ORIGINAL PAPER

QTL analysis of early stage heterosis for biomass in Arabidopsis

Rhonda Christiane Meyer · Barbara Kusterer · Jan Lisec · Matthias Steinfath ·
Martina Becher · Hanno Scharr · Albrecht E. Melchinger · Joachim Selbig ·
Ulrich Schurr · Lothar Willmitzer · Thomas Altmann

Abstract The main objective of this study was to identify genomic regions involved in biomass heterosis using QTL, generation means, and mode-of-inheritance classification analyses. In a modified North Carolina Design III we backcrossed 429 recombinant inbred line and 140 introgression line populations to the two parental accessions, C24 and Col-0, whose F_1 hybrid exhibited 44% heterosis for biomass. Mid-parent heterosis in the RILs ranged from −31 to 99% for dry weight and from −58 to 143% for leaf area. We detected ten genomic positions involved in biomass heterosis at an early developmental stage, individually explaining between 2.4 and 15.7% of the phenotypic variation. While overdominant gene action was prevalent in heterotic QTL, our results suggest that a combination of dominance, overdominance and epistasis is involved in biomass heterosis in this Arabidopsis cross.

R. C. Meyer (✉) · T. Altmann
Department of Molecular Genetics, Leibniz Institute of Plant
Genetics and Crop Plant Research (IPK), Corrensstraße 3,
06466 Gatersleben, Germany
e-mail: meyer@ipk-gatersleben.de

B. Kusterer · A. E. Melchinger
Institute of Plant Breeding, Seed Science, and Population
Genetics, University of Hohenheim, 70599 Stuttgart, Germany

J. Lisec · J. Selbig · L. Willmitzer
Max-Planck-Institute of Molecular Plant Physiology,
14476 Potsdam-Golm, Germany

M. Steinfath · J. Selbig
Department of Bioinformatics, Institute of Biochemistry
and Biology, University of Potsdam, 14476 Potsdam-Golm,
Germany

Introduction

The term 'heterosis' describes the superiority of F_1 hybrids manifested as increased size, fertility, or yield compared with the corresponding parental homozygous lines (Shull 1948). The phenomenon of heterosis has been exploited extensively in plant breeding (Birchler et al. 2003), however, the underlying genetic and molecular mechanisms are still largely unknown. Early studies put forward three genetic models as explanation for the extreme hybrid phenotype. The dominance hypothesis (Bruce 1910; Davenport 1908) attributes heterosis to the joint action of favorable

M. Becher
Department of Genetics, Institute of Biochemistry and Biology,
University of Potsdam, 14476 Potsdam-Golm, Germany

Present Address:
M. Becher · H. Scharr · U. Schurr
ICG-III Phytosphere, Forschungszentrum Jülich GmbH,
52425 Jülich, Germany

Present Address:
B. Kusterer
HYBRO Saatzucht GmbH & Co. KG,
17291 Schenkenberg, Germany

dominant alleles at multiple loci from both parents combined in the hybrid. The overdominance hypothesis (Crow 1948; Hull 1945) postulates the existence of loci at which the heterozygous state is superior to either homozygote, i.e. a positive interaction between alleles at a heterozygous locus. Pseudo-overdominance (Crow 1952), in contrast, refers to the situation of tightly linked genes with favorable dominant alleles linked in repulsion. The epistasis hypothesis (Powers 1944; Williams 1959) explains heterosis by the interaction of favorable alleles at different loci contributed by the two parents, which themselves may show additive, dominant, or overdominant action.

One of the most promising approaches to unravel the genetic basis for heterosis at the molecular level emerged through the availability of molecular markers. They provided the means to perform quantitative trait locus (QTL) analyses, a powerful approach to map and subsequently identify genes involved in complex traits. In recent studies evidence was provided for the occurrence of dominance (Abdelkhalik et al. 2005; Xiao et al. 1995) or overdominance (Li et al. 2001; Semel et al. 2006) at multiple loci, and several reports highlighted the importance of epistasis (Kusterer et al. 2007a; Luo et al. 2001; Mei et al. 2005; Melchinger et al. 2007a; Monforte and Tanksley 2000; Yu et al. 2005). These diverse results indicate that heterosis may be caused by combinations of these mutually non-exclusive mechanisms.

Heterosis has been shown to be a widespread occurring phenomenon in *Arabidopsis thaliana* and has been described for such varying traits as stem length (Griffing and Langridge 1963), photosynthetic efficiency (Sharma et al. 1979), seedling viability (Mitchell-Olds 1995), seed number (Alonso-Blanco et al. 1999), seed size (Stokes et al. 2007), phosphate efficiency (Narang and Altmann 2001), cold tolerance (Korn et al. 2008; Rohde et al. 2004), rosette diameter and plant height (Barth et al. 2003; Syed and Chen 2004), and biomass (Meyer et al. 2004). The occurrence of heterosis in the model plant Arabidopsis opens the possibility to combine QTL analyses and the rapid identification of the causal DNA variation with transcript and metabolite profiling to discover physiological and molecular processes involved in heterosis. This process should establish a precedence that may guide the analysis of heterosis in crops.

We study heterosis in *A. thaliana* with the aim to identify genomic sequences that contribute to heterosis. The first step toward this end was a forward genetics approach of defining genomic regions influencing heterosis for biomass with the use of QTL mapping experiments in segregating populations developed from two divergent accessions, C24 and Col-0, which in previous studies exhibited transgressive segregation and heterosis for biomass (Meyer et al. 2004). The main objective of this study

was to identify QTL for biomass heterosis and to characterize the genetic mode of action of heterotic QTL using the approach proposed by Melchinger et al. (2007b).

Materials and methods

Plant materials

Analyses were performed on homozygous mapping populations derived from the two *A. thaliana* accessions C24 and Col-0 (P_1 and P_2). The recombinant inbred line (RIL) population was derived from reciprocal crosses C24 × Col-0 (201 lines) and Col-0 × C24 (228 lines) as described in Törjék et al. (2006). Two large reciprocal sets of introgression lines (ILs) were created through backcrossing and selfing using marker assisted selection (Törjék et al. 2008). In both IL sets (78 ILs with Col-0 background and 62 ILs with C24 background) the donor segments cover almost the entire genome with an average substitution size of 18.3 cM.

To allow analyses of heterotic effects, testcrosses (TC) with parents Col-0 and C24 were produced for 429 RILs (422 RIL-TC$_{C24}$ and 416 RIL-TC$_{Col}$) and 42 ILs (20 IL-TC$_{C24}$ and 22 IL-TC$_{Col}$).

