the fixed effect of hybrid i, $\sum_i h_i=0$; e_j is the effect of environment j, $e_j \sim N(0, \sigma^2_e)$; $(he)_{ij}$ is the effect of the interaction of hybrid i with environment j, $(he)_{ij} \sim N(0, \sigma^2_{he})$ and ε_{ijk} is the residual variation, $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$. The hybrids were assumed to be fixed because they represented a selected group rather than a random sample. The environments were assumed to be a random sample of production environments for sorghum in Australia. Performance estimates for the hybrid groups from the cluster analysis were calculated by averaging the BLUEs of each member of the group.

Statistical analysis

Simple product-moment correlation coefficients were computed among the trait BLUEs for the hybrid groups and also between these estimates and the average Rogers distances between the parents of each hybrid group. The correlation coefficients with Rogers distances were computed on a whole genome basis and individually for each linkage group. Correlation coefficients were also calculated between the parental Rogers distances calculated for each linkage group to determine whether Rogers distances between parents calculated with markers from one linkage group, were correlated with the Rogers distances between parents calculated using markers from other linkage groups. Finally, correlations were computed between the estimates of hybrid group performance and the Rogers distance between the parents of each hybrid group. These correlations were computed for the whole genome and individually for each linkage group. Stepwise linear regression was used to select sets of independent variables (Rodgers distances and traits) that were used in regression models to predict yield. Coefficients of determination (R^2) were calculated for the various models using simple and multiple regression.

Results

Clustering of hybrids

The clustering grouped hybrids that had similar patterns of heterozygosity across linkage groups. A truncation level of 75 hybrid groups was chosen where 95% of the hybrid × Rogers distance (at each linkage group) sums of squares were retained. The percentage of the interaction sums of squares retained at a particular truncation level is equivalent to an R^2 value and is a measure of the proportion of the total information which is not lost when hybrids are grouped together. If each hybrid is considered as a separate group, 100% of the interaction sum of squares is retained. The 75 group truncation level was chosen because the increase in information retained with increasing numbers of groups began to plateau at this level. At this truncation level, 24 of the 75 groups consisted of single hybrids, 27 of two hybrids, 14 of three hybrids, three of four hybrids and seven of five hybrids. The hybrid groups related strongly to the known pedigree information, with most groups consisting of a single female or male tester crossed to a group of closely related lines.

Genetic distances among parents of hybrid groups

The distribution of Rogers distances for each linkage group using data from the hybrid groups is presented in Fig. 1. Distances for each linkage group ranged considerably among the hybrids, in most cases minimum distances were close to zero and maximum distances varied from 0.69 to 1 and average genetic distances varied from 0.37 to 0.50.

Correlations among values of parental diversity at different linkage groups

Correlations were carried out among the parental Rogers distances calculated for the individual linkage groups in order to determine the degree of independence of the distance information calculated for each linkage group (Table 1). Of the 45 correlations, nine were statistically significant ($p < 0.05$) and of these eight were positive and one was negative. Most of the correlations were relatively weak with the strongest correlation between any two linkage groups occurring between linkage groups B and

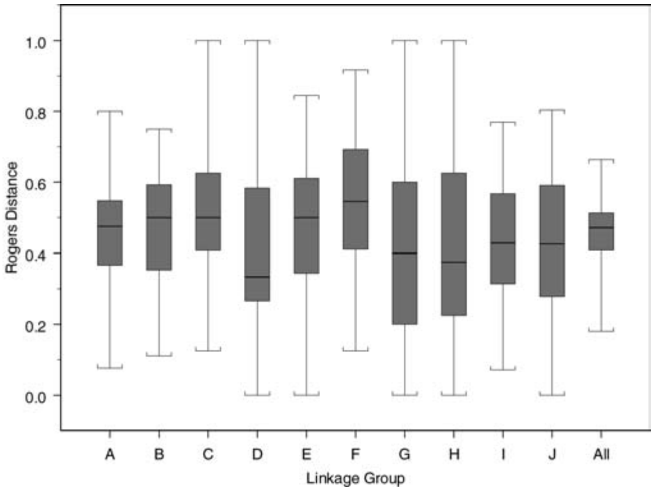

Fig. 1 Box plot showing upper and lower limits of Rogers distance between parent lines of the 75 hybrid groups for ten linkage groups considered independently and for all linkage groups considered simultaneously. The top and bottom of the boxes correspond to the 25th and 75th percentiles

F ($r = 0.38$). These results indicate that generally most of the linkage groups provided different information about Rogers distance among the hybrid groups.

