Genet. Res., Camb. (2001), **77**, *pp.* 95–106. *With* 1 *figure.* Printed in the United Kingdom © 2001 Cambridge University Press

Multi-trait QTL mapping in barley using multivariate regression

C. A. HACKETT[1]*, R. C. MEYER[2] AND W. T. B. THOMAS[2]

[1] *Biomathematics and Statistics Scotland, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK*
[2] *Scottish Crop Research Institute, Dundee DD2 5DA, Scotland, UK*

(*Received 20 December 1999 and in revised form 19 June and 21 August 2000*)

Summary

Many studies of QTL locations record several different traits on the same population, but most analyses look at this information on a trait-by-trait basis. In this paper we show how the regression approach to QTL mapping of Haley & Knott (1992) may be extended to a multi-trait analysis via multivariate regression, easily programmed in statistical packages. A procedure for identifying QTL locations using forward selection and bootstrapping is proposed. The method is applied to examine the locations for QTLs for six yield characters (the number of fertile stems, the grain number of the main stem, the main stem grain weight, the single plant yield, the plot yield and the thousand grain weight) in a doubled haploid population of spring barley. Several chromosomal locations with effects on more than one trait are found. The method is also suitable for examining a single trait measured in different years or environments, and is used here to examine data on heading date, a highly heritable trait, and plot yield, a trait with moderate heritability and showing QTL–environment interactions.

1. Introduction

Many genome-wide QTL studies have been published in which a number of traits have been measured on the same population. The genetic analysis to place QTLs affecting these traits on a linkage map is usually conducted on a trait-by-trait basis. However, such a trait-by-trait analysis may overlook much information of use to the plant breeder. For example, a trait-by-trait analysis often shows it is most likely that QTLs for the different traits are in very similar locations, but with different parents contributing favourable alleles (e.g. Thomas *et al*., 1995, 1996). The dilemma facing geneticists and breeders is whether such results indicate linkage or pleiotropy. If the former, then marker-assisted selection to break an unfavourable linkage would be worthwhile and the desirable population size could be formulated from the degree of linkage. If the latter, then much time and effort would be wasted trying to break an apparent linkage. A simultaneous analysis of all of the quantitative trait data from a cross is necessary to resolve this dilemma. Another example is when several traits, all known to relate to a complex trait of interest, are measured (e.g. hot water extract, grain nitrogen content, milling energy contributing to barley malting quality). A simultaneous analysis of these traits may give particular insight into the physiological processes occurring and may identify suitable targets for breeding and/or future research.

However, efficient statistical methods for combined trait analyses are at an early stage of development. Ronin *et al*. (1995) used mixtures of bivariate normal distributions to look at the joint likelihood of two correlated traits, affected by a QTL linked to a molecular marker, and Korol *et al*. (1995) extended this to full interval mapping for two correlated traits. They found, by simulation, that the power of QTL detection could be increased by taking into account correlations between traits, even if the QTL position under consideration affected only one of the traits. Jiang & Zeng (1995) have extended composite interval mapping to include the effects of a locus on several traits simultaneously. Their approach gives a hypothesis test for pleiotropy versus close linkage. By simulation, they demonstrated an increase in power of QTL detection and precision of parameter estimation.

* Corresponding author. Tel; +44 (0) 1382 562731. Fax; +44 (0) 1382 562426. e-mail; chacke@scri.sari.ac.uk

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press

Weller *et al.* (1996) derived principal components based on the values of three milk production traits (milk yield, fat yield and protein yield) and related these to a single locus associated with milk production in dairy cattle. Mangin *et al.* (1998) has derived theoretical results for a test, based on the principal components of the traits, to test the hypothesis of presence versus absence of a pleiotropic QTL.

The methods mentioned above for analysing several traits simultaneously by interval mapping all require complicated programming, and software is only available at present for the method of Jiang & Zeng (1995), in program JZmapqtl of QTL Cartographer (Basten *et al.*, 1994, 1999). In this report an alternative method for multi-trait mapping is presented. This is based on the regression mapping approach of Haley & Knott (1992), and can be programmed in standard statistical packages such as Genstat. It is applied to analyse data on several characters relating to yield in a doubled haploid population derived from a spring barley cross. The method is equally applicable to analyse traits measured in more than one environment or year, and is used here to analyse heading date, a highly heritable trait, and plot yield, with lower heritability.

2. Methodology

(i) *Genetic correlation between traits*

Correlation between traits can arise due to genetic or environmental effects. If a population of genetically identical plants is grown in a field trial, then some traits are likely to be correlated and this must be a purely environmental correlation, perhaps due to varying conditions in the field. If the population is not genetically identical then observed correlations may be a mixture of genetic and environmental effects. Alternatively, genetic and environmental correlations of opposite signs may cancel each other to some extent. Just as the variance of a trait is the sum of genetic effects and environmental effects, so is the covariance between two traits and it is of interest to separate these effects.

We will develop our model for a double haploid (DH) population derived from a cross between two inbred, homozygous parents. Consider two traits, X and Y, affected by QTLs Q_X and Q_Y respectively. Assume Q_X and Q_Y lie on the same chromosome, with a recombination fraction r between them and let the two parents have genotypes $\mathrm{Q_XQ_XQ_YQ_Y}$ and $\mathrm{q_Xq_Xq_Yq_Y}$ respectively. Table 1 shows the four possible offspring genotypes at the two loci Q_X and Q_Y, their frequencies (f) and the corresponding expected trait values.

The genetic covariance $\sigma_{\mathrm{G(XY)}}$ may be calculated as:

$$\sigma_{G(XY)} = \sum_{i=1}^{4} f_i(X_i - \bar{X}_i)(Y_i - \bar{Y}_i)$$

$$(\text{where } i = Q_XQ_XQ_YQ_Y, Q_XQ_Xq_Yq_Y, q_Xq_XQ_YQ_Y, q_Xq_Xq_Yq_Y)$$

$$= \frac{(1-r)}{2}a_Xa_Y - \frac{r}{2}a_Xa_Y - \frac{r}{2}a_Xa_Y + \frac{(1-r)}{2}a_X a_Y$$

$$= (1-2r)a_X a_Y.$$

The genetic covariance is largest if $r = 0$ and tends to zero as r approaches 0·5. If J chromosomes have a pair of linked QTLs for X and Y the genetic correlation becomes $\sum_{j=1}^{J}(1-2r_j)a_{X_j}a_{Y_j}$. Then the phenotypic covariance becomes $\sum_{j=1}^{J}(1-2r_j)a_{X_j}a_{Y_j} + \mathrm{cov}(e_X, e_Y)$, where the second term represents the environmental covariance. The expression will be complicated further if there are chromosomes where more than one QTL affects one or more of the traits under investigation.