Plant cultivation

Plants were grown in 1:1 mixture of GS 90 soil and vermiculite (Gebrüder Patzer, Sinntal-Jossa, Germany) in 96-well-trays under a long-day regime (16 h fluorescent light (120 µmol m^{-2} s^{-1}) at 20°C and 60% relative humidity/8 h dark at 18°C and 75% relative humidity). Six plants of the same line were grown per well. To avoid position effects, trays were rotated around the growth chamber every 2 days. In the first experiment, all RILs and RIL-TCs were cultivated together with parents C24 and Col-0 (P_1 and P_2) and their reciprocal F_1-hybrids C24 × Col-0 and Col-0 × C24 (F_{1-a} and F_{1-b}) in a split-plot design (Lisec et al. 2008). At least three replicates per line were grown.

For the second experiment, ILs were selected to cover the QTL regions determined in the RIL experiment. Plants were grown in two blocks and six subplots per block. Each subplot contained 42 ILs, 42 IL-testcrosses (IL-TCs) to the recurrent parent, and the controls twice (P_1, P_2, F_{1-a}, F_{1-b}). The position within the subplot was random. In addition, 'unselected' ILs without IL-TCs were grown in the same experiment. In this case, each subplot consisted of 56 ILs and 36 controls.

Molecular markers and linkage map

The RIL and IL mapping populations were genotyped with a set of 110 framework SNP markers (Törjék et al. 2003).

A linkage map was created from the RIL genotypes as described in detail in Törjék et al. (2006). Additional SNP markers were added to the framework map using Joinmap version 3.0 (Van Ooijen and Voorrips 2001) (Supplementary Fig. 1).

Determination of leaf area and shoot dry biomass

Images of plants were taken 6, 8 and 10 days after sowing (DAS) and whole leaf area was determined using the imaging system and software described in Walter et al. (2007). Due to different light source selection, parameter settings in the automated part of the analysis were adapted accordingly. Editing tools allowing for manual post-processing like removal of false positives (e.g. due to light reflection on vermiculite) were adapted to plant arrangement and their user-friendliness improved.

Shoot dry biomass was determined 15 DAS. The harvested aerial parts of the plants were placed in a vacuum oven at 80°C for 48 h. Dry biomass was measured using an analysis balance. Mean shoot dry biomass in mg plant^{-1} and mean leaf area in mm^2 plant^{-1} were estimated using a linear mixed model as described in Meyer et al. (2007).

Calculation of biomass heterosis

F_1 mid-parent-heterosis (MPH) was calculated from the estimated mean values of parents and reciprocal F_1-hybrids using the equation $\mathrm{MPH} = 100(\overline{F_1} - \overline{P})/\overline{P}$, where $\overline{F_1} = (F_{1-a} + F_{1-b})/2$ and $\overline{P} = (P_1 + P_2)/2$ (Falconer and Mackay 1996). The corresponding testcross relative mid-parent-heterosis of RIL-TCs was calculated as $r\mathrm{MPH}_{P_1} = 100(\mathrm{TC}_{P_1,i} - 0.5(\mathrm{RIL}_i + \overline{P_1}))/0.5(\mathrm{RIL}_i + \overline{P_1})$ for crosses to C24 and as $r\mathrm{MPH}_{P_2} = 100(\mathrm{TC}_{P_2,i} - 0.5(\mathrm{RIL}_i + \overline{P_2}))/0.5(\mathrm{RIL}_i + \overline{P_2})$ for crosses to Col-0. For statistical analyses absolute testcross mid-parent-heterosis values were calculated from RIL-TCs with C24 as $a\mathrm{MPH}_{P_1} = \mathrm{TC}_{P_1,i} - 0.5(\mathrm{RIL}_i + \overline{P_1})$ and with Col-0 as $a\mathrm{MPH}_{P_2} = \mathrm{TC}_{P_2,i} - 0.5(\mathrm{RIL}_i + \overline{P_2})$ (Lamkey and Edwards 1999). To determine the contribution of different genetic effects to MPH of the quantitative trait under study, the linear contrasts for augmented additive effect (additive effect confounded with epistasis, $Z_1 = \mathrm{TC}_{P_1,i} + \mathrm{TC}_{P_2,i}$) and augmented dominance effect (dominance effect confounded with epistasis, $Z_2 = \mathrm{TC}_{P_1,i} - \mathrm{TC}_{P_2,i}$) as defined by Melchinger et al. (2007b) were calculated.

Estimation of heritability

Broad-sense heritability (h^2) for dry weight and leaf area in the RILs was calculated as $\mathrm{Var}_G/(\mathrm{Var}_G + \mathrm{Var}_E)$, where $\mathrm{Var}_G =$ genotypic variance and $\mathrm{Var}_E =$ environmental variance estimated using the method of restricted maximum likelihood (REML) with genotype as random factor.

Correlation analyses

Procedures FCORRELATION and PCORRELATION in GenStat Release 10.2 were used for correlation analyses.

QTL analyses

Composite interval mapping (CIM) in RILs was performed using the software package PLABQTL (Utz and Melchinger 1996) with an extension for calculation of the Bayesian information criterion (BIC) according to the method of Burnham and Anderson (2004) to accommodate selection of cofactors and comparison of the models with and without digenic epistatic interactions (Kusterer et al. 2007b). Cofactors were automatically selected by forward stepwise regression. Significant LOD thresholds were determined by 5000 permutations. Data were also submitted to QTLNetwork-2.0 (Yang et al. 2008) to validate main-effect QTL and to perform a 2D-genome scan. QTLNetwork-2.0 implements a mixed-model based composite interval mapping method and allows simultaneous mapping of main-effect and interaction QTL. The genetic map used in all QTL analyses is based on the map presented by Törjék et al. (2006), with additional SNP markers (Supplementary Fig. 1). The augmented degree of dominance $(d_i^*/|a_i^*|)$ of significant per se or heterotic QTL was determined by estimating Z_1 and Z_2 at the respective positions (Kusterer et al. 2007b). QTL were classified according to their augmented dominance ratio as additive (A; $|d^*/a^*| < 0.2$), partially dominant (PD; $0.2 \leq |d^*/a^*| < 0.8$), dominant (D; $0.8 \leq |d^*/a^*| < 1.2$), and overdominant (OD; $|d^*/a^*| \geq 1.2$) (Stuber et al. 1987). In contrast to Kusterer et al. (2007b), our experimental design did not include testcrosses of the RILs to the F_1, therefore epistatic effects (linear contrast Z_3) could not be estimated.

