Genet. Res., Camb. (2005), **86**, *pp.* 97–106. *With 6 figures.* © 2005 Cambridge University Press
doi:10.1017/S0016672305007779 Printed in the United Kingdom

Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (*Oryza sativa* L.)

CHUNMING WANG[1†], CHENGSONG ZHU[1,2†], HUQU ZHAI[3] AND JIANMIN WAN[1,3*]
[1] *National Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China*
[2] *Department of Genetics, School of Life Science, Fudan University, Shanghai 200433, China*
[3] *Crop Science Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China*

(*Received 8 March 2005 and in revised form 29 July 2005*)

Summary

Markers with segregation ratio distortion are commonly observed in data sets used for quantitative trait locus (QTL) mapping. In this study, a multipoint method of maximum likelihood (ML) was newly developed to estimate the positions and effects of the segregation distortion loci (SDLs) in two F_2 populations of rice (*Oryza sativa* L.), i.e. Taichung65/Bhadua (TB; *japonica–indica* cross) and CPSLO17/W207-2 (CW; *japonica–japonica*). Of the four parents, W207-2 and Bhadua were found to be spikelet semi-sterile and stably inherited through selfing, and spikelet fertility segregated in the two populations. Therefore, recombination frequencies were recalculated after mapping the SDLs by using the multipoint method, and the molecular linkage maps of the two F_2 populations were constructed to detect QTLs underlying spikelet fertility. As a result, five SDLs in the TB population were mapped on chromosomes 1, 3, 8 and 9, respectively. Two major QTLs underlying spikelet fertility, namely *qSS-6a* and *qSS-8a*, were detected on chromosomes 6 and 8, respectively. In the CW population, a total of 12 SDLs were detected on all 12 chromosomes except 1, 5, 7 and 11. Three QTLs underlying spikelet sterility, namely *qSS-2, qSS-6b* and *qSS-8b* on chromosomes 2, 6 and 8, were determined on the whole genome scale. Interestingly, both *qSS-6a* and *qSS-6b*, detected in the two F_2 populations respectively, were located on a similar position as the *S5* gene on chromosome 6; while *qSS-8a* and *qSS-8b* were also simultaneously detected on similar positions of the short arm of chromosome 8 in the two populations, which should be a new sterility gene showing the same type of zygotic selection.

1. Introduction

Segregation distorting loci (SDLs), defined as chromosomal regions that cause distorted segregation ratios, are usually observed for some markers. Segregation distortion is known to bias estimation of recombination fractions in pair-wise analysis of markers (Lorieux *et al*., 1995). In the past, often a single marker was considered at a time, where only the linkage between one fully informative marker and a single SDL was tested (Xu *et al*., 1997; Harushima *et al*., 2001, 2002). Based on six F_2 populations of inter-subspecies crosses, chromosomal regions associated with marker segregation distortion in rice were determined and compared based on chi-square analysis ($P<0{\cdot}01$) given genotypic class within a segregating population (Xu *et al*., 1997). In a further study on reproductive barriers in F_2 populations regions causing deviation from Mendelian segregation ratios were mapped by a regression method and compared among three different *japonica–indica* crosses. Most of the barriers were mapped at different loci and demonstrated to form after *japonica–indica* differentiation (Harushima *et al*., 2001, 2002). In the single-marker chi-square test, the genotypic frequencies of the marker are affected by the recombination fraction in addition to the frequencies of the SDL's genotypic configurations. Hence, for a single-marker test, estimations of the position and effect are confounded.

* Corresponding author. Tel: +86 25 84396516. Fax: +86 25 84396516. e-mail: wanjm@njau.edu.cn
† The first two authors contributed equally to this paper.

Vogl & Xu (2000) pioneered development of a multipoint method for mapping multiple SDLs in the backcross design population. But there are several complications to be considered in practice. First, it is common to have less informative markers containing missing and dominant markers in experimental or natural populations, especially when fertility segregation exists. Second, segregation distortion is known to bias estimation of recombination distance between markers; Cheng *et al*. (1996) have already developed an expectation maximization (EM) algorithm to the position of two fully informative markers in the presence of a single SDL in the backcross population. When incomplete informative markers are used, the effects of the distortion are spread out to the much wider interval of fully informative markers flanking the distorted regions. Usually generation of a linkage map of marker loci precedes SDL analysis. The basic assumption of quantitative trait locus (QTL) analysis will not hold if the bias introduced by segregation distortion is worse. Hence, the recombination frequencies between consecutive markers and the recombination fraction between SDL and the adjacent markers should be re-estimated in the presence of SDL and missing or partial markers. Finally, Vogl & Xu's analysis of a backcross design population can be extended to the F_2 population; the dimension changes from two to three, binomial distribution changes to trinomial distribution and the transition probability matrix between adjacent loci changes. The changes are trivial but will complicate the analysis substantially.

In this paper, we derive a general algorithm to estimate systematically the position and the effects of SDLs with dominant and missing markers in the F_2 population. In particular, we re-estimate the recombination frequencies between consecutive markers and the recombination fraction between SDLs and the adjacent markers on the assumption that there are SDLs linked to markers, instead of excluding the markers that showed obvious segregation distortion in the previous study (Vogl & Xu, 2000). Our analysis is similar to the method of this previous study in spirit, but with derivation and sufficient details considering some complications in practice in the analysis. In addition, two different F_2 populations were constructed, one derived from a *japonica–indica* cross and the other one from *japonica–japonica*. Reproductive barriers causing deviation from Mendelian segregation ratios were mapped by the newly developed multipoint mapping method and compared between the two populations. With the linkage map constructed after multipoint mapping of SDLs, QTLs for spikelet fertility were analysed. Multiple reproductive barriers dispersed throughout the genome between *japonica–indica* and *japonica–japonica* crosses were compared through multipoint mapping of SDLs.

2. Materials and methods

(i) *Plant materials*

Two F_2 populations were derived from one cross of inter-subspecies of Taichung65/Bhadua (TB; *japonica–indica*) and one cross of two *japonica* parents CPSLO17/W207-2 (CW; *japonica–japonica*), respectively.

TB: Bhadua (*indica*) was found to be spikelet semi-sterile. It was crossed to Taichung65 as the female parent. A population of 96 F_2 plants was developed from the *indica–japonica* cross and planted together with the parental plants, F_1 hybrids in Kyushu University, Japan in 2001.