(ii) *Multi-trait analysis*

We will describe the regression mapping approach of Haley & Knott (1992) for a single variable, review

Table 1. *The four possible offspring genotypes at two loci* $\mathrm{Q_X}$ *and* $\mathrm{Q_Y}$ *affecting traits X and Y respectively for a doubled haploid population generated from parents with genotypes* $Q_XQ_XQ_YQ_Y$ *and* $q_Xq_Xq_Yq_Y$, *and the corresponding expected trait values.* m_X, a_X, m_Y *and* a_Y *are the mid-parent values and additive effects for traits X and Y*

Genotype	Frequency (f)	Expected value for X	Expected value for Y
$\mathrm{Q_XQ_XQ_YQ_Y}$	$(1-r)/2$	$m_X + a_X$	$m_Y + a_Y$
$\mathrm{Q_XQ_Xq_Yq_Y}$	$r/2$	$m_X + a_X$	$m_Y - a_Y$
$\mathrm{q_Xq_XQ_YQ_Y}$	$r/2$	$m_X - a_X$	$m_Y + a_Y$
$\mathrm{q_Xq_Xq_Yq_Y}$	$(1-r)/2$	$m_X - a_X$	$m_Y - a_Y$
Mean		m_X	m_Y

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press

Table 2. *Calculation of the expected trait value for each marker genotype in a DH population*

Genotype	Frequency	Proportion QQ (m_X+a_X)	Proportion qq (m_X-a_X)	Expected trait value	Explanatory variable l
AABB	$(1-r)/2$	$(1-r_A)(1-r_B)/2$	$r_A r_B/2$	$m_X+a_X(1-r_A-r_B)/(1-r)$	$(1-r_A-r_B)/(1-r)$
AAbb	$r/2$	$(1-r_A)r_B/2$	$r_A(1-r_B)/2$	$m_X+a_X(r_B-r_A)/r$	$(r_B-r_A)/r$
aaBB	$r/2$	$r_A(1-r_B)/2$	$(1-r_A)r_B/2$	$m_X-a_X(r_B-r_A)/r$	$-(r_B-r_A)/r$
aabb	$(1-r)/2$	$r_A r_B/2$	$(1-r_A)(1-r_B)/2$	$m_X-a_X(1-r_A-r_B)/(1-r)$	$-(1-r_A-r_B)/(1-r)$

some results on multivariate regression and combine the two to obtain a model for multi-trait analysis.

Haley–Knott regression

Consider a DH population with marker loci A and B flanking a QTL, Q. Let the recombination frequencies between A and Q, Q and B and A and B be r_A, r_B and r respectively. Table 2 shows the frequency of each of the four possible marker genotypes, the proportion of each QTL genotype associated with that marker genotype and the expected trait value associated with that marker genotype for a single trait, X. The expected trait value can always be expressed as the overall mean, m_X, plus the QTL effect, a_X, multiplied by a function of the QTL position. This function, l, is given in the final column of Table 2. Therefore, if we knew the position of the QTL, we could calculate l and then estimate the QTL parameters m_X and a_X by regression of the trait values on l. We do not know the QTL position, so, following Haley & Knott (1992), we try a series of positions along the chromosome.

Regression mapping, therefore, involves calculating the explanatory variable l at each position along the chromosome, regressing the trait value on this explanatory variable to calculate m_X, a_X and the variance ratio for the significance of the regression, and identifying the position with the largest variance ratio for the regression as the most likely position for a QTL. The same approach may be used if there is more than one QTL affecting a single trait on the chromosome, provided that QTLs are not in adjacent intervals.

This method for QTL mapping may be generalized to two or more traits simultaneously, using multivariate regression.

Multivariate regression

(a) *Model estimation*. The general multivariate regression model is

$$\mathbf{Y}=\mathbf{XB}+\mathbf{E},$$

where $\mathbf{Y}$ is an $(n\times p)$ matrix of p traits observed on n individuals, $\mathbf{X}$ is a matrix of explanatory variables, $\mathbf{B}$ is a matrix of regression coefficients to be estimated (in our context this corresponds to the values of m_X and a_X for each trait), and $\mathbf{E}$ is an $(n\times p)$ matrix of random deviations from the mean. In our context the rows of $\mathbf{E}$ correspond to different individuals and so are independent of each other. However, the columns of $\mathbf{E}$ correspond to the errors on different traits and these may be correlated, due to environmental correlations or to the effects of QTLs on other chromosomes. The covariance matrix corresponding to $\mathbf{E}$ is Σ. We will assume that $\mathbf{E}$ is multivariate normal $N(0,\Sigma)$. In this case we can estimate $\mathbf{B}$ and Σ by

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press

maximum likelihood estimation, with equations very similar to those for the simple linear regression model (Mardia *et al.*, 1979, p. 158):

$$\hat{\mathbf{B}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y},$$

where $\mathbf{X}^{\mathrm{T}}$ denotes the transpose of $\mathbf{X}$ and

$$\hat{\Sigma} = \frac{1}{n}(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})^{T}(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}) = \frac{1}{n}\hat{\mathbf{E}}^{\mathrm{T}}\hat{\mathbf{E}}.$$

(*b*) *Hypothesis testing*. Let $\bar{\mathbf{y}}$ be the $(p \times 1)$ vector of trait means, and let $\mathbf{H} = (\hat{\mathbf{Y}}^{\mathrm{T}}\hat{\mathbf{Y}} - n\bar{\mathbf{y}}\bar{\mathbf{y}}^{\mathrm{T}})$. Then we can write the MANOVA table as

Source	*d.f.*	*SSP matrix*
Multivariate regression	q	$\mathbf{H}$
Residual	$n-q-1$	$\hat{\mathbf{E}}^{\mathrm{T}}\hat{\mathbf{E}}$
Total (corrected)	$n-1$	$\mathbf{Y}^{\mathrm{T}}\mathbf{Y}-n\bar{\mathbf{y}}\bar{\mathbf{y}}^{\mathrm{T}}$

where q is the number of explanatory variables. Tests of significance are based on the matrix product $(\hat{\mathbf{E}}^{\mathrm{T}}\hat{\mathbf{E}})^{-1}\mathbf{H}$ in a similar way to simple linear regression. This is a $(p \times p)$ matrix and most tests are based on its eigenvalues $\theta_1, \theta_2, \ldots, \theta_{\mathrm{p}}$.