Heterosis data obtained in the IL population were analyzed using ANOVA in GenStat Release 10.2 with the appropriate contrasts, i.e. comparing the IL-TC with the mean of IL and recurrent parent.

Generation means analysis

An overall generation means analysis (GMA) including parental lines, F_1-hybrids, RILs and RIL-TCs was performed as described in Kusterer et al. (2007b). All parameters in the models were defined according to the F_2-metric (Cockerham 1954; Yang 2004). Neither cytoplasmic nor maternal effects were significant ($P < 0.01$) in the exploratory analysis and were excluded from further analyses. Model 1 contained the overall mean μ, the overall

additive effect [a] and the overall dominance effect [d]. In addition to these parameters, model 2 also included epistatic effects between unlinked pairs of loci [aa]. Adequacy of each model was tested with a Chi-square (χ^2) test (Mather and Jinks 1982). The coefficient of determination (R^2) was calculated to estimate the proportion of the variation among generation means accounted for by each model.

In the second experiment analyzing introgression lines, genetic effects for individual ILs were estimated by GMA as described in Melchinger et al. (2007a), but using only ILs and their testcrosses to the recurrent parent.

Qualitative mode-of-inheritance classification of QTL

To complement the generation means analysis of ILs, we performed QTL mapping and a classification of the mode-of-inheritance according to Semel et al. (2006) as described in Kusterer et al. (2007a). Each IL and IL-TC was compared by a t-test with the recurrent parent P_i as well as to each other. If either of them was significantly different from parent P_i, the corresponding IL was considered as harboring a QTL. Because the number of replications for the IL and IL-TC was identical but the number of replications of the parental lines was higher, the comparison of IL and IL-TC with the parent was tested at a significance level of $P < 0.01$, whereas the comparison of IL with IL-TC was tested at $P < 0.05$, adopting the procedure of Semel et al. (2006).

Results

Description of traits

We analyzed performance data [dry weight at 15 DAS (DW15) and leaf area at 6, 8, 10 DAS (LA06, LA08, LA10)], as well as data sets calculated from adjusted mean values of the performance data in RILs and RIL-TCs (relative mid-parent heterosis rMPH, absolute mid-parent heterosis aMPH, linear contrast Z_1 and linear contrast Z_2). Heritabilities (h^2) for the performance traits were calculated as proportion of total variance allocated to genotype (Supplementary Table 1). Values were moderate for DW15 (42%) and moderately high for leaf area (58, 59, 53%). Means and variances for all traits averaged over the RILs are given in Supplementary Tables 1 and 2. The distributions of the performance traits show a distinct shift toward higher values in the RIL-TCs compared to RILs for both leaf area and dry weight (Fig. 1). This is reflected in the mid-parent-heterosis values (Table 1). Significant ($P < 0.001$) differences in the performance (DW15, LA06, LA08, LA10) were detected between the two testcross
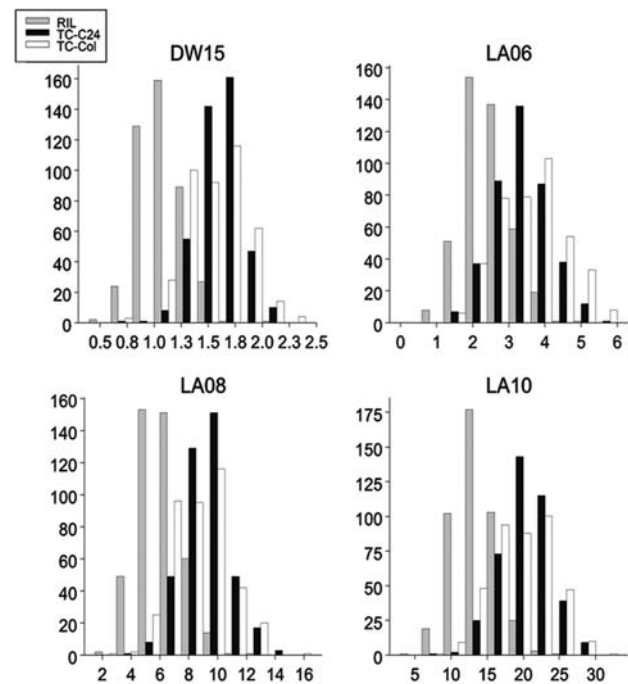
Fig. 1 Distribution of dry weight and leaf area in the RIL and RIL-TC populations. Shown are the distributions of the means (x axis) for dry weight at 15 DAS (mg plant^{-1}) and leaf area at 6, 8, 10 DAS (mm^2 plant^{-1}). The y axis indicates the number of lines

Table 1 Biometrical analyses of phenotypic data for the six traits in F_1 and RILs

	Trait	MPH (%) F_1	Mean rMPH (%) RILs	rMPH (%) Range in RILs
MPH-C24	DW15	43.71	36.60 ± 18.39	−31.13 to 88.32
	LA06	28.11	29.68 ± 32.10	−45.38 to 126.50
	LA08	24.43	28.42 ± 25.14	−50.48 to 105.60
	LA10	26.15	25.94 ± 22.71	−49.96 to 92.15
MPH-Col	DW15	43.07	34.45 ± 24.99	−36.33 to 99.10
	LA06	32.45	28.59 ± 34.95	−57.93 to 133.03
	LA08	22.02	28.20 ± 30.67	−52.54 to 142.94
	LA10	18.55	26.44 ± 28.01	−46.35 to 92.87

Mid-parent-heterosis (MPH) for the original cross (F_1, with $F_{1-Col \times C24}$ in MPH-C24 and $F_{1-C24 \times Col}$ in MPH-Col), mean testcross relative MPH (rMPH) calculated from RILs and RIL-TCs and averaged over the whole population (RIL) and the range of rMPH detected using RILs and RIL-TCs (range)

populations but not the reciprocal RIL populations using Kolmogorov-Smirnov two-sample tests (Conover 1971).