CW: A semi-sterile progeny line W207-2 (*japonica*) was crossed to CPSLO17. In the summer season of 2002, the parental plants, F_1 hybrids and F_2 population of 157 individuals were grown in the experimental field in Nanjing Agricultural University, China.

(ii) *Linkage map construction*

The CTAB method (Rogers & Bendish, 1988) was used with minor modifications for isolating total DNA from rice leaves. Simple sequence repeats (SSR) analysis was performed in Nanjing Agricultural University, China following the procedure of Chen *et al*. (1997) with minor modifications. The original sources and motifs for all the SSR markers used in this study are given in the gramene database (http://www.gramene.org/) and in the linkage maps constructed according to Temnykh *et al*. (2000) and McCouch *et al*. (2002). Amplification reactions were carried out in 10 μl containing 10 mM Tris-HCl pH 8·3, 50 mM KCl, 1·5 mM $MgCl_2$, 50 μM each of dNTPs, 0·2 μM each of primers, 0·5 U *Taq* polymerase (TaKaRa, Dalian, China) and 20 ng of DNA template. Reactions were performed using a PTC-200 thermal cycler (MJ Research) programmed at 94 °C for 5 min, followed by 35 cycles of 30 s at 94 °C, 30 s at 55 °C, and 1 min at 72 °C with a final extension of 7 min at 72 °C. Amplified products were separated in 8% non-denaturing PAGE gels, and observed by the silver staining method according to Sanguinetti & Simpson (1994). Restriction fragment length polymorphisms (RFLPs) between the parents were surveyed and 94 polymorphic markers were selected to construct the linkage map, of which the genotype data were provided by Dr Yasui in Kyushu University, Japan. Linkage groups and the order of SSR and RFLP markers were preliminarily determined using MAPMAKER/EXP 3.0 (Lander *et al*., 1987), and the linkage distances were recalculated using our newly developed method.

(iii) *SDL analysis*

We assume that all markers are neutral and codominant in the sense that their segregations would be

https://doi.org/10.1017/S0016672305007779 Published online by Cambridge University Press

Mendelian if there were no linked SDL on the same chromosome. The observed segregation distortions on these neutral markers, however, are caused by one or more SDLs near the markers. The degree of segregation distortion in the experimental population is determined by the effect (size) of the SDLs. Before estimating the effects and locations of SDLs, we test all the markers by chi-square with codominant markers according to Lorieux *et al.* (1995) with minor modification in order to infer whether the segregation distortion was caused by gametic or zygotic selection.

We consider an F_2 population from a cross between two inbred lines, P_1 and P_2. In the population, each marker or SDL has three possible genotypes. Let φ_{ij} be the indicator function, taking the value of 1 if the ith individual belongs to one of the jth possible genotypes in the F_2 population, otherwise taking the value of zero. Parameters of interest are the effects, denoted by π_j ($j=1, 2, 3$), and location, denoted by λ, of the segregation distorting locus. The distribution of φ_{ij} is

$$\Pr(\varphi_{ij}/\pi)=\pi_1^{\varphi i1}\pi_2^{\varphi i2}\pi_3^{\varphi i3}$$

for $i=1, \ldots, n$ with $\pi_j=\Pr(\varphi_{ij}=1)$, where n is the population size.

In the absence of segregation distortion, we have $\pi_1=0{\cdot}25$, $\pi_2=0{\cdot}5$ and $\pi_3=0{\cdot}25$. If φ_{ij} were observable, we could directly estimate and test π_j. The maximum likelihood estimate would be

$$\hat{\pi}_j=\frac{1}{n}\sum \Pr(\phi_{ij}=1) \text{ for } j=1, 2, 3.$$

However, φ_{ij} is not observable, while the phenotypic value marker genotypes are now observable. Therefore, an alternative model using multiple markers that are already mapped on the genome should be developed. Suppose there are m markers on a chromosome in the order of $M_1, M_2, \ldots, M_m$ with each marker having three possible genotypes. Let x_k denote the genotype of marker M_k for the ith individual, which takes a value 1, 0, -1 if the marker is homozygote of P_1 type, heterozygote or homozygote of P_2 type, respectively. Let Z_k be the phenotypic observation of marker M_k. The markers are under consideration for a linkage group under the assumption of no crossing-over interference; the likelihood is defined with matrix notation as

$$L=\prod_{i=1}^{n} q_1' H_{i1}(r_1)H_{i2}(r_2)\ldots H_{im-1}(r_{m-1})c$$

where

$$H(r_k)=\begin{bmatrix} (1-r_k)^2 & 2r_k(1-r_k) & r_k^2 \\ r_k(1-r_k) & (1-2r_k^2+2r_k^2) & r_k(1-r_k) \\ r_k^2 & 2r_k(1-r_k) & (1-r_k)^2 \end{bmatrix}$$

which denotes a transition probability matrix from M_k to M_{k+1}, where r_k is the recombination frequency between M_k and M_{k+1}, q_1 denotes the row vector of the prior probability $p(x_1)$, $q_1'=[\Pr(x_1=1), \Pr(x_1=0), \Pr(x_1=-1)]$ and $c'=[1,1,1]$, where $'$ denotes transposition (Jiang & Zeng, 1997). It is noted that we should specify appropriate marker matrix elements of zero depending on the information content of the phenotype of the marker, i.e. only one element in the transition matrix takes value and zero everywhere else when fully informative markers occurred.

The likelihood is a function of recombination frequencies between adjacent markers for a given marker linkage order without considering segregation distortion of neutral markers. Now we consider an SDL and the k'th SDL is located between markers k and $k+1$. The logarithm likelihood defined with matrix notation will be

$$\log L=\sum_{i=1}^{n}\{[H'(r_{ik,k'})\prod_{j=k-1}^{1} H'(r_{ij})q_1]'[H(r_{(ik',(k+1)})\times\prod_{j=k+1}^{m-1} H(r_{ij})c\}\Pr(\phi_{ij}/\pi)$$

There are several ways to find the maximum likelihood estimate of π_j. We adopt an expectation maximization (EM) algorithm and treat φ_{ij} as missing data. We treat λ as constant for the moment. As SDLs may affect the estimation of recombination frequencies between adjacent markers, the recombination frequencies between the adjacent markers and the recombination frequency between SDLs and the adjacent markers, the recombination frequencies at some or all intervals are all regarded as parameters and will be re-estimated.