One particularly useful test statistic is Wilks' $\Lambda = |\hat{\mathbf{E}}^{\mathrm{T}}\hat{\mathbf{E}}| \,/\, |\hat{\mathbf{E}}^{\mathrm{T}}\hat{\mathbf{E}}+\mathbf{H}|$, derived as a likelihood ratio test of $\mathbf{B} = \mathbf{0}$. This can be written in terms of the eigenvalues $\theta_1, \theta_2, \ldots, \theta_{\mathrm{p}}$ as:

$$\Lambda = \prod_{i=1}^{p} \frac{1}{1+\theta_i}$$

(Chatfield & Collins, 1980, p. 147) and it can be transformed to an approximate F distribution, with pq degrees of freedom for the numerator. Rao (1973) gives details of the transformation and equations for the denominator degrees of freedom. If either p or q are less than or equal to 2 the transformation is exact. Procedures for fitting a multivariate regression model are found in several standard statistical packages. Here multivariate regressions were fitted in Genstat (Genstat 5 Committee, 1995), using the procedure FITMULTIVARIATE.

(*c*) *Model selection*. Selection of explanatory variables from a large set has received much attention for univariate regression, but less for multivariate regression. Bedrick & Tsai (1994) conducted a simulation study of different criteria for model selection in multivariate regression, and found that the commonly used Akaike information criterion *AIC* (Akaike, 1973) is prone to overfit the data. They proposed an adjusted criterion, the minimum AIC_C, which avoids this problem. They define the AIC_C to be:

$$AIC_C = n \log|\hat{\sum}| + Dp(n+q),$$

where $D = n/(n-(q+p+1))$. The addition of an extra explanatory variable will increase q by 1, and so increase the second term of the AIC_C. Only if this increase is outweighed by the decrease in the first term will the explanatory variable be included. We will use this criterion to compare different subsets of explanatory variables. The AIC_C does not indicate the significance of each of the selected explanatory variables: the F test referred to in the previous section is used for this.

Multi-trait regression mapping

If we think a single QTL may affect p (> 1) traits, we can represent this as a multivariate regression model by generalizing the Haley–Knott model:

$$\begin{bmatrix} Y_{11} & \cdots & Y_{p1} \\ Y_{12} & \cdots & Y_{p2} \\ \vdots & \cdots & \vdots \\ Y_{1n} & \cdots & Y_{pn} \end{bmatrix} = \begin{bmatrix} 1 & l_1(d) \\ 1 & l_2(d) \\ \vdots & \vdots \\ 1 & l_n(d) \end{bmatrix} \times \begin{bmatrix} m_1 & \cdots & m_p \\ a_1 & \cdots & a_p \end{bmatrix} + \begin{bmatrix} E_{11} & \cdots & E_{p1} \\ E_{12} & \cdots & E_{p2} \\ \vdots & \cdots & \vdots \\ E_{1n} & \cdots & E_{pn} \end{bmatrix}$$

where l is the explanatory variable for regression mapping, a function of the distance d along the chromosome and the genotypes of the markers flanking position d. The position with the lowest value of the AIC_C should indicate the position of the QTL.

However, it may be that there are two or more QTLs affecting the traits on this chromosome. In this case we look for the set of positions $\{d_1, d_2, \ldots\}$ that minimize the AIC_C statistic, and see which effects are significant. For two traits and two QTLs, with one of the QTLs affecting each trait, we would expect to see some effects not significantly different from zero, for example:

$$\begin{bmatrix} Y_{11} & Y_{21} \\ Y_{12} & Y_{22} \\ \vdots & \vdots \\ Y_{1n} & Y_{2n} \end{bmatrix} = \begin{bmatrix} 1 & l_1(d_1) & l_1(d_2) \\ 1 & l_2(d_1) & l_2(d_2) \\ \vdots & \vdots & \vdots \\ 1 & l_n(d_1) & l_n(d_2) \end{bmatrix} \times \begin{bmatrix} m_1 & m_2 \\ a_1 & 0 \\ 0 & a_2 \end{bmatrix} + \begin{bmatrix} E_{11} & E_{21} \\ E_{12} & E_{22} \\ \vdots & \vdots \\ E_{1n} & E_{2n} \end{bmatrix}$$

The joint distribution of the estimated QTL effects is multivariate normal (Mardia *et al.*, 1979, p. 160) and the significance of each effect may be assessed by a t-test.

Stepwise selection of QTL position

Likely QTL locations can be identified by a stepwise analysis. We can calculate a series of explanatory variables corresponding to points at regular intervals along each chromosome according to the equation in the last column of Table 2, and use a stepwise