Correlation analyses

Pearson correlation coefficients between dry weight and leaf area within RILs were positive and moderately high (Table 2). The respective means across both RIL-TCs

Table 2 Correlation coefficients between RIL per se traits (below diagonal) and between RIL per se and the mean across RIL-TCs (equivalent to Z_1) (diagonal)

	DW15	LA06	LA08	LA10
DW15	*0.17****			
LA06	0.72***	*0.11**		
LA08	0.75***	0.93***	*0.10**	
LA10	0.73***	0.91***	0.96***	*0.15***

***,**,* Significant at $P < 0.001$, $P < 0.01$, $P < 0.05$, respectively

Table 3 Mean absolute correlation coefficients between heterosis (aMPH$_{C24}$, aMPH$_{Col}$) and linear transformations (Z_1, Z_2), and marker heterozygosity

Trait	aMPH$_{C24}$	aMPH$_{Col}$	Z_1	Z_2
DW15	0.342***	0.204***	0.063^{ns}	0.436***
LA06	0.170**	0.122*	0.037^{ns}	0.258***
LA08	0.195***	0.107^{ns}	0.022^{ns}	0.271***
LA10	0.194***	0.119*	0.022^{ns}	0.283***

ns Not significant

***,**,* Significant at $P < 0.001$, $P < 0.01$, $P < 0.05$, respectively

(equivalent to Z_1) showed significant but weak positive correlations with the RIL per se performance (Table 2). The linear contrast Z_1 shows no correlation to marker heterozygosity, whereas the linear contrast Z_2 and aMPH show moderate and weak correlations, respectively, with marker heterozygosity (Table 3).

Analyses of RILs and RIL-TCs

A global generation means analysis for dry weight and leaf area was performed on parental lines, F_1-hybrids, RILs and RIL-TCs (Table 4). Model 1 (not including additive × additive effects) accounted for 96–98%, and model 2 (including [aa] for 99% of the variation among generation means for all traits. The χ^2 values for the goodness-of-fit of both models were significant. Both models yielded similar estimates of [a] and [d]. Dominance effects were significant ($P < 0.01$) and had a positive sign for all traits. Estimates of [aa] under model 2 were significant ($P < 0.01$) and positive for all traits.

In order to identify genomic regions responsible for the occurrence of heterosis in the Col-0/C24 cross, absolute mid-parent-heterosis (aMPH) and the linear contrasts Z_1 and Z_2 for shoot biomass and leaf area were used as traits in QTL analyses (composite interval mapping) in PLABQTL and QTLNetwork-2.0. A complete list and description of QTL detected for biomass heterosis and leaf growth is given in Supplementary Table 3. For biomass heterosis at 15 DAS, six QTL for aMPH$_{C24}$, one QTL for aMPH$_{Col}$ and six QTL for Z_2 were found, jointly accounting for 20.6 ± 3.5, 18.0 ± 3.4 and $30.5 \pm 3.8\%$ of the phenotypic variation, respectively. Seven QTL for DW15 per se explain $20.2 \pm 3.5\%$, and four QTL for Z_1 account for $18.9 \pm 3.5\%$ of the phenotypic variation. Individual QTL effects range from 1.95 to 18.03%. We could also detect per se and heterotic QTL for leaf area, with joint R^2 ranging from 2.4 to 17.6% (Supplementary Table 3). Augmented dominance ratios for significant QTL were

Table 4 Generation means analysis using RILs and RIL-TCs

Parameters	DW15	LA06	LA08	LA10
Model 1				
μ	1.56 ± 0.02**	3.61 ± 0.07**	8.99 ± 0.19**	20.25 ± 0.45**
[a]	-0.01 ± 0.04	-0.15 ± 0.14	-0.33 ± 0.38	-0.42 ± 0.87
[d]	0.90 ± 0.06**	2.21 ± 0.19**	5.78 ± 0.53**	13.06 ± 1.23*
R^2 (%)	98.2	96.3	96.0	95.9
$\chi^2_{(5)}$ [a]	103.0****	76.6****	127.0****	144.9****
Model 2				
μ	1.52 ± 0.02**	3.45 ± 0.05**	8.48 ± 0.15**	19.09 ± 0.26**
[a]	-0.01 ± 0.02	-0.16 ± 0.06	-0.32 ± 0.17	-0.40 ± 0.31
[d]	0.82 ± 0.03**	1.96 ± 0.21**	4.91 ± 0.32**	11.12 ± 0.55**
[aa]	0.20 ± 0.05*	0.67 ± 0.15*	2.19 ± 0.49*	4.95 ± 0.85**
R^2 (%)	99.6	99.4	99.3	99.6
$\chi^2_{(4)}$ [a]	20.7***	13.3*	21.6***	15.2*

Generation means were calculated without cytoplasma effect c, which was not significant in the exploratory analysis. Parameters included are overall mean μ, overall additive effect [a], overall dominance effect [d], overall additive × additive epistasis [aa] (model 2 only)

****,***,**,* Significant at $P < 0.0001$, $P < 0.001$, $P < 0.01$, $P < 0.05$, respectively

[a] χ^2 statistic with degrees of freedom in parentheses

Table 5 Augmented degree of dominance of significant QTL estimated from linear contrasts Z_1 and Z_2

Interval	DW						LA06					
	per se	aMPH$_{C24}$	aMPH$_{Col}$	Z_1	Z_2	Gene action	per se	aMPH$_{C24}$	aMPH$_{Col}$	Z_1	Z_2	Gene action
1/4–28	0.15	−0.15			−0.20	OD	0.16					OD
1/82–102	−0.13	−0.22		−0.30		A	−0.13			−0.23		A
2/42–60		−0.20			−0.20	OD						
3/0–12	0.26					PD	0.16					D
3/32–38					−0.14	OD						
3/50–72	−0.21	−0.29		−0.18	−0.26	PD		−0.18			−0.23	D
3/82–86												
4/0–13	−0.22	0.13	0.44	0.27	-0.30	PD	−0.15		0.37	0.25	−0.23	PD
4/30–72	−0.20			−0.17		PD	−0.27					PD
5/32–42											−0.16	OD
5/70–94	0.19	−0.20			−0.20	OD	0.16					PD

| Interval	LA08						LA10					
	per se	aMPH$_{C24}$	aMPH$_{Col}$	Z_1	Z_2	Gene action	per se	aMPH$_{C24}$	aMPH$_{Col}$	Z_1	Z_2	Gene action
1/4–28	0.17					OD	0.15					OD
1/82–102		−0.18		−0.23		A	−0.17	−0.19		−0.26		A
2/42–60												
3/0–12	0.24					D	0.22					D
3/32–38												
3/50–72		−0.18			−0.26	D		−0.16			−0.23	D
3/82–86							−0.19					OD
4/0–13			0.37	0.25	−0.23	PD			0.36	0.26	−0.23	PD
4/30–72	−0.26					PD	−0.19			−0.15		PD
5/32–42					−0.16	OD					−0.18	OD
5/70–94	0.19					OD						