E step:

Conditional on the data, the position and the initial value of the parameters, $\pi_j^{(0)}$ (for $j=1, 2, 3$) and the initial recombination fraction $r_k^{(0)}$ between consecutive markers computed by Mapmaker 3.0 (Lander *et al.*, 1987), the posterior probabilities of φ_{ij} are

$$\Pr(\phi_{ij}=1/\lambda, \pi_j)=\frac{\Pr(\phi_{i1}, \ldots\phi_{iM}/\phi_{ij}=1, \lambda)\pi_j^{(0)}}{\sum_{j=1}^{3}\Pr(\phi_{i1}, \ldots\phi_{iM}/\phi_{ij}=1, \lambda)\pi_j^{(0)}}.$$

M step:

$$\pi_j^{(1)}=\frac{\sum_{i=1}^{n}\Pr(\phi_{ij}=1)}{n}.$$

To update the estimate of recombination frequency at some or all intervals, we need to calculate it for each individual. An SDL may be treated as a missing marker when estimating the recombination frequency

https://doi.org/10.1017/S0016672305007779 Published online by Cambridge University Press

Table 1. *Loci showing segregation distortion in the two F_2 populations*

Population	Chromosome	Position (cM)	CI_{95}[a]	Nearest marker to SDL	θ_1[b]	θ_2[c]	θ_3[d]	LOD[e]	$T_{0.95}$[f]	$T_{0.99}$[g]	Selection types
TB	1	173	9 (169–178)	XNbp346	0·776	0·224	0·000	67·0**	7·24	11·34	Gametic
	3	9	27 (4–31)	R518	0·160	0·708	0·131	8·7*	6·01	8·89	Zygotic
	8	5	18 (1–19)	XNpb278	0·241	0·660	0·099	7·9*	5·94	9·85	Zygotic
		45	42 (22–64)	G2132	0·412	0·453	0·135	7·4*	5·94	9·85	Gametic
	9	36	35 (27–62)	XNpb103	0·034	0·588	0·378	17·6**	6·18	9·15	Gametic
CW	2	3	17 (2–19)	RM110	0·139	0·622	0·239	6·7*	5·90	9·34	Gametic
	3	154	38 (137–174)	RM135	0·126	0·657	0·217	9·8**	6·05	9·26	Zygotic
		199	23 (181–204)	RM148	0·167	0·658	0·175	8·0*	6·05	9·26	Gametic
	4	53	24 (38–62)	RM564	0·243	0·616	0·141	6·4*	6·17	9·29	Zygotic
		87	18 (77–95)	RM317	0·226	0·638	0·136	7·8*	6·17	9·29	Zygotic
	6	39	25 (29–54)	RM314	0·168	0·701	0·132	13·3**	7·12	11·18	Zygotic
	8	36	15 (34–49)	RM6863	0·213	0·270	0·517	27·4**	7·12	11·18	Zygotic
		120	32 (99–131)	RM256	0·140	0·506	0·354	7·4*	7·12	11·18	Zygotic
		146	29 (134–163)	RM281	0·129	0·479	0·392	11·0**	5·95	9·31	Gametic
	9	34	21 (25–46)	RM219	0·141	0·618	0·241	6·5*	5·88	9·26	Gametic
	10	107	31 (90–121)	RM228	0·221	0·676	0·103	13·4**	6·13	10·24	Zygotic
	12	130	18 (113–131)	RM463	0·412	0·462	0·126	13·1**	5·88	11·37	Gametic

[a] CI_{95}, the mean width of the 95% confidence interval (in cM).
[b] θ_1, frequency of AA genotype in the populations.
[c] θ_2, frequency of Aa genotype in the populations.
[d] θ_3, frequency of aa genotype in the populations.
[e] LOD, likelihood ratio; * and ** indicate significant difference at the 0·05 and 0·01 probability level, respectively.
[f] $T_{0.95}$, critical values based on 1000 permutations of the original data with $\alpha = 0{\cdot}05$.
[g] $T_{0.99}$, critical values based on 1000 permutations of the original data with $\alpha = 0{\cdot}01$.

between SDL and the adjacent markers:

$$P(x_k x_{k+1}/z_1, \ldots, z_m) = \frac{p(x_k x_{k+1}) p(z_1, \ldots, z_m/x_k, x_{k+1})}{\sum_{x_k, x_{k+1}} p(x_k x_{k+1}) p(z_1, \ldots, z_m/x_k, x_{k+1})}$$

which can be expressed in a 3×3 matrix form A_k. Defining a_{icd} as the cth row and dth column element of A_k for individual i, the recombination frequency for the kth interval can be updated as

$$r_k^{(1)} = \frac{1}{2n} \sum_{i=1}^{n} \left[a_{i12} + 2a_{i13} + a_{i21} + \frac{r_k^{(0)2}}{(1 - r_k^{(0)})^2 + r_k^{(0)2}} 2a_{i22} + a_{i23} + 2a_{i31} + a_{i32} \right].$$

E and M steps are iterated until convergence. We can now test the null hypothesis that there is no segregation distortion for the particular location λ. The null hypothesis is formulated as H_0: $\pi_1 = 0{\cdot}25$, $\pi_2 = 0{\cdot}5$ and $\pi_3 = 0{\cdot}25$, which can be tested using the likelihood-ratio (LR) test statistic $B = -2(\log L(0{\cdot}25, 0{\cdot}5, 0{\cdot}25, \lambda) - \log L(\pi_1, \pi_2, \pi_3, \lambda))$. Though under the null model, B is approximately distributed as a chi-square variable with 2 degrees of freedom. The LOD (log of the odds ratio) threshold significance level was determined by computing 1000 permutations (Churchill & Doerge, 1994) and the confidence intervals of a SDL and QTL location were determined by the bootstrapping method (Visscher *et al.*, 1996).

The maximum likelihood estimate of the position of the SDL can be obtained by examining the likelihood-ratio profile along the chromosome as is commonly done in interval mapping of QTL.

(iv) *Fertility observation and QTL analysis*

F_2 spikelet fertility was calculated as the percentage of normal spikelets over total spikelets per panicle for each of the F_2 plants. Six panicles for each plant of the two populations were observed and averaged for the spikelet fertility score.