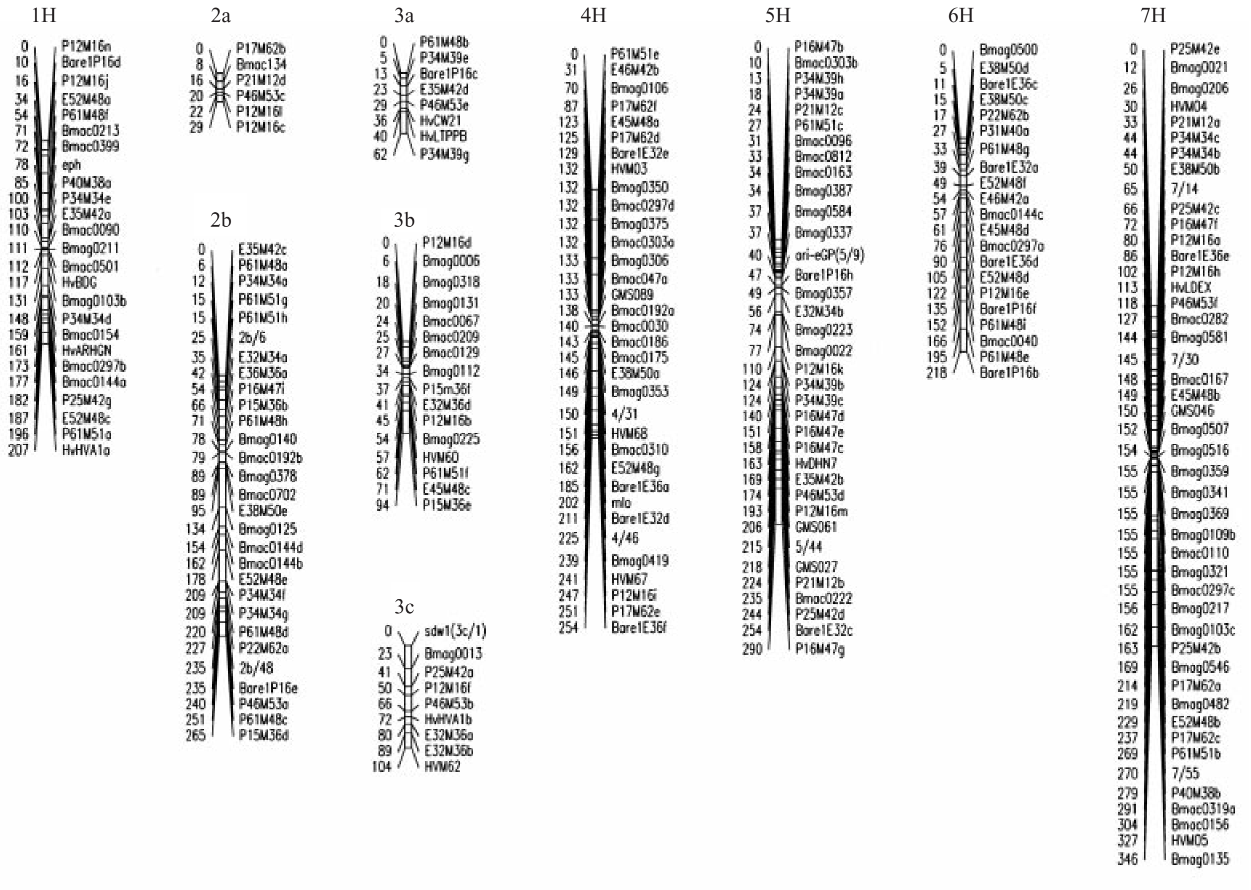

Fig. 1. Linkage map of Derkado × B83-12/21/5. The positions selected as associated with yield characters are also indicated (also see Table 5). If closely linked positions were selected for different years, the positions on the map are for the 1994 trait data.

approach, with forward selection, to identify the subset of explanatory variables for which the AIC_C is minimized. Explanatory variables were calculated at 5 cM intervals for this analysis (i.e. we are evaluating for the best QTL location on a 5 cM grid for each chromosome). A smaller grid was found to give explanatory variables which were too highly correlated to be useful for a stepwise analysis. The notation l(i/j) will be used to denote the jth explanatory variable on chromosome i, i.e. a test for a QTL on chromosome i at a position $5(j-1)$ cM from the first marker. Note that l(i/1) is at 0 cM, the location of the first marker. Derkado was regarded as the parent contributing the upper-case alleles in the notation of Table 2, and B83-12/21/5 as the parent contributing the lower-case alleles, so the estimated QTL effect a_X represents (Derkado allele − B83-12/21/5 allele)/2.

A difficulty in all QTL analyses is the question of multiple QTLs on a chromosome. A ghost QTL (Martinez & Curnow, 1992) may be detected in place of two genuine, linked QTLs. The following strategy was used to investigate whether the fit of the model was improved by a small shift in the position of a selected explanatory variable, or by replacement of a single explanatory variable by two linked variables. For each selected explanatory variable l(i/j) in turn, the effect on the AIC_C of substituting l(i/j) by l(i/j$\pm$1), l(i/j$\pm$2), l(i/j$\pm$3), l(i/j$\pm$4) and l(i/j$\pm$5) (i.e. explanatory variables representing QTL locations up to 25 cM to each side of the original position) was investigated, keeping the other selected positions unchanged. The effect on the AIC_C of including each of these 11 positions in combination with every other position on the chromosome was also investigated.

Bootstrap analysis to test for influential individuals

A preliminary exploration of the above modelling strategy with simulated data (not shown here) indicated that occasionally the significance of an explanatory variable in the model was due to the influence of a small number of individuals. To ensure that the final model was not affected by the inclusion or omission of a few individuals, a bootstrapping approach was used. Five hundred bootstrap samples were drawn from the dataset, the selected explanatory variable were refitted and the significance of each using Rao's F test was recorded. Only the explanatory

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press

Table 3. *Outline of MANOVA table for calculation of genetic and environmental variance and covariance*

Source	d.f.	SSP	Mean SSP	Expected mean SSP
Replicate	1	SSP_R	SSP_R	
Genotype	152	SSP_G	$SSP_G/152$	$2\Sigma^2_G+\Sigma^2_E$
Residual	152	SSP_E	$SSP_E/152$	Σ^2_E
Total	305	SSP_T		

variables that were significant in at least 95% of the bootstrap resamples were retained in the model.

(iii) *Experimental data for the mapping study*

The mapping population consisted of 156 lines from an F1 doubled haploid population, derived from a spring barley cross between Derkado (which carries the *sdw1* dwarfing gene and the *mlo11* powdery mildew resistance gene) and the SCRI breeding line B83-12/21/5 (which carries the Golden Promise dwarfing gene *ari-eGP*). Three lines were subsequently excluded from this study due to extreme outliers among some trait values. The genotyping of the population for the *sdw1* and *mlo11* genes and for S-SAP, AFLP and SSR markers are described in detail in Thomas *et al.* (1998). A further 87 SSRs comprising four GMS markers (Struss & Plieske, 1998), eight markers derived from database sequences, an HVM marker (Saghai-Maroof *et al.*, 1994) and 74 SCRI markers were scored on the population and the maps were reconstructed using JoinMap 2.0 (Stam & Van Ooijen, 1995). At a LOD score of 5, ten linkage groups were obtained. Some of the markers used here have been mapped in other populations, and this information was used to identify linkage groups. Markers from chromosomes 2H and 3H formed several groups, which are indicated as 2a, 2b and 3a, 3b and 3c. The resulting map is shown in Fig. 1.

The population was grown in field trials at SCRI in 1994, 1995 and 1996, using row-column designs with two replicates for each trial. The traits heading date (HD), plot yield (PY) and thousand grain weight (TGW) were recorded for each plot. Five plants were taken from each plot prior to harvest in 1994, and four in 1995 and 1996 and the number of fertile stems (TN), the grain number of the main stem (GN), and its grain weight (MSY), and the single plant yield (SPY) were recorded. The population was also grown in a similar field trial at Advanta Seeds UK, Boothby Graffoe, Lincs., UK in 1996, but the only yield characters used here are PY and TGW. More detail about the trials and the traits can be found in Thomas *et al.* (1998). Each trait in each trial was analysed using REML (Genstat 5 Committee, 1995) to check for row and column effects, but these were generally small, and were not included in further analyses. The estimated trait means for each DH line in each trial were used here in the QTL analysis.

A multivariate analysis of variance was carried out for each trial separately to resolve the matrix of total sums of squares and products (SSP) into components for replicate, genotype and residual, following the format of Table 3. The genetic and environmental variance–covariance matrices Σ_G and Σ_E were estimated by equating observed and expected SSP matrices. The genetic correlation between traits X and Y was then calculated as

$$C_{G(XY)} = \Sigma_{G(XY)}/\surd(\Sigma_{G(XX)}\Sigma_{G(YY)}),$$

where $\Sigma_{G(XY)}$ is the genetic covariance between traits X and Y and $\Sigma_{G(XX)}$ and $\Sigma_{G(YY)}$ are the genetic variances. The environmental correlations were calculated similarly.