Only significant effects ($P < 0.01$) are shown. A negative value signifies an increasing effect of the Col-0 allele. The dominance ratio $d*/|a*|$ for all significant QTL was calculated using effects estimated in Z_1 ($a*$) and Z_2 ($d*$). QTL were classified as additive (A; $|d*/a*| < 0.2$), partially dominant (PD; $0.2 \leq |d*/a*| < 0.8$), dominant (D; $0.8 \leq |d*/a*| < 1.2$), and overdominant (OD; $|d*/a*| \geq 1.2$)

estimated from the linear contrasts Z_1 and Z_2 (Table 5). For shoot biomass four of six QTL detected in Z_2 show overdominant effects and two QTL show partial dominance. The QTL for aMPH$_{C24}$ at position 1/90, where there is no corresponding QTL for Z_2, has additive effect. Gene actions defined for leaf area are consistent with the biomass data. We tested for presence of digenic epistatic effects between all pairs of marker loci using QTLNetwork-2.0, and could find two significant additive × additive interactions between unlinked pairs of markers on chromosomes 1 and 5 for DW15 and LA08, and on chromosomes 1 and 4 for DW15, LA08 and LA10, respectively (Supplementary Table 4).

Verification of detected QTL in ILs and IL-TCs

The QTL detected in the RIL population were verified using lines of the two corresponding, reciprocal IL populations (Törjék et al. 2008). In a first step, heterosis data were analyzed as linear contrasts in ANOVA. All six biomass heterosis QTL at positions 1/12, 1/92, 2/46–56, 3/56–62, 4/4, 5/74 could be verified (Supplementary Table 5). Additional biomass heterosis QTL could be identified in the intervals 1/65–68, 1/77–83, 2/74, 3/74–79, 4/60–78 and 5/14–44. Most (80.5%) IL/IL-TC combinations showed significant heterosis (Fig. 2), revealing a highly complex situation. The IL data were also analyzed using the generation means approach (Melchinger et al. 2007a). Significant dominance effects were detected for previously identified heterotic QTL at positions 1/90–92, 2/46, 2/56, 3/56–62 and 4/4–6 (Supplementary Table 5). Significant additive effects were detected for previously identified biomass QTL (Lisec et al. 2008) at positions 3/60 and 5/86. The mode-of-inheritance classification according to Semel et al. (2006) identified a total of 12 heterotic (overdominant) QTL, across both reciprocal IL subpopulations, most with a positive overdominance effect (Supplementary Table 6).
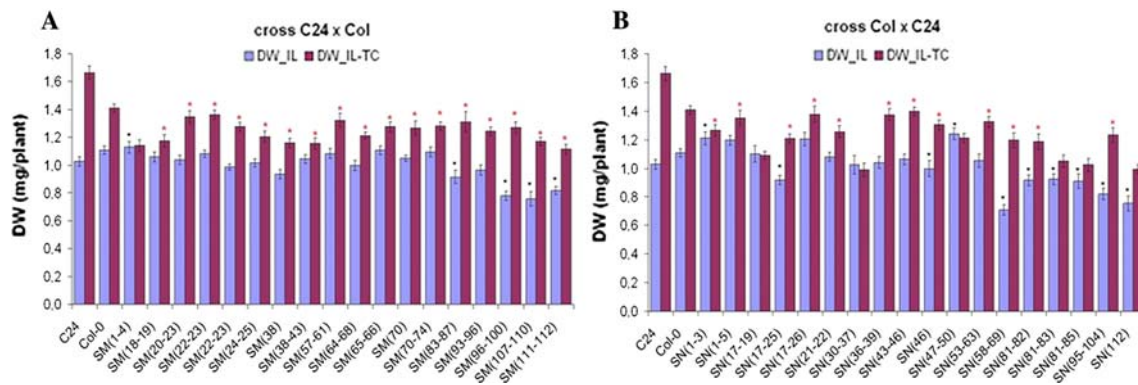
Fig. 2 Biomass in ILs and ILs-TC from the C24xCol (**a**) and ColxC24 (**b**) IL populations. Mean values and standard deviations of 24 replicates are shown. ILs are denoted with the substituted marker intervals, with SM indicating C24 background and SN Col-0

background. ILs with biomass significantly ($P < 0.05$) different from the recurrent parent are indicated by a black asterisk; IL-TC displaying significant ($P < 0.05$) mid-parent-heterosis are marked with a *red asterisk* above the *bars*

Discussion

The distributions of the performance traits show clear differences between RILs and their testcrosses for dry weight and leaf area, identifying them as heterotic traits. Testcross relative mid-parent-heterosis (rMPH) for all traits behaves like a segregating trait in the RILs, displaying a broad range of values and thus making it possible to use it as a trait in QTL analyses (Stuber et al. 1992). The two testcross populations differ significantly for all performance traits. The differences are also reflected by the fact that only 1 (of 10) heterotic QTL region, at the top of chromosome 4, contributes to both $aMPH_{C24}$ and $aMPH_{Col}$ for dry weight. This can be explained by different heterozygous regions involved in the realization of heterosis between the two populations. Xiao et al. (1995) described a similar situation in rice, where they detected most of the QTL (72%) in only one of the two testcrosses. The authors concluded that dominance complementation is the major genetic basis of heterosis in rice. In a recent maize study, one testcross was significantly higher for all eight traits analyzed, confirming the prevalence of alleles with increasing effect in one of the parents (Frascaroli et al. 2007). The situation is not as clear-cut in the analyzed Arabidopsis populations, with both parents contributing 'increasing alleles' as illustrated by the observed transgressive segregation. In our study, augmented dominance ratios of the detected QTL and mode of inheritance classification provide evidence that both dominance and overdominance are involved in biomass heterosis in Arabidopsis. We identified more QTL for performance per se than for Z_1. This is consistent with the low correlation detected between the two measures. Z_1 represents the augmented additive effect, i.e. the additive effect for the respective QTL minus half the sum of dominance × additive epistatic interactions (Melchinger et al. 2007b). The discrepancy between QTL detected for performance per se

and for Z_1 points toward the existence of dominance × additive epistasis, which reduces the power of Z_1 to detect QTL.