To normalize the variance, the spikelet sterility percentage of each individual was transformed to arcsine [= arcsine $x^{1/2}$]. Composite interval mapping was performed to identify QTLs by using the software package QTL CARTOGRAPHER with forward-backward regression (Basten *et al.*, 1998). The additive and dominance effect and the percentage of variation explained by individual QTLs were estimated by following Zeng (1994). The experiment-wise LOD (log of the odds ratio) threshold significance level was determined by computing 1000 permutations (Churchill & Doerge, 1994), as implemented by QTL CARTOGRAPHER. These permutations can account for non-normality in marker distribution and trait

https://doi.org/10.1017/S0016672305007779 Published online by Cambridge University Press

values. The levels of significance for QTLs in this study were determined to be $P \leqslant 0{\cdot}05$: LOD 3·22 for the TB population and 2·03 for CW. For the designation of QTLs, we followed McCouch *et al*. (1997).

3. Results

(i) *Linkage map construction and SDL analysis*

With the TB and CW F_2 populations containing 96 and 157 individuals respectively, two frameworks of linkage map with 94 RFLP markers and 118 SSR markers respectively were constructed. The orders of the markers determined were consistent with those of the maps published by Harushima *et al*. (1998), Temnykh *et al*. (2000) and McCouch *et al*. (2002), though there was a gap on the short arm of chromosome 7 due to a limited number of markers in the CW population. The linkage distances between the markers were adjusted after considering the segregation distortions.

There should be 9024 and 18526 marker genotypes in the TB and CW populations, respectively. Of them, 179 and 524 were missing, accounting for 1·98% and 2·83% of the total marker genotypes. One dominant marker (RM5068, chromosome 8) was genotyped in the CW population. Single-marker chi-square tests showed that there were 24 and 47 markers which deviated from Mendelian segregation ratios in the TB and CW populations, respectively (data not shown), but the multiple test would result in higher false positives. Henceforth, the LR test was made in the genome scan to verify whether there is segregation distortion for the particular location in the two F_2 populations with missing marker genotypes.

The results of LR tests showed that five SDLs in the TB population were mapped on chromosomes 1, 3, 8 and 9, respectively, with two SDLs located on chromosome 8. Zygotic selection was detected near XNpb346 on the long arm of chromosome 3 and XNpb278 on the short arm of chromosome 8, with gametic selection at the other SDLs. The SDL on chromosome 1 showed strongest distortion, with the lowest $\theta_3 = 0$ and highest LOD = 67·0 (Table 1, Figs. 1 and 2). In the CW population, a total of 12 SDLs were detected on all the 12 chromosomes except for 1, 5, 7 and 11, with more than one SDL located on each of chromosomes 3, 4 and 8. Zygotic selection was detected on chromosomes 3, 4, 6, 8 and 10, and gametic selection on chromosomes 2, 3, 8, 9 and 12 (Table 1, Figs. 3 and 4) The SDL near RM6863 on chromosome 8 showed strongest distortion with lowest $\theta_2 = 0{\cdot}270$, highest $\theta_3 = 0{\cdot}517$ and highest LOD = 27·4. It is noteworthy that the SDL at SSR marker RM6863 on the short arm of chromosome 8 appeared to correspond to the one at RFLP marker XNpb278 in the TB population, since the linkage distance between the two markers was only 3·6 cM in the integrated map showing relative positions of newly developed SSR markers in relation to genetically mapped RFLP markers (McCouch *et al*., 2002). Furthermore, both the SDLs showed the same zygotic selection type in the two populations (Table 1, Figs. 2 and 4).

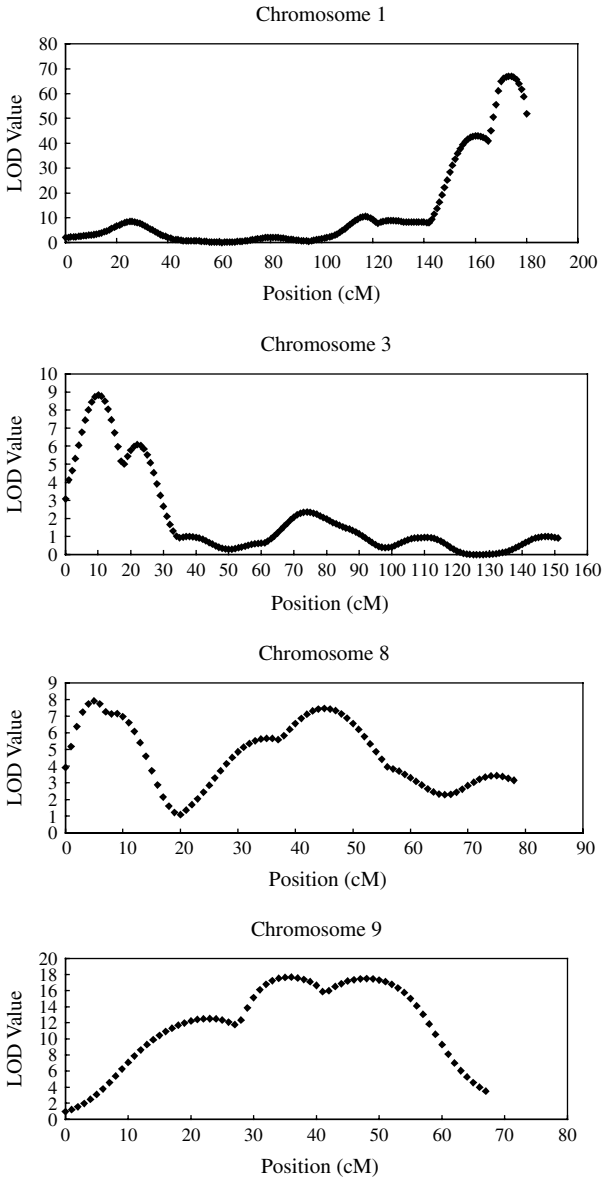

Fig. 1. SDLs in the Taichung65/Bhadua F_2 population. One SDL was detected on each of the chromosomes 1, 3 and 9, and two SDLs on chromosome 8.