3 Results

(i) *Biometrical analysis of yield characters*

The genetic and environmental correlations and the heritabilities for traits TN, SPY, MSY, GN, PY and TGW at SCRI in 1994, 1995 and 1996 are shown in Table 4. (In 1996, the data on TN and SPY were erratic for one replicate due to some poor establishment and so these traits were excluded from the biometrical analysis, although the data from the remaining replicate were used for the QTL analysis.)

Except for the genetic correlations for TN, the trait with the lowest heritability, the correlations are quite consistent over trials. MSY and SPY have large genetic and environmental correlations. The environmental correlations between GN and both MSY and SPY are slightly larger than the genetic correlations. The genetic correlation between GN and PY is high for all three trials, but the corresponding environmental correlation is very low. This is also true for the correlation of TGW with SPY and MSY.

(ii) *QTL analysis by multivariate regression*

Analysis of yield characters

A preliminary analysis of individual yield characters indicated that the two dwarfing genes *sdw1* and *ari-eGP* were significantly associated with several traits, and the explanatory variables corresponding to these (3c/1 and 5H/9 respectively, see Fig. 1) were included in all the multivariate regression models. Further explanatory variables were selected successively to minimize the AIC_C.

The explanatory variables selected for inclusion for the 1994 yield characters were, in order of selection

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press

Table 4. *Correlations between six yield characters: number of fertile stems (TN), single plant yield (SPY), main stem yield (MSY), grain number (GN), plot yield (PY) and thousand grain weight (TGW). Environmental correlations are above the diagonal, genetic correlations below the diagonal and the heritabilities (%) are on the diagonal. The three figures in each cell are for trials in 1994, 1995 and 1996 at SCRI*

Trait	TN	SPY	MSY	GN	PY	TGW
TN	12% 18% —	0·83 0·86 —	0·37 0·43 —	0·33 0·36 —	0·15 0·17 —	0·15 0·19 —
SPY	0·68 0·15 —	39% 23% —	0·74 0·75 —	0·56 0·59 —	0·24 0·20 —	0·11 0·26 —
MSY	0·27 −0·48 —	0·88 0·75 —	71% 60% 63%	0·78 0·82 0·84	0·21 0·13 0·07	0·12 0·16 0·41
GN	0·02 −0·08 —	0·44 0·38 —	0·54 0·43 0·66	75% 59% 62%	0·21 −0·04 0·10	0·11 −0·06 0·21
PY	0·11 −0·28 —	0·42 0·23 —	0·48 0·42 0·32	0·63 0·68 0·60	64% 77% 55%	−0·03 −0·45 −0·09
TGW	0·49 −0·35 —	0·68 0·57 —	0·64 0·74 0·71	−0·19 −0·27 −0·02	0·09 −0·10 −0·15	65% 95% 84%

(after 3c/1 and 5H/9): 4H/28, 2b/7, 7H/31, 4H/46, 7H/14, 2b/48, 5H/44, 6H/8 and 4H/39. Three of these are on chromosome 4H, and two, 4H/39 and 4H/46, are only 35 cM apart. The effect on the AIC_C of small shifts in the QTL locations was investigated as described above, and it was found that the replacement of explanatory variables 4H/28, 2b/7 and 7H/31 by 4H/31, 2b/6 and 7H/33 reduced the AIC_C, indicating an improvement in the fit of the model.

Inspection of the QTL effects revealed that the only significant effect for 4H/39 was on the trait TGW, and that the effect on TGW for 4H/46 was of a very similar size, but had the opposite sign. These two effects will therefore cancel each other except for individuals where there has been a recombination on this chromosome between the two positions. Inspection of the fitted values from the models with and without position 4H/39 reveals a substantial change in the fit for just two individuals, which had a recombination between the two positions, together with outlying trait values for more than one of the yield characters. A bootstrap analysis with 500 samples found that the dwarfing genes and the first seven explanatory variables listed above were significant ($P < 0{\cdot}05$) in at least 95% of the samples. However, the last two selected positions, 6H/8 and 4H/39, were significant in fewer than 95% of the bootstrap resamples. This suggests that a small proportion of the individuals may be contributing to the selection of these positions and they were dropped from the model. The QTL effects and their significances in the revised model for the 1994 trial are given in the first line of each cell of Table 5.

The explanatory variable selection and bootstrap analysis were repeated for the SCRI trials in 1995 and 1996, and the results are also shown in Table 5. There was a close correspondence in the explanatory variables selected for each year, although the order of selection varied and selected positions varied by up to 25 cM. Explanatory variable 5H/44 was selected only in 1994 and explanatory variable 7H/55 was only significant by the bootstrap analysis in 1995, although 7H/56 was selected but subsequently excluded by the bootstrap analysis in 1996.

Traits PY and GN had large genetic correlations but low environmental correlations in all three years. The QTL detected at position 7H/14–7H/16 (closest marker P16M47f) had significant negative effects on both these traits in all three years. The negative sign indicates that the allele for increased PY and GN comes from the B83-12/21/5 parent. Powell *et al.* (1997) also report a QTL affecting PY on the short arm of chromosome 7H in a cross between Blenheim and E224/3. The QTL detected at position 4H/46–4H/48 lies closer than the *mlo* locus to the end of the long arm of chromosome 4H. This has a consistent, highly significant negative effect on PY and GN, indicating that a QTL for reduced GN and PY is

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press

Table 5. *The QTL effects (Derkado allele − B83-12/21/5 allele)/2 of the explanatory variables selected for the yield characters. The overall significance of the selected position (by an F-test) and the significance of the individual QTL effects (by a t-test) are shown by asterisks, with *, **, *** indicating significance with P < 0·05, 0·01, 0·001 respectively. The upper, middle and lower figures in each cell are for trials in 1994, 1995 and 1996 at SCRI*