Correlations between traits and genetic distances

Table 2 contains three types of data: (1) linear correlation coefficients between the hybrid group performance estimates for the four traits; (2) correlation coefficients between hybrid group heterozygosity estimates and hybrid group performance estimates for the four traits; and (3) the linkage group location of relevant QTLs previously mapped in sorghum. Height and stay green were positively correlated with grain yield while maturity was negatively correlated with yield. A significant positive correlation was observed between whole genome heterozygosity and yield. Significant positive correlations were also observed between heterozygosity and yield for five of the ten linkage groups. Whole genome heterozygosity was also positively correlated with height. Significant

Table 1 Correlations of parental Rogers distance among individual linkage groups for 75 hybrid groups

	A	B	C	D	E	F	G	H	I	J
A	1									
B	–0.08	1								
C	0.28*	0.05	1							
D	–0.07	0.27*	0.01	1						
E	0.09	0.01	0.10	–0.08	1					
F	0.21	0.38***	–0.05	0.02	–0.15	1				
G	0.21	0.12	0.14	–0.13	0.32**	0.12	1			
H	0.34**	0.09	0.14	0.04	–0.25*	0.22	–0.09	1		
I	0.13	–0.03	0.01	0.11	0.25*	0.08	0.06	0.13	1	
J	0.11	0.16	–0.01	–0.07	0.27*	0.11	0.26*	–0.11	0.03	1

Significance levels $*P < 0.05$, $**P < 0.01$, $***P < 0.001$

Table 2 Correlations between hybrid performance for various traits, between hybrid performance and parental Rogers distances and the location of relevant QTLs. The letters T, Z, R, P, N indictae that QTLs for those traits have been located on those linkage groups in the following studies T = Tao et al. (2000), Tao YZ unpublished data, N = Subudhi et al. (2000), Xu et al. (2000), Crasta et al. (1999), V = Tuinstra et al. (1997), R = Rami et al. (1998), P = Lin et al. (1995) and Paterson et al. (1995), Z = Pereira and Lee (1995). Superscripts indicate the number of QTLs detected if multiple QTLs were identified on a linkage group by a particular research group

Trait and linkage groups	Grain yield	Maturity		Plant height		Stay green	
Maturity	–0.28***						
Height	0.47***	0.14					
Stay green	0.40***	0.02		0.16			
All	0.42***	–0.25*		0.47***		0.13	
A	0.30**	–0.23*		0.18	TP	0.17	TN^{2}V
B	0.07	–0.09	TP	0.27*	TR	–0.40***	TNV
C	0.08	–0.09		0.16	TPR	0.12	T
D	–0.16	0.08	P	0.17	PZ	–0.16	
E	0.23*	–0.08		0.19		0.55***	
F	0.32**	0.07	T	0.29*	R	–0.08	
G	0.51***	–0.21	P	0.31*	P^{2}Z	0.21	T
H	–0.19	0.06		0.07		–0.26*	N
I	0.09	–0.16	T	0.07	TZ	0.28*	T
J	0.47***	–0.34**		0.31*	PRZ	0.24*	

Significance levels $*P < 0.05$, $**P < 0.01$ $***P < 0.001$

Table 3 Coefficients of determination for simple and multiple linear regression equations predicting hybrid group yield from variables including genetic diversity measures and various traits. Trait abreviations SG = stay green, HGT = height, DF = maturity

Independent variable(s)		R^2%
Markers	All markers	18***
	J	22***
	G and J	38***
	A, F, G, H and J	51***
Traits	DF	9*
	SG	19***
	HGT	23***
	SG HGT DF	57***
Markers and traits	SG, HGT, DF, D	61***
	SG, HGT, DF, F	60***
	SG, HGT, DF, G	63***
	SG, HGT, DF, H	59***
	SG, HGT, DF, I	61***
	SG, HGT, DF, F,G,I	71***

Significance levels $*P < 0.05$, $**P < 0.01$, $***P < 0.001$

positive correlations between height and heterozygosity were observed for five of the ten linkage groups. In contrast, the correlations between heterozygosity and maturity were predominantly negative. Significant negative correlations were observed for whole genome heterozygosity and for heterozygosity on linkage groups A and J. Whole genome heterozygosity was not correlated with stay green. However, significant positive correlations between heterozygosity and stay green were observed for three of the ten linkage groups and significant negative correlations were observed for two linkage groups. Considerable variation was apparent in the magnitude of correlations and their direction among linkage groups for each of the traits. However, most of the correlations were moderate to weak with the strongest correlation of 0.55 occurring between heterozygosity on linkage group E and stay green.