(ii) *Phenotypic variation*

The spikelet fertilities and their segregations in the TB F_2 populations are shown in Fig. 5. The fertility of Taichung65 was normal while the fertilities of Bhadua and F_1 plants were around 50·4% and 70·0%, respectively. Fertility segregation was observed in the F_2 population with average of 69·0%, SD 24·0% and *cv*

https://doi.org/10.1017/S0016672305007779 Published online by Cambridge University Press

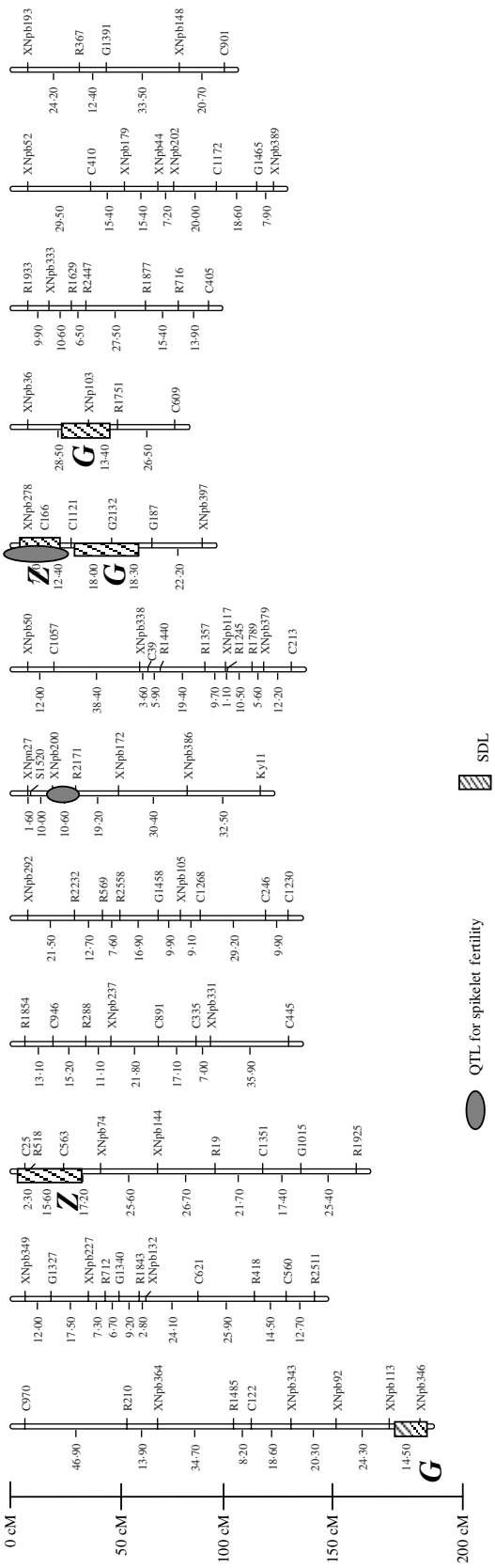

Fig. 2. Linkage map based on the F_2 (Taichung65/Bhadua) and QTLs for spikelet fertility. SDLs are marked with G or Z, denoting gametic or zygotic selection, respectively.

34·8%; some of the F_2 plants showed a very low spikelet fertility of less than 30·0%.

CPSLO17 showed normal spikelet fertility while W207-2 and F_1 plants showed low spikelet fertilities of 35·1% and 79·2%, respectively. Frequency distribution of spikelet fertilities in the F_2 population is shown in Fig. 6. The distribution was bimodal with one peak assumed at 90% and another one around 35%. This result suggested that the spikelet semi-sterility of W207-2 is controlled by major genes.

(iii) *QTL analysis for spikelet sterility*

Across the entire genome map of the TB population, two QTLs for spikelet sterility were detected on chromosomes 6 and 8, designated as *qSS-6a* and *qSS-8a*, respectively (Table 2, Fig. 2). Of these, *qSS-6a* was linked to the RFLP marker R2171 with LOD score of 3·46 and percentage of variance explained (PVE) of 18·9%. Another QTL, *qSS-8a* with LOD score of 3·21, was detected on chromosome 8, linked to the marker XNpb278, and accounted for 11·7% of the phenotypic variance of spikelet fertility.

In the CW population, three QTLs controlling spikelet fertility were identified in the entire genome map, of which *qSS-8b* near the marker RM6863 on chromosome 8 was detected with a relatively large LOD score of 40·31 and a PVE of 74·8% (Table 2, Fig. 4). Interestingly, *qSS-6a* and *qSS-6b* were located at the corresponding positions in the two populations; meanwhile *qSS-8a* and *qSS-8b* were also located at a similar position showing the same type of zygotic selection.

4. Discussion

A distorted segregation ratio is one of the common phenomena observed in both experimental and natural populations. The distortions are caused either by differential representation of SDL genotypes in gametes before fertilization or by viability differences of SDL genotypes after fertilization but before genotyping scoring. For codominant markers, it is possible to determine what type of selection (gametic versus zygotic) occurred at a locus by using two successive chi-square tests. For a single marker test, estimation of the position and effect are confounded, so errors in marker genotyping may cause systematic deviations from the expected segregation ratio; and if segregation distortion is caused by an SDL all markers in the vicinity of the SDL will be affected. Maximum likelihood methods using flanking markers were developed for mapping SDLs by Fu & Ritland (1994). But on the one hand more than one SDL per chromosome may be present, and on the other hand, markers may be dominant or partially informative; furthermore estimates of map distance may be biased

https://doi.org/10.1017/S0016672305007779 Published online by Cambridge University Press