Location	TN	SPY	MSY	GN	PY	TGW
5H/9***	0·085*	0·188***	0·067***	−0·379*	0·191***	3·362***
(*ari-eGP*)	−0·067*	0·026	0·031***	−0·724***	−0·006	3·481***
	−0·050	0·070	0·067***	−0·188	−0·005	2·998***
3c/1***	−0·002	−0·144***	−0·062***	0·216	0·184***	−1·361***
(*sdw1*)	0·071*	−0·058*	−0·069***	0·011	0·041	−3·072***
	−0·015	−0·121**	−0·086***	−0·326	0·096**	−2·234***
4H/31***	0·042	0·034	−0·006	−0·995***	−0·034	1·168***
4H/26***	0·006	0·053	0·024**	−0·153	0·056	1·331***
4H/29***	0·075	0·054	−0·007	−0·764***	−0·030	0·679***
2b/6***	0·087*	−0·028	−0·036**	−0·266	0·068	−0·647*
2b/6***	0·008	−0·055	−0·029**	−0·062	−0·070	−1·369***
2b/6***	−0·047	−0·111*	−0·049***	−0·409	−0·001	−0·869***
7H/30***	−0·023	0·052	0·040***	0·323	0·109*	0·746*
7H/33***	−0·085*	−0·024	0·020*	0·431*	0·054	0·764***
7H/33***	−0·026	0·066	0·020	0·347	0·011	0·707***
4H/46***	−0·107*	−0·201***	−0·051***	−1·093***	−0·269***	0·003
4H/48***	−0·079*	−0·077*	−0·016	−0·612**	−0·182***	0·621**
4H/46***	0·083	0·001	−0·052***	−1·014***	−0·129**	0·172
7H/14***	0·051	−0·052	−0·040***	−1·038***	−0·215***	0·379
7H/16**	−0·035	−0·058	−0·022*	−0·676***	−0·152**	0·564*
7H/16***	0·008	−0·053	−0·019	−0·938***	−0·121**	0·687***
2b/48***	0·101**	0·116**	0·022*	−0·051	0·030	1·354***
2b/48***	−0·024	0·004	0·010	−0·209	−0·041	0·920***
2b/48***	0·032	0·011	−0·005	−0·370	−0·019	0·771***
5H/44**	−0·106**	−0·150**	−0·030**	−0·516**	−0·074	−0·880**
—	—	—	—	—	—	—
—	—	—	—	—	—	—
—	—	—	—	—	—	—
7H/55***	0·031	0·093***	0·041***	0·438**	0·076	0·660**
—	—	—	—	—	—	—

linked to the *mlo* allele in Derkado. Tinker *et al.* (1996) and Bezant *et al.* (1997) have reported yield QTLs in barley which are likely to lie in the same region. Unlike the results of Kjaer *et al.* (1990), there was little evidence of an effect of this QTL on TGW, with a significant result ($P < 0{\cdot}01$) in one of the three years.

A second QTL was detected on chromosome 4H, affecting GN and TGW, but not PY. In this case the QTL effects were of opposite sign, so that the allele from Derkado was associated with an increase in TGW and a decrease in GN. A QTL affecting TGW on this chromosome was also reported by Langridge *et al.* (1996) in a cross Chebec × Harrington. A negative relationship has also been reported for GN and TGW (Rasmussen & Cannell, 1970; Riggs & Hayter, 1975) and probably reflects pleiotropy at a common locus affecting the balance between the two yield components.

TGW had a high genetic correlation with MSY, and to a lesser extent with SPY, and low environmental correlations. The *sdw1* dwarfing gene carried by the Derkado parent at position 3c/1 on the long arm of chromosome 3H had a negative effect on TGW, MSY and SPY in all three years. This locus also had a positive effect on PY in two of the three years. Similar associations with *sdw1* were reported in the Blenheim × E224/3 cross (Thomas *et al.*, 1995; Powell *et al.*, 1997). The *ari-eGP* dwarfing gene carried by the B83-12/21/5 parent at position 5H/9 on chromosome 5H is also associated with a reduction in TGW and MSY. This gene was also associated with a reduction in SPY and PY in 1994, and with an increase in GN in 1994 and 1995. These associations are similar to those found by Powell *et al.* (1985) in the crosses Golden Promise × Mazurka, Golden Promise × Ark Royal and BH4/143/2 × Ark Royal, and by Thomas *et al.* (1991) in a cross TS43/3/5 × Apex.

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press

Table 6. *Correlations between* (*a*) *heading date and* (*b*) *plot yield for different environments. Environmental correlations are above the diagonal, genetic correlations below the diagonal and the heritabilities* (%) *are on the diagonal*
(*a*)

Heading	1994, SCRI	1995, SCRI	1996, SCRI
1994, SCRI	97%	−0·17	0·02
1995, SCRI	0·95	95%	0·01
1996, SCRI	0·92	0·99	90%

(*b*)

Yield	1994, SCRI	1995, SCRI	1996, SCRI	1996, Boothby Graffoe
1994, SCRI	64%	0·11	0·17	−0·09
1995, SCRI	0·69	77%	0·20	0·06
1996, SCRI	0·67	0·84	55%	−0·03
1996, Boothby Graffoe	0·69	0·61	0·67	36%

Two QTLs affecting TGW were found on chromosome 2H. The QTL at position 2b/6 had significant negative effects on TGW and MSY in all three seasons, while that at position 2b/48 had a positive effect on TGW but no consistently significant effects on other traits. Langridge *et al.* (1996) found two QTLs of opposite effects affecting TGW on chromosome 2H, which are likely to be in the same region as those found in the present study. Bezant *et al.* (1997) also report QTLs of opposite signs affecting TGW, one on the long arm and one on the short arm of chromosome 2H in a cross between Blenheim and Kym, but there are no common markers to align the maps.

Multi-environment analysis

The previous analysis can also be used to analyse data on a single trait scored for different environments, i.e. different years and/or sites. In particular, this approach enables us to see which associations are consistent across environments and where there are QTL–environment interactions. We will illustrate this for two traits: heading date (with high heritability) and plot yield (with moderate heritability: see Table 4). Heading dates are available for the three trials at SCRI in 1994, 1995 and 1996, and plot yields are available for these trials and for a fourth trial in Lincolnshire, UK in 1996. Table 6(a) and (b) show

Table 7. *The QTL effects* (*Derkado allele* − *B83-12/21/5 allele*)/2 *of the explanatory variables selected for the multivariate regression analysis of (a) heading date and (b) plot yield in different environments. The overall significance of the selected position (by an F-test) and the significance of the individual QTL effects (by a t-test) are shown by asterisks, with *, **, *** indicating significance with $P < 0{\cdot}05$, 0·01, 0·001 respectively*
(*a*)

Heading	1994, SCRI	1995, SCRI	1996, SCRI
5H/9***	1·04***	0·97***	1·09***
3c/1***	1·93***	2·15***	2·22***
7H/16***	−2·44***	−2·37***	−2·56***
4H/44***	−0·76***	−0·97***	−1·23***
6H/19***	0·94***	0·72***	0·60**

(*b*)

Yield	1994, SCRI	1995, SCRI	1996, SCRI	1996, Boothby Graffoe
5H/9***	0·16***	0·02	−0·01	−0·19***
3c/1***	0·20***	0·03	0·09*	0·08
4H/45***	−0·28***	−0·20***	−0·15***	−0·22***
7H/13***	−0·22***	−0·13***	−0·15***	−0·10