Prediction of hybrid performance

The results of the linear regression and stepwise linear regression are presented in Table 3. Using average genetic diversity as a predictor, 18% of the genetic variation for grain yield could be explained. However, 51% of the variation in grain yield could be accounted for by using variation in heterozygosity on linkage groups A, F, G, H and J. An R^2 of 38% could be obtained by using data from linkage groups G and J alone. This is compared to a model with three traits (maturity height and stay green) which explains 57% of the variation. The best multiple regression models using five linkage groups, D, F, G, H and I, significantly improved the R^2 of the multiple regression model based on the traits alone, with the R^2 value increased by between 2 to 6%. The inclusion of average genetic diversity rather than individual linkage group did not significantly improve the model. The best model combining the three traits and multiple linkage groups included three linkage groups and this explained 71% of the variation in yield.

Discussion

Heterosis makes an important contribution to the yield of hybrid sorghum and as a result sorghum breeders find that the yield per se of inbred parents is a poor indicator of hybrid yields (Duvick 1999). Estimates of mid-parent heterosis for grain yield in sorghum range from 6 to 54% (Kirby and Atkins 1968; Liang and Walter 1968, 1969; Wenzel 1988). Generally the advantage of hybrids is in the order of 30–40% which is less than the figure of approximately 50% observed in maize (Duvick 1999). Maturity, height and stay green also exhibit heterosis with estimates of mid-parent heterosis varying from –1 to –6% for maturity and 5 to 19% for height (Kirby and Atkins 1968; Liang and Walter 1968; Wenzel 1988). van Oosterom et al. (1996) observed heterosis of 27% for absolute green leaf area duration which is highly correlated with stay green.

Plant height, stay green and maturity are all associated with grain yield in Australian sorghum growing environments. The strong correlation ($r = 0.47$) we observed between increased plant height and grain yield in this study is often noted in sorghum (Casady 1965; Graham and Lessman 1966; Liang et al. 1969; Henzell et al. 1982). The variation in plant height among the genotypes in this study was not extreme, with all hybrids being of acceptable height for commercial production. Maturity was negatively correlated ($r = –0.28$), with the grain yield, with later maturing hybrids tending to have lower yields than earlier maturing hybrids. Under conditions where water is not limiting, later maturing genotypes generally produce higher grain yields. However, in Australia, post-anthesis drought stress is relatively common (Chapman et al. 2000) and under these conditions earlier maturing hybrids are favoured because their water use is better matched to the available water. Post-anthesis stay green of sorghum is positively correlated with increased grain yield for environments where water stress occurs during grain filling (Henzell et al. 1992; Borrell and Douglas 1996; Borrell et al. 2000b). The positive correlation between grain yield and stay green was expected because of the frequency of post-anthesis water stress in Australian sorghum growing environments.

Clustering of the hybrids

The decision to cluster the hybrids into groups was based on our intention to observe the general relationship between parental diversity and performance among the germplasm used in this study. The over-representation of particular patterns of diversity across linkage groups had the potential to distort this relationship. The hybrid groups generated from the clustering procedure related strongly to the known pedigree information, with most groups consisting of a single female or male line crossed to a group of closely related lines. This result gave us confidence in the clustering procedure. As expected, some patterns of heterozygosity across linkage groups were over represented in our sample of hybrids. Some of the groups contained up to five hybrids, while most groups contained only one individual. The associations found with the grouped data were generally similar to those found using individual hybrids (data not shown). However, when hybrid groups were used, the magnitude of the stronger correlations generally increased while the significance of the weaker associations declined due to reduced sample size (n = 75 vs n = 162) and reduced magnitude of the correlations (data not shown).

Correlation between genome wide heterozygosity and hybrid performance

Average, minimum and maximum Rogers distances between the parents in this study are considerably lower than those calculated for the groups of maize hybrids used by Godshalk et al. (1990) and Melchinger et al. (1990a, b) but comparable with, though lower than, those calculated by Lee et al. (1989). The hybrids used by Lee et al. (1989) were the result of crosses between maize inbred lines from the same and different heterotic groups, whereas those hybrids in the former studies involved crosses primarily among inbred lines from different heterotic groups. Our results reflect the lower level of polymorphism in grain sorghum compared with maize (Tao et al. 1998) and the known shared ancestry of the male and female breeding populations used in this study (Jordan et al. 1998; R.G. Henzell, personal communication).