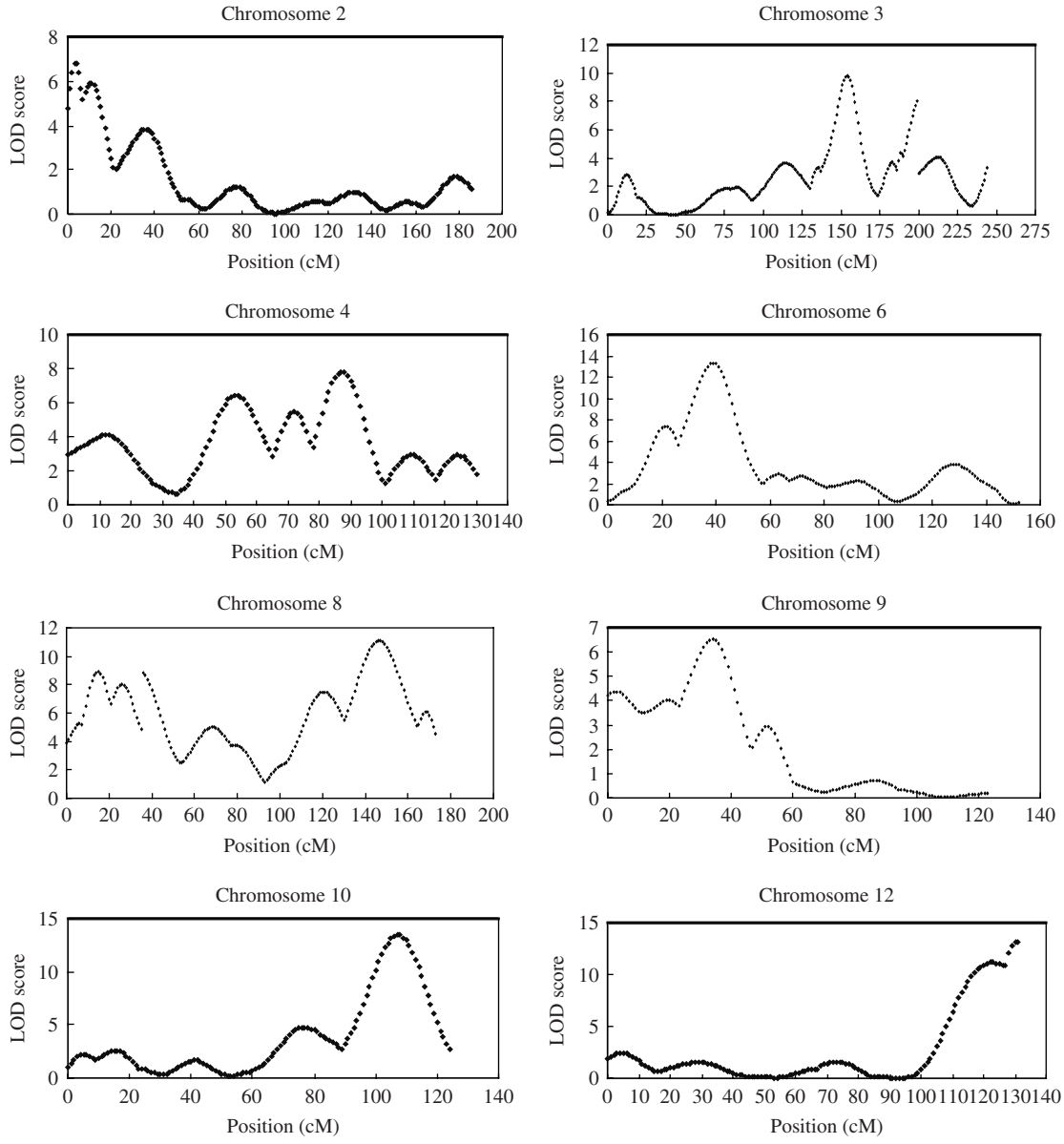

Fig. 3. SDLs in the CPSLO17/W207-2 F_2 population. One SDL was detected on each of chromosomes 2, 6, 9, 10 and 12, and more than one SDL on chromosomes 3, 4 and 8.

due to the effect of SDLs. In this paper, a multipoint method with dominant and missing markers was developed for mapping multiple SDLs using F_2 populations on the assumption that there is no interaction among the linked SDLs. In particular, the recombination frequencies between consecutive markers and the recombination fractions between SDL and the adjacent markers were re-estimated. The adjusted linkage map will be used for further QTL analysis. There is no doubt that our method will be biased when there is interaction among the multiple SDLs on the same chromosome. In that case, a Bayesian method is suggested (Vogl & Xu, 2000).

In a previous study based on chi-square analysis ($P<0{\cdot}01$), chromosomal regions associated with marker segregation distortion in rice were determined and compared with six F_2 populations of inter-subspecies crosses (Xu *et al.*, 1997). Another study was conducted to map reproductive barriers causing deviation from Mendelian segregation ratios in F_2 populations of three different *japonica–indica* crosses, and most of the barriers were mapped at different loci (Harushima *et al.*, 2002). In comparing the results in this study with previous studies, fewer SDLs were detected and most of those SDLs detected in this study corresponded to the ones reported in the previous studies. It is clear that there would be considerable numbers of false positive SDLs in the single-marker chi-square test used in the previous studies.

Through our approach we efficiently analysed the number, positions and effects of SDLs in *japonica–indica* and *japonica–japonica* F_2 populations. In

https://doi.org/10.1017/S0016672305007779 Published online by Cambridge University Press