Treating mid-parent-heterosis as a quantitative trait, we could identify heterotic QTL for biomass and leaf area. As expected from the high phenotypic correlation between dry weight and leaf area we identified similar performance and heterotic QTL for these traits, not only concerning position, but also effect, increasing allele and gene action, with LA10 being most similar to DW15. There is also good coincidence between QTL detected for aMPH and Z_2, in agreement with the correlation coefficients. In these cases, the QTL representing the augmented dominance effect (dominance effect for respective QTL minus half the sum of additive × additive epistatic interactions (Melchinger et al. 2007b)) for the trait under consideration also indicates a heterotic effect. QTL detected for transformation Z_2 show at least partial dominance; most (4 of 6) have an overdominant gene action. The additive effect estimated for the aMPH QTL for DW15 at position 1/90 (no overlap with Z_2) can be attributed to the fact that aMPH is confounded with additive effect (Melchinger et al. 2007b). The same genomic region was found to significantly influence heterosis in the IL/IL-TC study, with overdominant gene action identified in the generation means and mode of inheritance analyses. Positive dominance effects can be counterbalanced by positive additive × additive effects, thus reducing the power of Z_2 to detect heterotic QTL (Melchinger et al. 2007b). The existence of positive additive × additive effects is suggested by the overall generation means analysis in the RILs and substantiated by the mixed model QTL analysis. In GMA the model including epistasis detected significant dominance and additive × additive effects for all traits. The inclusion of epistasis in the model considerably improved the fit, explaining around 99% of the variation among generation means for all traits (substantially increased R^2,
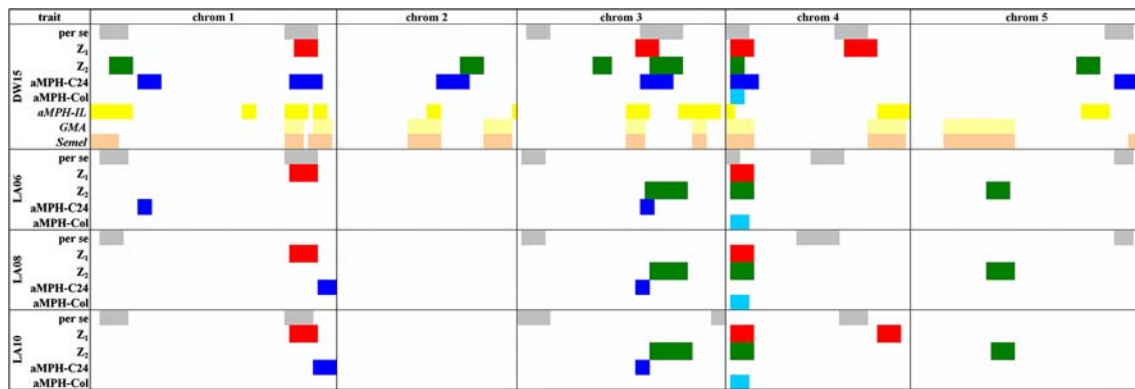
Fig. 3 Schematic representation of growth related per se and heterotic QTL. QTL positions are indicated by *boxes* corresponding to the intervals covered by the respective confidence intervals. *DW15* Shoot dry weight at 15 DAS, Z_1 augmented additive effect, Z_2 augmented dominance effect, *aMPH*$_{C24}$, *aMPH*$_{Col}$ absolute mid-parent-heterosis in cross with C24 and Col-0, respectively, *LAn* leaf area at n DAS

while the χ^2 test remained significant). Furthermore, Kusterer et al. (2007b) using a triple test cross approach on the same RIL population at a later growth stage, did detect additive × additive epistatic effects. Overdominance and epistasis have been reported to play an important role in heterosis in a number of organisms for various traits (reviewed in Lippman and Zamir (2007)). In *Brassica rapa*, 15 of 23 QTL detected for biomass and associated traits showed overdominance, and 444 digenic epistatic interactions were found (Dong et al. 2007). Analyzing heterosis for yield and its component traits in *Brassica napus*, Radoev et al. (2008) concluded that epistasis together with all levels of dominance form the genetic basis of heterosis in this crop plant. QTL analyses of heterosis in maize revealed an association of QTL with additive to dominance effects with traits with low heterosis, whereas QTL for traits with high heterosis mostly showed dominance to overdominance effects (Frascaroli et al. 2007). While our experimental design did not allow us to estimate directly the involvement of digenic epistatic effects on heterosis via the transformation Z_3 proposed by Melchinger et al. (2007b), we found evidence for additive × additive epistasis between two pairs of loci. The interaction between loci on chromosomes 1 and 5, with parental allele combinations increasing biomass, does not include main-effect QTL. In contrast, the two interacting loci on chromosomes 1 and 4 have been identified previously as main-effect QTL. Furthermore, both have been detected as aMPH QTL and a recombinant allele combination increased biomass, consistent with a role in heterosis (increased performance of a double heterozygote over both parental homozygotes).

QTL for different traits and transformations co-locate in three main genomic regions on chromosomes 1, 3, and 4. In a parallel study (Lisec et al. 2009) focusing on metabolites we identified hot spots for heterotic metabolite QTL in similar positions. The cluster containing the strong heterotic

biomass QTL identified at position 4/4 corresponds to the 'hot spot' described for metabolic and biomass QTL in the same Arabidopsis population (Lisec et al. 2008) and co-locates with a cluster of 23 metabolic heterotic QTL described in the metabolic study (Lisec et al. 2009). Interestingly, this 'hot spot' is also involved in additive × additive epistasis (cf. Supplementary Table 3). This could indicate the presence of a major growth/metabolism regulator in this region, although we cannot rule out tightly linked QTL, as demonstrated by Kroymann and Mitchell-Olds (2005) who detected two interacting QTL within 210 kb. The 'hot spot' on top of chromosome 4 currently spans 155 kb and contains 66 genes, including the *FRIGIDA* gene (*FRI*, At4g00650), a well-known major determinant of natural variation in Arabidopsis for flowering time (Johanson et al. 2000). However, the location of the QTL peak and preliminary analyses of subILs covering this region suggest a minor role for FRI in the manifestation of heterosis.