Yield of the hybrid groups in this study was significantly correlated with whole genome parental genetic diversity (r = 0.42). The magnitude of this correlation is comparable to correlations observed in maize studies that included hybrids produced from crosses among lines from the same heterotic group (Lee et al.1989; Melchinger et al. 1990b) and considerably greater than the correlations obtained when the experiments involved crosses primarily among inbreds from different heterotic groups (Godshalk et al. 1990; Melchinger et al. 1990b; Dudley et al. 1991).

Plant height was also positively correlated with whole genome heterozygosity (r = 0.47), while maturity was negatively correlated (r = –0.25). The sign of the correlation coefficients for these traits is consistent with the direction of heterosis found in other studies. No association was apparent between stay green and whole genome heterozygosity. These results indicate that significant levels of gametic phase linkage disequilibrium exist between QTLs and marker alleles within the set of parents used in this study. However, variation in average heterozygosity explained only 18% of the variation for yield, which is of limited value for identifying high yielding hybrid combinations. The phenotypic traits stay green (r = 0.40) and height (r = 0.47) were as useful for predicting yield as marker diversity.

Correlation between heterozygosity of particular linkage groups and hybrid performance

In the absence of selection, positive associations might be expected between parental genetic diversity on different chromosomes. Selection for favourable alleles and their associated linkage blocks on different chromosomes has the potential to alter this expectation in unpredictable ways. We found a general lack of strong correlations between the parental Rogers distances on different linkage groups. This indicated that most of the linkage groups provided independent sources of information about genetic diversity among the parents of the hybrid groups. Polymorphic genes that affect yield are unlikely to be evenly distributed over the genome, making it likely that the association between heterozygosity and yield will vary between linkage groups and the evidence from this study supported this idea.

Correlations between hybrid group heterozygosity and grain yield showed variation among linkage groups. Heterozygosity was positively correlated with yield for five (A, E, F, G and J) of the ten linkage groups. The correlation coefficients of 0.51 for linkage group G and 0.47 for linkage group J were higher than the correlation associated with whole genome heterozygosity. Variation in genetic distance between parent lines for these two linkage groups can be used to explain 38% of the variation in F1 yield compared with 18% for whole genome heterozygosity. Similarly, variations in associations were observed for the other traits. Height was positively correlated with parental heterozygosity (linkage groups B, F, G and J). Stay green was significantly correlated with heterozygosity for linkage groups B, E, H, I and J. The correlations were positive for linkage groups E, I and J and negative for linkage groups B and H. In the case of maturity, heterozygosity was only associated with the trait for linkage groups A and J.

Association between QTLs identified for particular traits identified in other studies and linkage groups associated with those traits in this study

Common RFLP probes were used to align the linkage groups in the map of Tao et al. (2000) with linkage groups of sorghum maps used in a number of sorghum QTL studies (Table 2). All four linkage groups where height was significantly correlated with heterozygosity had previously been observed to contain QTLs for plant height. The genetic control of plant height in sorghum is complex, with height to the tip of the panicle affected by internode length (four major genes with modifiers), the number of internodes (four major genes), peduncle length and panicle length (Doggett 1988). It is not surprising, therefore that most linkage groups have been observed to contain genes that affect plant height. Linkage group J showing a strong association with grain yield is of particular interest. Pereira and Lee (1995) and Rami et al. (1998) both presented evidence that this linkage group contained the dwarfing gene Dw3 that is known to have a pleiotropic effect on a number of important characteristics including grain yield (Casady 1965). Our results suggest that this linkage group is strongly associated with hybrid performance for a number of traits, including plant height, stay green, maturity and grain yield.

Neither of the linkage groups that were associated with maturity were observed to contain QTLs for maturity in other studies. Stay green was significantly correlated with heterozygosity for linkage groups B, E, H, I and J. Three of the five groups have previously been shown to contain QTLs for stay green (non-senesence). Linkage groups E and J have not been shown to contain stay green QTLs. These correlations may be spurious associations due to correlations among the linkage groups (see Table 1.) alternatively they might be driven by QTLs that are yet to be mapped. All of the stay green QTLs mapped previously have been derived from the line B35; at least one different source of stay green derived from the line KS19 exists and appears to have a different mode of action at the physiological level (Borrell et al. 2000a). KS19 was used extensively in the development of the male parents employed in this study. Three out of five significant correlations were positive, which is consistent with the general observation of dominance of stay green (non-senescence) over senescence observed by Walulu et al. (1994) and van Oosterom et al. (1996).