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press

the genetic and environmental correlations and the heritability for each trial for heading date and yield respectively. The environmental correlations are very low, as would be expected, while the genetic correlations are larger. Explanatory variables were selected as described above, and the significance of the selected explanatory variables was examined using bootstrapping. For heading date, three explanatory variables were selected in addition to the *sdw1* and *ari-eGP* genes. One of these (4H/44) lies close to the *mlo* locus, an association discussed by Thomas *et al.* (1998). Bootstrapping confirmed that all these markers had significant coefficients for all three seasons. The signs of the coefficients agreed across seasons, and the coefficients were of similar sizes (Table 7a). For plot yield, four explanatory variables (4H/45, 7H/13, 7H/65 and 5H/55) were originally selected in addition to the *sdw1* and *ari-eGP* genes but the last two of these were not consistently significant in the bootstrap samples and have been excluded from Table 7b. Explanatory variables 4H/45 and 7H/13 are in very similar positions to those selected in the multi-trait analysis (Table 5): small differences in the coefficients are due to the extra explanatory variables included in the models of Table 5. We see that explanatory variable 4H/45, situated close to the *mlo* gene, has a significant effect in every environment. No other explanatory variable has a significant effect in every environment, indicating QTL–environment interactions. This is most apparent for the *ari-eGP* gene at position 5H/9, where the effects at SCRI in 1994 and Boothby Graffoe, Lincs., in 1996 have opposite signs.

4. Conclusions

This paper presents a straightforward approach for multivariate QTL analysis, as a multivariate extension of the regression approach of Haley & Knott (1992). This regression approach is computationally faster than mapping by maximum likelihood, and permits greater flexibility in modelling. Comparisons of the two approaches in the univariate case show that the likelihood profile and the estimates of QTL locations and effects are very similar. Xu (1995) has shown, however, that the regression method can overestimate the residual variance, particularly for large QTL effects or widely spaced markers. This overestimation is due to the inclusion in the residual variance term of QTL variation among individuals with the same marker genotype. In the data set presented here, the marker spacing is quite dense and so biases are likely to be small, and have not been considered here. Unbiased estimators of the residual variance for the multivariate case could, however, be developed in a similar way to that proposed by Xu (1995) for the univariate case.

The multivariate regression approach may be programmed in Genstat or other statistical packages with very little extra complexity compared with the univariate approach of Haley & Knott (1992), and it may be applied to several traits and/or several environments. We propose that a forward stepwise approach is used to select a first set of explanatory variables by minimizing the AIC_C criterion. Positions and pairs of positions close to the selected explanatory variables should then be examined to test whether the AIC_C can be reduced further (for example, a ghost QTL may have been selected in the first set in place of two genuine, linked QTLs). The significance of each explanatory variable in the final set may be assessed by Wilks' Λ statistic, or more easily by transforming this as described by Rao (1973) to obtain a test statistic with an approximate F distribution. Influential individuals in the population may be responsible for the detection of spurious associations with markers or pairs of linked markers (as demonstrated by Hackett (1994) for univariate QTL analysis), and bootstrapping of the selected explanatory variables should be used to confirm whether the significance is dependent on particular individuals.

The question of whether effects on two traits are due to two linked QTLs or to a single QTL with pleiotropic effects is difficult to resolve. In the analyses presented here, more than one QTL was selected on some chromosomes, but the separation was never less than 75 cM. The exception to this were the two explanatory variables selected at positions 4H/39 and 4H/46 for the 1994 data. Both were significant with $P < 0{\cdot}01$ according to Rao's F-test, but an examination of the QTL effects and the fit of the model suggested that overfitting to two individuals was responsible for this significance, rather than two linked QTLs. The ability of our approach to resolve linkage or pleiotropy will in general depend on the population type and size, the heritability of the traits and the degree of genetic and environmental correlation. Lebreton *et al.* (1998) have proposed a test of linkage versus pleiotropy using a combination of bootstrapping and univariate QTL analysis, and a thorough comparison of the approaches would be of great interest.

QTL mapping of a set of correlated traits by mapping their principal components has been proposed by Weller *et al.* (1996) and Mangin *et al.* (1998). This approach is appealing, in that correlated traits are replaced by uncorrelated components. However, it is important in this context to distinguish between phenotypic, environmental and genetic correlation. Mangin *et al.* (1998) work with the phenotypic correlation matrix. As the traits in their example (bacterial wilt in tomato caused by *Ralstonia solanacearum*, measured at 6, 14 and 28 days after inoculation) are intuitively likely to have a high level of genetic correlation their analysis seems appropriate,

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press

but it would not be appropriate for our yield characters, with a mixture of genetic and environmental correlation. A small simulation exercise (data not shown) indicated that QTL mapping of principal components of traits correlated due to a mixture of environmental and genetic causes led to inferences of QTLs where none had been simulated. It is important to use a canonical transformation to set both the genetic and the residual correlations to zero (see Lynch & Walsh, 1998, p. 778 and references therein). The multivariate approach proposed in this paper avoids this problem and should be appropriate for any combination of environmental and genetic correlations.

The analysis uses the same data as the single trait analysis of Thomas *et al.* (1998), who concentrated on detecting QTLs for yield characters on chromosome 4H, especially close to the *mlo* locus. By considering the yield characters together in a multi-trait analysis, we gain more insight into the processes operating at each locus. For example, the Derkado allele at location 4H/46–4H/48 near *mlo* is associated with a decrease in both grain number and plot yield, while the Derkado allele at 4H/26–4H/31 is associated with a decrease in grain number and an increase in thousand grain weight, resulting in no overall change in plot yield. Such information will be useful in the choice of markers for use in marker-assisted breeding schemes.

The work of C.A.H. and W.T.B.T. is supported by the Scottish Executive Rural Affairs Department. R.C.M. is funded under the DTI/LINK scheme. We thank J. Israel and N. Jones of Advanta Seeds for carrying out the trial at Boothby Graffoe, Lincs., UK. We are grateful for the comments of three referees, which improved the presentation of the paper.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (ed. B. N. Petrov & F. Csàki), pp. 267–281. Budapest: Akademia Kiadó.

Basten, C. J., Weir, B. S. & Zeng, Z.-B. (1994). Zmap: a QTL cartographer. In *Proceedings of the Fifth World Congress on Genetics Applied to Livestock Production: Computing Strategies and Software* (ed. C. Smith, J. S. Gavora, B. Benkel, J. Chesnais, W. Fairfull, J. P. Gibson, B. W. Kennedy & E. B. Burnside), vol. 22, pp. 65–66.