The parallel use of RIL and IL populations and several analysis methods proved very beneficial, as both populations and the application of the various methods complemented each other. This did not only allow us to validate QTL found in one population, but also to detect additional QTL in the other. The main difference between RILs and ILs is the absence of variation for 'background' epistasis in the ILs, thereby increasing the power to detect QTL (Keurentjes et al. 2007; Reif et al. 2009). However, as epistasis has been shown to play an important role in heterosis, either both types of populations (Keurentjes et al. 2007), or specially selected double or multiple introgression ILs (Reif et al. 2009) are needed to detect and quantify digenic, or higher order interactions.

In our study QTL for growth traits co-locate at only two positions (on chromosome 3) with growth QTL (Z_1) identified previously in the same populations in a different environment at a later developmental stage (Kusterer et al. 2007b).

The specifically detected QTL indicate that different factors are of relevance under different conditions. In contrast, eight of the ten heterotic QTL (Z_2 and Z_3) reported by Kusterer et al. (2007b) and Melchinger et al. (2007a) coincide with heterotic QTL identified in this study (cf. Fig. 3), including the main candidate regions at the top of chromosome 4 and bottom of chromosome 3. In these regions, the earlier study identified QTL for the linear contrast Z_3, estimating additive $\times$ additive epistatic effects, further underlining the importance of epistasis for heterosis. The complex trait 'heterosis' is expected to be controlled by many genes, the combination and interaction of which may depend on the organism and trait under study (Korn et al. 2008; Li et al. 2008). In each cross, only a subset of specific combinations may be relevant for heterosis, as illustrated by the seemingly contradictory findings concerning the genetic basis of heterosis reported in different studies (reviewed in Lippman and Zamir (2007)). The joint detection of heterotic QTL in our two studies suggests the existence of genomic regions of more general importance for the manifestation of heterosis. These regions constitute primary targets for further fine-mapping with the ultimate goal to identify and characterize the underlying genes. The identification of heterotic QTL in Arabidopsis and the superior tools for the identification of natural genetic determinants of trait variation (Peters et al. 2003; Weigel and Nordborg 2005), including the rapid advances in genome re-sequencing in this species (Ossowski et al. 2008) opens the opportunity to identify the DNA variation responsible for or contributing to heterosis.

References

Abdelkhalik AF, Shishido R, Nomura K, Ikehashi H (2005) QTL-based analysis of heterosis for grain shape traits and seedling characteristics in an indica-japonica hybrid in rice (Oryza sativa L.). Breed Sci 55:41–48

Alonso-Blanco C, Blankestijn-de Vries H, Hanhart CJ, Koornneef M (1999) Natural allelic variation at seed size loci in relation to other life history traits of Arabidopsis thaliana. Proc Natl Acad Sci USA 96:4710–4717

Barth S, Busimi AK, Utz HF, Melchinger AE (2003) Heterosis for biomass yield and related traits in five hybrids of Arabidopsis thaliana L. Heynh. Heredity 91:36–42

Birchler JA, Auger DL, Riddle NC (2003) In search of the molecular basis of heterosis. Plant Cell 15:2236–2239

Bruce AB (1910) The Mendelian theory of heredity and the augmentation of vigor. Science 32:627–628

Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res 33:261–304

Cockerham CC (1954) An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39:859–882

Conover WJ (1971) Practical nonparametric statistics. Wiley, New York

Crow JF (1948) Alternative hypotheses of hybrid vigor. Genetics 33:477–487

Crow JF (1952) Dominance and overdominance. In: Gowen JW (ed) Heterosis. Iowa State College Press, Ames, pp 282–297

Davenport C (1908) Degeneration, albinism and inbreeding. Science 28:454–456

Dong D-K, Cao J-S, Kai S, Liu L-C (2007) Overdominance and epistasis are important for the genetic basis of heterosis in Brassica rapa. HortScience 42:1207–1211

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Prentice-Hall, Harlow

Frascaroli E, Canè MA, Landi P, Pea G, Gianfranceschi L, Villa M, Morgante M, Pè ME (2007) Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. Genetics 176:625–644

Griffing B, Langridge J (1963) Phenotypic stability of growth in the self-fertilized species Arabidopsis thaliana. In: Hanson WD, Robinson HF (eds) Statistical genetics and plant breeding. National Research Council Publication, Washington, DC, pp 368–394

Hull FH (1945) Recurrent selection for specific combining ability in corn. Agron J 37:134–145

Johanson U, West J, Lister C, Michaels S, Amasino R, Dean C (2000) Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time. Science 290:344–347

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-de Vries H, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population. Genetics 175:891–905

Korn M, Peterek S, Mock H-P, Heyer AG, Hincha DK (2008) Heterosis in the freezing tolerance, and sugar and flavonoid contents of crosses between Arabidopsis thaliana accessions of widely varying freezing tolerance. Plant Cell Environ 31:813–827

Kroymann J, Mitchell-Olds T (2005) Epistasis and balanced polymorphism influencing complex trait variation. Nature 435:95–98

Kusterer B, Muminovic J, Utz HF, Piepho H-P, Barth S, Heckenberger M, Meyer RC, Altmann T, Melchinger AE (2007a) Analysis of a triple testcross design with recombinant inbred lines reveals a significant role of epistasis in heterosis for biomass-related traits in Arabidopsis. Genetics 175:2009–2017

Kusterer B, Piepho H-P, Utz HF, Schön CC, Muminovic J, Meyer RC, Altmann T, Melchinger AE (2007b) Heterosis for biomass-related traits in Arabidopsis investigated by quantitative trait loci analysis of the triple testcross design with recombinant inbred lines. Genetics 177:1839–1850

Lamkey KR, Edwards JW (1999) The quantitative genetics of heterosis. In: Coors JG, Pandey S (eds) Genetics and exploitation of heterosis in crops. American Society of Agronomy, Madison, pp 31–48

Li ZK, Luo LJ, Mei HW, Wang DL, Shu QY, Tabien R, Zhong DB, Ying CS, Stansel JW, Khush GS, Paterson AH (2001) Over-dominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. Genetics 158:1737–1753

Li L, Lu K, Chen Z, Mu T, Hu Z, Li X (2008) Dominance, over-dominance and epistasis condition heterosis in two heterotic rice hybrids. Genetics 180:1725–1742

Lippman ZB, Zamir D (2007) Heterosis: revisiting the magic. Trends Genet 23:60–66

Lisec J, Meyer RC, Steinfath M, Redestig H, Becher M, Witucka-Wall H, Fiehn O, Törjék O, Selbig J, Altmann T, Willmitzer L (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations. Plant J 53:960–972