In general, the data from the QTL studies was of limited value in identifying linkage groups likely to be associated with hybrid performance. This is likely to be due to a range of factors including the paucity of QTL data for these traits, the wide crosses used in many of the mapping populations, the frequency of particular QTLs segregating in this sample of lines and the size of particular QTL effects.

The significant negative correlations observed between stay green and heterozygosity for linkage groups B and H are of interest. One possible explanation for these correlations is that these linkage groups contain genes for stay green with recessive gene action present at relatively high frequency in this group of sorghum inbreds. Recessive gene action has been reported for particular sources of stay green. (Tenkouano et al. 1993). All of the linkage groups that showed significant correlations with grain yield also showed significant correlations with one or more of the traits (maturity, plant height or stay green). Linkage group B contrasts to this finding. The heterozygosity on this group shows a significant correlation with both stay green and height but not with grain yield. In this case heterozygosity on this linkage group is associated with reduced stay green (unfavourable for yield) and increased height (favourable for yield), potentially negating any association between heterozygosity and yield.

Implications for identification of high yield hybrid combinations

Yield of the hybrid groups in this study was significantly correlated with whole genome heterozygosity, but the strength of the association is of limited value for identifying high yielding hybrids. Genetic distance between parents for a combination of specific linkage groups or regions of linkage groups seems to have promise. Zhang et al. (1994, 1995) proposed a method whereby heterozygosity at selected markers that were found to be associated with grain yield was correlated with hybrid performance in rice. They found that correlation coefficients between the selected markers and hybrid performance were much greater than the correlations between distance measures based on a larger number of unselected marker loci. Our results support this type of approach. Variation in genetic distance of the two most important linkage groups (G and J) can be used to explain 38% of the variation in F1 yield compared with 18% for whole genome heterozygosity. A model combining information from linkage groups A,F,G,H and J explained 51% of the variation in hybrid performance. However, most of the variation in yield appears to be associated with variation in height (23%), stay green (19%) and maturity (9%). When combined, the three traits explained 57% of the variation for yield. The inclusion of markers as well as trait data improved the prediction of hybrid performance. Diversity data from the linkage groups D, F, G, H and I combined individually in a model that included the three phenotypic traits resulted in a statistically significant improvement to the prediction equation. The R^2 of the multiple regression based on the traits alone was increased by 2–6%. A multiple regression equation including the three traits and linkage groups F, G and I explained 71% of the variation in performance.

The figure of 71% of the variation in yield suggests that this information could be used to develop a prediction system for sorghum hybrid performance. Patterns of parental diversity across linkage groups have the potential to be exploited in predictions systems. However, the correlation between grain yield and the traits of stay green, height and maturity sugest that a more precise system could be developed. These traits have relatively high heritabilities and are under predominantly major gene control making the development of allele-specific markers for QTLs feasable. Access to allele specific markers combined with a physiological understanding of how these traits condition yield should allow a useful prediction system to be developed.

Acknowledgements The Authors wish to acknowledge the financial support of the Australian Grain Research and Development Corporation.

References

Ahnert D, Lee M, Austin DF, Livini C, Woodman WL, Openshaw SJ, Smith JSC, Porter K, Dalton G (1996) Genetic diversity among elite sorghum inbred lines assessed with DNA. Crop Sci 36:1385–1392

Bernardo R (1992) Relationship between single-cross performance and molecular marker heterozygosity. Theor Appl Genet 83: 628–634

Borrell AK, Douglas ACL (1996) Maintaining green leaf area in grain sorghum increases yield in a water-limited environment. Proc 3rd Australian Sorghum Conf. Australian Institute of Agricultural Science, Tamworth, pp 315–322

Borrell AK, Hammer GL, Douglas ACL (2000a) Does maintaining green leaf area in sorghum improve yield under drought? I. Leaf growth and senescence. Crop Sci 40:1026–1037

Borrell AK, Hammer GL, Henzell RG (2000b) Does maintaining green leaf area in sorghum improve yield under drought? II. Dry matter production and yield. Crop Sci 40:1037–1048