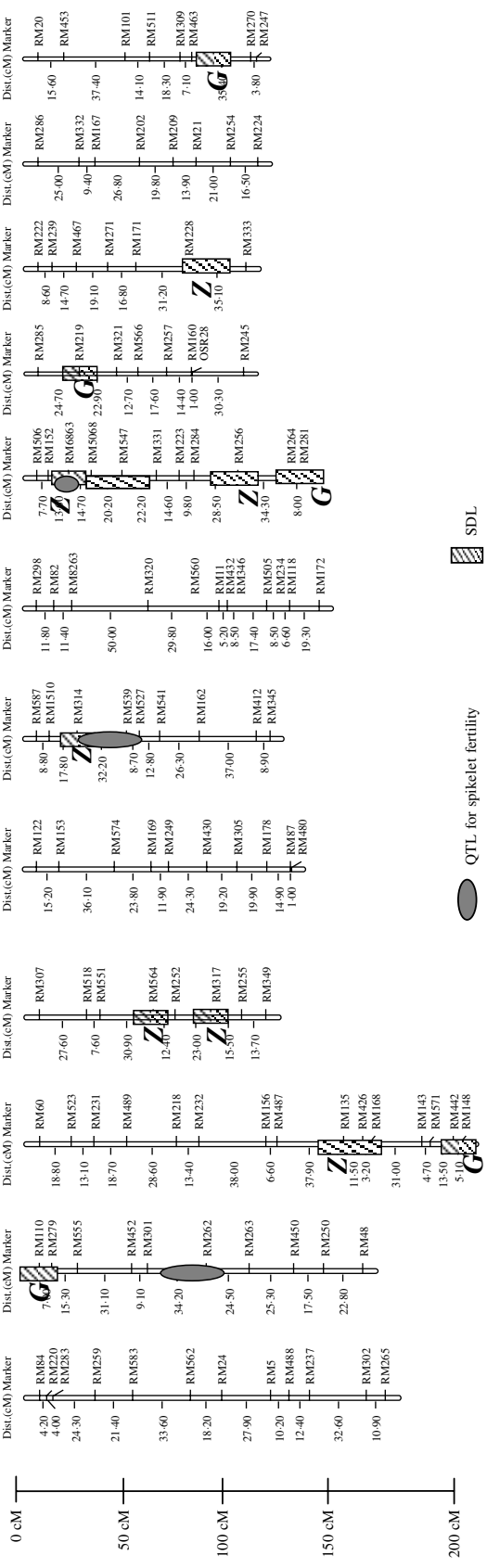

Fig. 4. Marker linkage map based on the F_2 (W207-2/CPSLO17) and QTLs for spikelet fertility. The notation is the same as in Fig. 2.

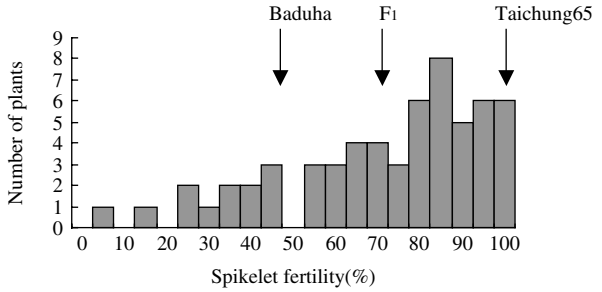

Fig. 5. Frequency distribution for spikelet fertility in the F_2 population of Taichung65/Bhadua.

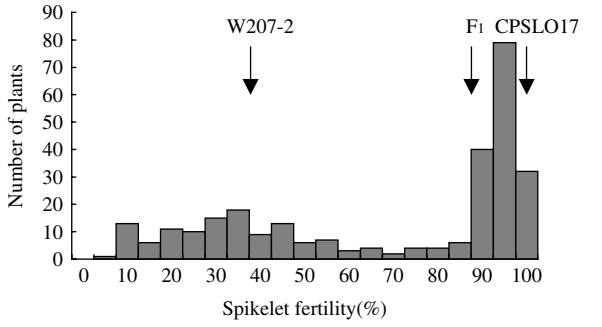

Fig. 6. Frequency distribution for spikelet infertility in the F_2 population of CPSLO17/W207-2.

addition to the loci for visible spikelet sterility in the inter-subspecies cross, i.e. the TB population, three other SDLs were found causing segregation distortion. To solve the problem of hybrid sterility, the present results indicate the need to restore distortion-neutral alleles on a whole genome basis by introducing wide-compatibility types of rice to active hybridization plans, and the molecular markers linked to these alleles in this study should be useful for the restoration.

As early as 1974, various genic sterilities were found in rice, many of which were reported in inter-subspecific hybrids in rice (Oka, 1974). The genetic control of distorted segregation has previously been studied in rice using morphological and isozyme markers. Our previous work has shown a set of genes for female gamete abortion in hybrids between distantly related varieties (Wan *et al.*, 1995, 1996*a*, *b*). Ikehashi & Araki (1986) were able to relate the hybrid sterility to an allelic interaction at the *S-5* locus. Following that work, similar allelic interactions were identified at other loci. According to the study, *indica* and *japonica* varieties carry $S\text{-}5^i$ and $S\text{-}5^j$ alleles, respectively, at the *S-5* locus on chromosome 6. Some cultivars, such as Ketan Nangka and Dular, have a neutral allele $S\text{-}5^n$. The $S\text{-}5^i/S\text{-}5^j$ genotype is semi-sterile due to partial abortion of female gametes carrying $S\text{-}5^j$, whereas both $S\text{-}5^n/S\text{-}5^i$ and $S\text{-}5^n/S\text{-}5^j$ genotypes are fertile. The $S\text{-}5^n$ allele is thus called a 'wide-compatibility gene' (WCG), and used to develop subspecific hybrids. Following that work,

https://doi.org/10.1017/S0016723005007779 Published online by Cambridge University Press

Table 2. *QTLs detected for spikelet fertility in F_2 populations*

Population	Loci	Chromosome	Nearest marker to putative QTL	LOD[a]	Position (cM)	CI_{95}[b]	a[c]	d[d]	PVE[e] (%)
TB	*qSS-6a*	6	R2171	3·46	22·8	7·6 (19·5–27·1)	−0·206	0·116	0·189
	qSS-8a	8	XNpb278	3·21	3·5	20·1 (0–20·1)	−0·057	0·258	0·117
CW	*qSS-2*	2	RM263	2·22	95·9	35·9 (69·8–105·7)	0·063	0·192	0·063
	qSS-6b	6	RM539	2·15	50·7	16·9 (44·3–61·2)	−0·014	0·155	0·036
	qSS-8b	8	RM6863	40·31	20·1	2·7 (18·5–21·2)	−0·369	0·371	0·748

[a] LOD, likelihood ratio.
[b] CI_{95}, mean width of the 95% confidence interval (in cM).
[c] a, additive effects of Bhadua and W207-2 alleles.
[d] d, dominant effect.
[e] PVE, percentage of phenotypic variance explained.

similar allelic interactions were identified at other loci – *S-7*, *S-8*, *S-9, S-15*, *S-16*, and *S-17* – together with respective WCGs, although they are less frequent than *S-5* but found to cause female gamete abortion in hybrids between distantly related varieties (Wan & Ikehashi, 1996*a*). In rice, 28 sterility gene loci (*S*) in 10 chromosomal regions were previously identified (summary by Nagato *et al*., 1998).