Lisec J, Steinfath M, Meyer RC, Selbig J, Melchinger AE, Willmitzer L, Altmann T (2009) Identification of heterotic metabolite QTL in Arabidopsis thaliana RIL and IL populations. Plant J. doi: 0.1111/j.1365-313X.2009.03910.x

Luo LJ, Li ZK, Mei HW, Shu QY, Tabien R, Zhong DB, Ying CS, Stansel JW, Khush GS, Paterson AH (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. Genetics 158:1755–1771

Mather K, Jinks JL (1982) Biometrical genetics. Chapman & Hall, London

Mei HW, Li ZK, Shu QY, Guo LB, Wang YP, Yu XQ, Ying CS, Luo LJ (2005) Gene actions of QTLs affecting several agronomic traits resolved in a recombinant inbred rice population and two backcross populations. Theor Appl Genet 110:649–659

Melchinger AE, Piepho H-P, Utz HF, Muminovic J, Wegenast T, Törjék O, Altmann T, Kusterer B (2007a) Genetic basis of heterosis for growth-related traits in Arabidopsis investigated by testcross progenies of near-isogenic lines reveals a significant role of epistasis. Genetics 177:1827–1837

Melchinger AE, Utz HF, Piepho HP, Zeng ZB, Schön CC (2007b) The role of epistasis in the manifestation of heterosis: a systems-oriented approach. Genetics 177:1815–1825

Meyer RC, Törjék O, Becher M, Altmann T (2004) Heterosis of biomass production in Arabidopsis: establishment during early development. Plant Physiol 134:1813–1823

Meyer RC, Steinfath M, Lisec J, Becher M, Witucka-Wall H, Törjék O, Fiehn O, Eckardt Ä, Willmitzer L, Selbig J, Altmann T (2007) The metabolic signature related to high plant growth rate in Arabidopsis thaliana. Proc Natl Acad Sci USA 104:4759–4764

Mitchell-Olds T (1995) Interval mapping of viability loci causing heterosis in Arabidopsis. Genetics 140:1105–1109

Monforte AJ, Tanksley SD (2000) Fine mapping of a quantitative trait locus (QTL) from Lycopersicon hirsutum chromosome 1 affecting fruit characteristics and agronomic traits: breaking linkage among QTLs affecting different traits and dissection of heterosis for yield. Theor Appl Genet 100:471–479

Narang RA, Altmann T (2001) Phosphate acquisition heterosis in Arabidopsis thaliana: a morphological and physiological analysis. Plant Soil 234:91–97

Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res. doi:10.1101/gr.080200.108

Peters JL, Cnudde F, Gerats T (2003) Forward genetics and map-based cloning approaches. Trends Plant Sci 8:484–491

Powers L (1944) An expansion of Jones's theory for the explanation of heterosis. Am Nat 78:275

Radoev M, Becker HC, Ecke W (2008) Genetic analysis of heterosis for yield and yield components in rapeseed (Brassica napus L.) by quantitative trait locus mapping. Genetics 179:1547–1558

Reif JC, Kusterer B, Piepho H-P, Meyer RC, Altmann T, Schön CC, Melchinger AE (2009) Unravelling epistasis with triple testcross progenies of near isogenic lines. Genetics 181:247–257

Rohde P, Hincha DK, Heyer AG (2004) Heterosis in the freezing tolerance of crosses between two Arabidopsis thaliana accessions (Columbia-0 and C24) that show differences in non-acclimated and acclimated freezing tolerance. Plant J 38:790–799

Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, Pleban T, Lippman Z, Gur A, Zamir D (2006) Overdominant quantitative trait loci for yield and fitness in tomato. Proc Natl Acad Sci USA 103:12981–12986

Sharma RK, Griffing B, Scholl RL (1979) Variations among races of Arabidopsis thaliana for survival in limited carbon dioxide. Theor Appl Genet 54:11–16

Shull GH (1948) What is heterosis? Genetics 33:439–446

Stokes D, Morgan C, O'Neill C, Bancroft I (2007) Evaluating the utility of Arabidopsis thaliana as a model for understanding heterosis in hybrid crops. Euphytica 156:157–171

Stuber CW, Edwards MD, Wendel JF (1987) Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. Crop Sci 27:639–648

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Syed NH, Chen ZJ (2004) Molecular marker genotypes, heterozygosity and genetic interactions explain heterosis in Arabidopsis thaliana. Heredity 94:295–304

Törjék O, Berger D, Meyer R, Müssig C, Schmid K, Rosleff-Sörensen T, Weisshaar B, Mitchell-Olds T, Altmann T (2003) Establishment of a high-efficiency SNP-based framework marker set for Arabidopsis. Plant J 36:122–140

Törjék O, Witucka-Wall H, Meyer RC, von Korff M, Kusterer B, Rautengarten C, Altmann T (2006) Segregation distortion in Arabidopsis C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. Theor Appl Genet 113:1551–1561

Törjék O, Meyer RC, Zehnsdorf M, Teltow M, Strompen G, Witucka-Wall H, Blacha A, Altmann T (2008) Construction and analysis of 2 reciprocal Arabidopsis introgression line populations. J Hered 99:396–406

Utz HF, Melchinger AE (1996) PLABQTL: a program for composite interval mapping of QTL. J QTL 2

Van Ooijen JW, Voorrips RE (2001) JoinMap® version 3.0: software for the calculation of genetic linkage maps

Walter A, Scharr H, Gilmer F, Zierer R, Nagel KA, Ernst M, Wiese A, Virnich O, Christ MM, Uhlig B, Jünger S, Schurr U (2007) Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species. New Phytol 174:447–455

Weigel D, Nordborg M (2005) Natural variation in Arabidopsis. How do we find the causal genes? Plant Physiol 138:567–568

Williams W (1959) Heterosis and the genetics of complex characters. Nature 184:527–530

Xiao J, Li J, Yuan L, Tanksley SD (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745–754

Yang RC (2004) Epistasis of quantitative trait loci under different gene action models. Genetics 167:1493–1505

Yang J, Hu C, Hu H, Yu R, Xia Z, Ye X, Zhu J (2008) QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. Bioinformatics 24:721–723

Yu CY, Wan JM, Zhai HQ, Wang CM, Jiang L, Xiao YH, Liu YQ (2005) Study on heterosis of inter-subspecies between indica and japonica rice (*Oryza sativa* L.) using chromosome segment substitution lines. Chin Sci Bull 50:131–136