Burr EJ (1968) Cluster sorting with mixed character types. I. Standardization of character values. Aust Computer J 1:97–99

Burr EJ (1970) Cluster sorting with mixed character types. II. Fusion strategies. Aust Computer J 2:98–103

Casady AJ (1965) Effect of a single height (Dw) gene of sorghum on grain yield, grain yield components and test weight. Crop Sci 5:385–388

Chapman SC, Cooper M, Hammer GL, Butler DG (2000) Genotype by environment interactions affecting grain sorghum. II Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. Aust J Agric Res 51:209–221

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

Charcosset A, Lefort-Buson M, Gallais A (1991) Relationship between heterosis and heterozygosity at marker loci: a theoretical computation. Theor Appl Genet 81:571–575

Crasta OR, Xu WW, Rosenow DT, Mullet J, Nguyen HT (1999) Mapping of post-flowering drought resistance traits in grain sorghum: association between QTLs influencing premature senescence and maturity. Mol Gen Genet 262:579–588

Doggett H (1988) Sorghum, 2nd edn. Longman Scientific and Technical. Copublished in the USA with John Wiley and Sons Inc, New York

Dudley JW, Saghai Maroof MA, Rufener GK (1991) Molecular markers and grouping of parents in maize breeding programs. Crop Sci 31:718–723

Duncan RR, Bramel-Cox PJ, Miller FR, Shands HL, Wiesner LE (1991) Contributions of introduced sorghum germplasm to hybrid development in the USA. Use of plant introductions in cultivar development. Part 1. Proc Symposium Division C-1 Crop Sci Society of America, Las Vegas, Nevada, USA. Crop Sci Soc America, Inc, pp 69–102

Duvick DN (1999) Heterosis: feeding people and protecting natural resources. In: Coors JG Pandey S (ed) The genetics and exploitation of heterosis in crops. Crop Sci Society of America, pp 19–30

Everson EH, Schaller CW (1955) The genetics of yield differences associated with awn barbing in the barley hybrid (Lion × Atlas10) × Atlas. Agron J 47:276–280

Frei OM, Stuber CW, Goodman MM (1986) Use of allozymes as genetic markers for predicting performance in maize single cross hybrids. Crop Sci 26:37–42

Gilmour AR, Thompson R, Cullis BR (1995) Average Information REML, an efficient algorithm for REML estimation in linear mixed models. Biometrics 51:1440–1450

Godshalk EB, Lee M, Lamkey KR (1990) Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. Theor Appl Genet 80:273–280

Graham D, Lessman KJ (1966) Effect of height on yield and yield components of two isogenic lines of *Sorghum vulgare* Pers. Crop Sci 6:372–374

Henzell RG, Persley DM, Fletcher DS, Greber RS, van Slobbe L (1982) Development of grain sorghum lines with resistance to sugarcane mosaic virus and other sorghum diseases. Plant Dis 66:900–901

Henzell RG, Brengman RL, Page FD, van Slobbe FDSL, Foster G (1986) Breeding for midge resistance in grain sorghum in Queensland. Proc 1st Australian Sorghum Conference, Gatton, February 1986, pp 7.10–7.24

Jordan DR Tao YZ, Godwin ID, Henzell RG, Cooper M, McIntyre CL (1998) Loss of genetic diversity associated with selection for resistance to sorghum midge in Australian sorghum. Euphytica 102:1–7

Kirby JS, Atkins RE (1968) Heterotic response for vegetative and mature plant characters in grain sorghum, *Sorghum bicolor* (L.) Moench. Crop Sci 8:335–339

Lanza LLB, Souza CL Jr, Ottoboni LMM, Vieira ML (1997) Genetic distance of inbred lines and prediction of maize single-cross. Theor Appl Genet 94:1023–1030

Lee M, Godshalk EB, Lamkey KR, Woodman WW (1989) Association of restriction fragment length polymorphisms among maize inbreds with the agronomic performance of their crosses. Crop Sci 29:1067–1071

Liang GHL, Walter TL (1968) Heritability estimates and gene effects for agronomic traits in grain sorghum, *Sorghum vulgare* Pers. Crop Sci 8:77–81

Liang GHL, Overley CB, Casady AJ (1969) Interrelations among agronomic characters in grain sorghum (*Sorghum bicolor* Moench). Crop Sci 9:299–302

Lin Y, Schertz KF, Paterson AH (1995) Comparative analysis of QTLs affecting plant height and maturity across the Poaceae, in reference to an interspecific sorghum population. Genetics 141:391–411