Compared with the previous studies, *qSS-6a* and *qSS-6b* might harbour a gene showing allelic interaction for spikelet sterility (*S-5*) (Wan *et al*., 1995; summary by Nagato *et al*., 1998), while QTLs of *qSS-8a* and *qSS-8b* were newly detected at the similar chromosome region in the two different populations. Interestingly, zygotic selection was detected at both *qSS-8a* and *qSS-8b*. Therefore this chromosome region appeared to be associated with the rice's spikelet sterility and zygotic selection, or some genes related to spikelet sterility and zygotic selection might cluster in this chromosome region. *qss-2* was newly detected in the CW population but not in the TB population. These loci for segregation distortion indicate that the application of Mendelian segregation ratios may not always be justified in *indica–japonica* hybrids. The results obtained can also be used to introduce WCG-type genes into *indica–japonica* hybrids by the use of new markers. To understand the underlying mechanisms that cause segregation distortion, it would be interesting to evaluate the effect of these factors systematically in different genetic backgrounds and environments with near-isogenic lines (NILs) containing individual SDLs and QTLs related to spikelet sterility.

The source codes for a C++ program, with which the above calculations can be performed, are available from Chengsong Zhu (cszhu@sina.com) or Chunming Wang (wangchm@njau.edu.cn).

We are grateful to Professor W. G. Hill and two anonymous reviewers for their thoughtful criticisms, comments and suggestions on early versions of the manuscript. This major revision of the manuscript has been greatly improved both in scientific merit and presentation by incorporating their suggestions. The research was sponsored by National high-tech project (No. 2003AA207020, 2001AA241024), China to C.M.W. and J.M.W.

References

Basten, C. J., Weir, B. S. & Zeng, Z. B. (1998). QTL CARTOGRAPHER: a reference manual and tutorial for QTL mapping. Raleigh, NC: Department of Statistics, North Carolina State University.

Chen, X., Temnykh, S., Xu, Y., Cho, Y. G. & McCouch, S. R. (1997). Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theoretical Applied Genetics* **95**, 553–567.

Cheng, R., Saito, A. & Ukai Y. (1996). Estimation of the position and effect of a lethal factor locus on a molecular marker linkage map. *Theoretical Applied Genetics* **93**, 494–502.

Churchill G. A. & Doerge R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Fu, Y. B. & Ritland K. (1994) On estimating the linkage of marker genes to viability genes controlling inbreeding depression. *Theoretical Applied Genetics* **88**, 925–932.

Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S. Y., Antonio, B. A., Parco, A., Kajiya, H., Huang, N., Yamamoto, K., Nagamura, Y., Kurata, N., Khush, G. S. & Sasaki, T. (1998). A high-density rice genetic linkage map with 2275 markers using a single F_2 population. *Genetics* **148**, 479–494.

Harushima, Y., Nakagahra, M., Yano, M., Sasaki, T. & Kurata, N. (2001). A genome-wide survey of reproductive barriers in an intra-specific hybrid. *Genetics* **159**, 883–892.

Harushima, Y., Nakagahra, M., Yano, M., Sasaki, T. & Kurata, N. (2002). Diverse variation of reproductive barriers in three intraspecific rice crosses. *Genetics* **160**, 313–322.

Ikehashi, H. & Araki, H. (1986). Genetics of F_1 sterility in remote crosses of rice. In *Rice Genetics*, pp. 119–130. PO Box 933, Manila, Philippines: IRRI.

Jiang, C. & Zeng, Z. (1997). Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. *Genetica* **101**, 47–58.

Lander, E., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. & Newburg, L. (1987).

MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and nature populations. *Genomics* **1**, 174–181.

Lorieux, M., Perrier, X., Goffinet, B., Lanaud, C. & Gonzalez de Leon, D. (1995). Maximum likelihood models for mapping genetic markers showing segregation distortion. 2. F2-population. *Theoretical Applied Genetics* **90**, 81–89.

McCouch, S. R., Cho, Y. G., Yano, M., Paul, E., Blinstrub, M., Morishima, H. & Kinoshita, T. (1997). Report on QTL nomenclature. *Rice Genetics Newsletter* **14**, 11–13.

McCouch, S. R., Teytelman, L., Xu, Y., Lobos, K. B., Clare, K., Walton, M., Fu, B., Maghirang, R., Li, Z., Xing, Y. Z., Zhang, Q., Kono, I., Yano, M., Fjellstrom, R., Declerck, G., Schneider, D., Cartinhour, S., Ware, D. & Stein, L. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Research* **9**, 199–207.

Nagato, Y., Yoshimura, A. & Conveners (1998). Report of the committee on gene symbolization, nomenclature and linkage groups. *Rice Genetics Newsletter* **15**, 34–74.

Oka, H. I. (1974). Analysis of genes controlling F_1 sterility in rice by the use of isogenic lines. *Genetics* **77**, 521–534.

Rogers, S. O. & Bendich, A. (1988). Extraction of DNA from plant tissues. *Plant Molecular Biology Manual* **A6**, 1–10.

Sanguinetti, C. J. & Simpson, N. E. (1994). Rapid silver staining and recover of PCR products separated on polyacrylamide gels. *Biotechology* **17**, 915–919.

Temnykh, S., Park, W. D., Ayres, N., Cartinhour, S., Hauck, N., Lipovich, L., Cho, Y. G., Ishii, T. & McCouch, S. R. (2000). Mapping and genome organization of microsatellite sequence in rice (*Oryza sativa* L.). *Theoretical Applied Genetics* **100**, 697–712.

Visscher, P. M., Thompson, P. & Haley, C. S. (1996). Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**, 1013–1020

Vogl, C. & Xu, S. (2000). Multipoint mapping of viability and segregation distorting loci using molecular markers. *Genetics* **155**, 1439–1447.

Wan, J. & Ikehashi, H. (1995). Identification of a new locus *S-16* causing hybrid sterility in native rice varieties (*Oryza sativa* L.) from Tai-hu Lake region and Yunnan province, China. *Breeding Science* **45**, 161–170.

Wan, J. & Ikehashi, H. (1996*a*). List of hybrid sterility gene loci (HSGLi) in cultivated rice (*Oryza sativa* L.). *Rice Genetics Newsletter* **13**, 110–113.

Wan, J., Yamaguchi, Y., Kato, H. & Ikehashi, H. (1996*b*). Two new loci for hybrid sterility in cultivated rice (*Oryza sativa* L.). *Theoretical Applied Genetics* **92**, 183–190.

Xu, Y., Zhu, L., Xiao, J., Huang, N. & McCouch, S. R. (1997). Chromosomal regions associated with segregation distortion of molecular markers in F_2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). *Molecular General Genetics* **253**, 535–545.

Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics* **136**, 1457–1468.