Basten, C. J., Weir, B. S. & Zeng, Z.-B. (1999). *QTL Cartographer, version 1.13: A Reference Manual and Tutorial for QTL Mapping*. Department of Statistics, North Carolina State University, Raleigh, NC.

Bedrick, E. J. & Tsai, C.-L. (1994). Model selection for multivariate regression in small samples. *Biometrics* **50**, 226–231.

Bezant, J., Laurie, D., Pratchett, N., Chojecki, J. & Kearsey, M. (1997). Mapping QTL controlling yield and yield components in a spring barley (*Hordeum vulgare* L.) cross using marker regression. *Molecular breeding* **3**, 29–38.

Chatfield, C. & Collins, A. J. (1980). *Introduction to Multivariate Analysis*. London: Chapman & Hall.

Genstat 5 Committee (1995). *Genstat 5 Release 3.2 Manual Supplement*. Oxford: Numerical Algorithms Group.

Hackett, C. A. (1994). Selection of markers linked to quantitative trait loci by regression techniques. In *Biometrics in Plant Breeding: Applications of Molecular Markers. Proceedings of the Ninth Meeting of the EUCARPIA Section Biometrics in Plant Breeding* (ed. J. W. van Ooijen & J. Jansen), pp. 99–106. Wageningen: CPRO-DLO.

Haley, C. S. & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Jiang, C. & Zeng, Z.-B. (1995). Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* **140**, 1111–1127.

Kjaer, B., Jensen, H. P., Jensen, J. & Helms Jørgensen, J. (1990). Associations between three *ml-o* powdery mildew resistance genes and agronomic traits in barley. *Euphytica* **46**, 185–193.

Korol, A. B., Ronin, Y. I. & Kirzhner, V. M. (1995). Interval mapping of quantitative trait loci employing correlated trait complexes. *Genetics* **140**, 1137–1147.

Langridge, P., Karakousis, A., Kretschmer, J., Manning, S., Boyd, R., dao Li, C., Islam, R., Logue, S. & Lance, R. (1996). Waite Agricultural Research Institute. http://greengenes.cit.cornell.edu/WaiteQTL/CxH.html

Lebreton, C. M., Visscher, P. M., Haley, C. S., Semikhodskii, A. & Quarrie, S. A. (1998). A nonparametric bootstrap method for testing close linkage vs pleiotropy of coincident quantitative trait loci. *Genetics* **150**, 931–943.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.

Mangin, B., Thoquet, P. & Grimsley, N. (1998). Pleiotropic QTL analysis. *Biometrics* **54**, 88–99.

Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

Martinez, O. & Curnow, R. N. (1992). Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theoretical and Applied Genetics* **85**, 480–488.

Powell, W., Thomas, W. T. B., Caligari, P. D. S. & Jinks, J. L. (1985). The effects of major genes on quantitatively varying characters in barley. 1. The *Gpert* locus. *Heredity* **54**, 343–348.

Powell, W., Thomas, W. T. B., Baird, E., Lawrence, P., Booth, A., Harrower, B., McNicol, J. W. & Waugh, R. (1997). Analysis of quantitative traits in barley by the use of amplified fragment length polymorphisms. *Heredity* **79**, 48–59.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd edn. New York: Wiley.

Rasmussen, D. C. & Cannell, R. Q. (1970). Selection for grain yield and components of yield in barley. *Crop Science* **10**, 51–54.

Riggs, T. J. & Hayter, A. M. (1975). A study of the inheritance and inter-relationships of some agronomically important characters in spring barley. *Theoretical and Applied Genetics* **46**, 257–264.

Ronin, Y. I., Kirzhner, V. M. & Korol, A. B. (1995). Linkage between loci of quantitative traits and marker loci: multi-trait analysis with a single marker. *Theoretical and Applied Genetics* **90**, 776–786.

Saghai-Maroof, M. A., Biyashev, R. B., Yang, G. P., Zhang, Q. & Allard, R. W. (1994). Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations and population dynamics. *Proceedings of the National Academy of Science of the USA* **81**, 5466–5470.

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press

Stam, P. & Van Ooijen, J. W. (1995). *JoinMap version 2.0: Software for the Calculation of Genetic Linkage Maps*. Wageningen: CPRO-DLO.

Struss, D. & Plieske, J. (1998). The use of microsatellite markers for detection of genetic diversity in barley Populations. *Theoretical and Applied Genetics* **97**, 308–315.

Thomas, W. T. B., Powell, W. & Swanston, J. S. (1991). The effects of major genes on quantitatively varying characters in barley. 4. The *Gpert* and *denso* loci and quality characters. *Heredity* **66**, 381–388.

Thomas, W. T. B., Powell, W., Waugh, R., Chalmers, K. J., Barua, U. M., Jack, P., Lea, V., Forster, B. P., Swanston, J. S., Ellis, R. P., Hanson, P. R. & Lance, R. C. M. (1995). Detection of quantitative trait loci for agronomic, yield, grain and disease characters in spring barley (*Hordeum vulgare* L.) *Theoretical and Applied Genetics* **91**, 1037–1047.

Thomas, W. T. B., Powell, W., Swanston, J. S., Ellis, R. P., Chalmers, K. J., Barua, U. M., Jack, P., Lea, V., Forster, B. P., Waugh, R. & Smith, D. B. (1996). Quantitative trait loci for germination and malting quality characters in a spring barley cross. *Crop Science* **36**, 265–273.

Thomas, W. T. B., Baird, E., Fuller, J. D., Lawrence, P., Young, G. R., Russell, J., Ramsay, L., Waugh, R. & Powell, W. (1998). Identification of a QTL decreasing yield in barley linked to Mlo powdery mildew resistance. *Molecular Biology* **4**, 381–393.

Tinker, N. A., Mather, D. E., Rossnagel, B. G., Kasha, K. J., Kleinhofs, A., Hayes, P. M., Falk, D. E., Ferguson, T., Shugar, L. P., Legge, W. G., Irvine, R. B., Choo, T. M., Briggs, K. G., Ullrich, S. E., Franckowiak, J. D., Blake, T. K., Graf, R. J., Dofing, S. M., Saghai-Maroof, M. A., Scoles, G. J., Hoffman, D., Dahleen, L. S., Kilian, A., Chen, F., Biyashev, R. M., Kudrna, D. A. & Steffenson, B. J. (1996). Regions of the genome that affect agronomic performance in two-row barley. *Crop Science* **36**, 1053–1062.

Weller, J. I., Wiggans, G. R., Van Raden, P. M. & Ron, M. (1996). Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. *Theoretical and Applied Genetics* **92**, 998–1002.

Xu, S. (1995). A comment on the simple regression method for interval mapping. *Genetics* **141**, 1657–1659.

https://doi.org/10.1017/S0016672300004869 Published online by Cambridge University Press