Martin JM, Talbert LE, Lanning SP, Blake NK (1995) Hybrid performance in wheat as related to parental diversity. Crop Sci 35:104–108

Melchinger AE, Lee M, Lamkey KR, Woodman WL (1990a) Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize inbreds. Crop Sci 30:1033–1040

Melchinger AE, Lee M, Lamkey KR, Hallauer AR, Woodman WL (1990b) Genetic diversity for restriction fragment length polymorphisms and heterosis for two diallel sets of maize inbreds. Theor Appl Genet 80:488–496

Oosterom van EJ, Jayachandran R, Bidinger FR (1996) Diallel analysis of the stay green trait and its components in sorghum. Crop Sci 36:549–555

Paterson AH, Lin Y, Li Z, Schertz KF, Doebley JF, Pinson SRM, Liu S, Stansel JW, Irvine JE (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science 269:1714–1718

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrica 58: 545–554

Pereira MG, Lee M (1995) Identification of genomic regions affecting plant height in sorghum. Theor Appl Genet 90: 380–388

Poehlman JM (1987) Breeding field crops. AVI publishing Company, Wesport Connecticut

Rami JF, Dufour P, Trouche G, Fliedel G, Mestres C, Davrieux F, Blanchard P, Hamon P (1998) Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). Theor Appl Genet 97:605–616

Rogers JS (1972) Measures of genetic similarity and genetic distance. Studies in genetics VII. University of Texas Publishing. Co, pp 145–153

Smith OS, Smith JSC, Bowen SL, Tenborg RA, Wall SJ (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield, heterosis, and RFLPs. Theor Appl Genet 80:833–840

Stephens JC, Holland RF (1954) Cytoplasmic male-sterility for hybrid sorghum seed production. Agron J 46:20–23

Subudhi PK, Rosenow DT, Nguyen HT (2000) Quantitative trait loci for the stay green trait in sorghum (*Sorghum bicolor* L. Moench): consistency across genetic backgrounds and environments. Theor Appl Genet 101:733–741

Tao YZ, McIntyre CL, Henzell RG (1996) Applications of molecular markers to Australian sorghum breeding programs. I: Construction of a RFLP map using sorghum recombinant inbred lines. AIAS Occasional Publication: pp 443–450

Tao YZ, Jordan DR, Henzell RG, McIntyre CL (1998) Construction of a genetic map in a sorghum recombinant inbred line using probes from different sources and its comparison with other sorghum maps. Aust J Agric Res 49:729–736

Tao YZ, Henzell RG, Jordan DR, Butler DG, Kelly AM, McIntyre CL (2000) Identification of genomic regions associated with stay green in sorghum by testing RILs in multiple environments. Theor Appl Genet 100:1225–1232

Tenkouano A, Miller FR, Fredriksen RA, Rosenow DT (1993) Genetics of nonsenescence and charcoal rot resistance in sorghum. Theor Appl Genet 85:644–648

Tuinstra MR, Grote EM, Goldsbrough PB, Ejeta G (1997) Genetic analysis of post-flowering drought tolerance and components. Mol Breed 3:439–448

Vierling RA, Xiang Z, Joshi CP, Gilbert ML, Nguyen HT (1994) Genetic diversity among elite Sorghum lines revealed by restriction. Theor Appl Genet 87:816–820

Walulu RS, Rosenow DT, Wester DB, Nguyen HT (1994) Inheritance of the stay green trait in sorghum. Crop Sci 34:970–972

Ward JH (1963) Hierarchical grouping to optimise an objective function. J Am Stat Assoc 58:263–44

Wenzel WG (1988) Inheritance and interrelationships of quantitative traits in sorghum. Sorghum Newslett 31:1–3

Wishart D (1969) An algorithm for hierarchical classification. Biometrics 25:165–170

Xu WW, Subudhi PK, Crasta OR, Rosenow DT, Mullet JE, Nguyen HT (2000) Molecular mapping of QTLs conferring stay green in grain sorghum (*Sorghum bicolor* L. Moench). Genome 43:461–469

Zhang QF, Gao YJ, Yang SH, Ragab RA, Saghai MA, Li ZB (1994) A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor Appl Genet 89: 185–192

Zhang QF, Gao YJ, Maroof MAS, Yang SH, Li JX (1995) Molecular divergence and hybrid performance in rice. Mol Breed 1:133–142