

Multi-environment QTL mapping in blackcurrant (*Ribes nigrum* L.) using mixed models

C. A. Hackett · J. Russell · L. Jorgensen ·
S. L. Gordon · R. M. Brennan

Received: 28 April 2010 / Accepted: 5 July 2010 / Published online: 23 July 2010
© Springer-Verlag 2010

Abstract The first genetic linkage map of blackcurrant, published by Brennan et al. (*Euphytica* 161:19–34, 2008), identified regions where quantitative trait loci (QTLs) for some important traits were located. The analysis was complicated by the fact that the mapping population was found to contain two subgroups, with segregation ratios consistent with these being crossed and selfed offspring. The QTL analysis was based on the trait mean over 3 years and focused on the crossed offspring. Here we proposed a mixed model multi-environment approach for this population. The 3 years are considered as three separate environments, the data from both the selfed and crossed offspring are combined and different residual terms are explored to model the correlation between the years. This permits tests for interactions between the QTLs, the year and the type of offspring (selfed or crossed). This is applied to re-analyse two important traits, anthocyanin concentration and budbreak. Several additional QTLs were identified, some affecting the trait in both the selfed and crossed offspring, others in just one.

Introduction

The cultivated blackcurrant (*Ribes nigrum* L.) is grown throughout temperate areas of Europe and New Zealand, and the fruit is valued for its high levels of ascorbic acid

and anthocyanins (Brennan 2008), particularly as a processing crop for juice production. However, the fresh market for blackcurrant is also increasing due to the perceived health benefits associated with high antioxidant capacity and polyphenolic content (Ghosh et al. 2007; McDougall et al. 2008). The breeding of new cultivars aims to select phenotypes with elevated levels of nutritionally important compounds, but has also recently looked to address emerging issues regarding the ability of some cultivars to withstand increasingly mild winter temperatures that are occurring more regularly as a result of climate change (Jones and Brennan 2009). The mild winters can lead to erratic budbreak in some cultivars, due to non-fulfillment of the relatively high chilling requirement of blackcurrant bushes for dormancy break; this can ultimately cause a reduction in fruit quality, including uniformity of ripeness which then affects anthocyanin content.

For both fruit quality and agronomic traits, whilst significant diversity exists within the *Ribes* germplasm available to breeders (Bordonaba and Terry 2008), the phenotyping of breeding progenies can be a very lengthy and therefore costly process. This is particularly true for fruit quality-related traits, since blackcurrant plants take 2 years to produce any fruit and at least 3 years to come into full cropping. As a result, the identification of QTLs for key traits and the development of linked markers for deployment in breeding programmes offer real opportunities to increase breeding efficiency and accelerate cultivar development.

The first linkage map of blackcurrant was published by Brennan et al. (2008); the analysis was complicated by the discovery that the population used in the study had two distinct subgroups of individuals, with segregation ratios consistent with one subgroup being derived from a cross of the two parents, and the other being derived by selfing. Information from the two subgroups was combined to estimate the linkage map.

Communicated by H. Nybom.

C. A. Hackett (✉)
Biomathematics and Statistics Scotland, Invergowrie,
Dundee DD2 5DA, UK
e-mail: christine@bioss.ac.uk

J. Russell · L. Jorgensen · S. L. Gordon · R. M. Brennan
SCRI, Invergowrie, Dundee DD2 5DA, UK

The QTL mapping presented in Brennan et al. (2008) focused on the crossed offspring, and used phenotypic trait values that were calculated as the means over 3 years of field trials. Marker-trait associations were explored for the selfed offspring, but few associations were detected. This analysis did not address the question of consistency of the QTL effects over the years and between the crossed and selfed offspring. In QTL studies it is important to know whether the effect of a QTL is consistent over different seasons, environments, etc. or whether there is differential expression in different environments. In the latter case, there may be environmental measurements that predict the QTL expression. Significant environmental effects have been noted in the levels of various fruit quality components, including vitamin C (Brennan and Graham 2009) and anthocyanins (Jenkins 2008).

Several authors have addressed the question of QTL \times environment interactions using extensions to interval mapping (Jansen et al. 1995; Tinker and Mather 1995; Jiang and Zeng 1995). Knott and Haley (2000) and Hackett et al. (2001) have used approaches based on multivariate regression and Romagosa et al. (1996) used an additive main effects and multiplicative interaction (AMMI) model.

Mixed models provide a particularly flexible way to model QTL \times environment interactions. The general use of these in crop experiments has been reviewed by Piepho et al. (2003) and Smith et al. (2005). Mixed models can be used to combine data over different environments and to incorporate extra terms such as spatial effects. They can model heterogeneity in the residual variances in different environments, and correlation between trait values across environments. Piepho (2000) and Verbyla et al. (2003) have used mixed models for QTL mapping, applying their analyses to barley populations. Malosetti et al. (2004) focus in particular on the formulation of the random terms for the residual genetic effects not modelled by QTLs, and their interaction with the environment. This work is extended to multi-trait multi-environment models by Malosetti et al. (2008), who apply it to maize.

In this paper, we have developed a mixed model approach to combine the data from the selfed and crossed offspring, and to model the year to year variation in more detail. This has been applied to re-analyse two important traits in blackcurrant, anthocyanin concentration and budbreak.

Materials and methods

Mapping population

The mapping population is described in detail in Brennan et al. (2008). It was designed as a F1 full-sib progeny between diverse parents:

1. SCRI S36/1/100: a seed parent with commercially acceptable fruit quality parameters including anthocyanin and vitamin C content, later budbreak and susceptibility to gall mite (*Cecidophyopsis ribis*).
2. EMRS B1834: a pollen parent with lower fruit quality, earlier budbreak and resistance to gall mite.

One hundred and twenty-five individuals were raised from these parents and grown in an unreplicated trial at SCRI using standard agronomic practises. Fruit samples (500 g) were hand harvested from individual 2-year-old plants in three consecutive years (2001–2003). Juice was extracted with the addition of 2 ml pectinase using a blender, after which the juice was filtered through Whatman No. 3 filter paper to remove particulate matter. Anthocyanin content was then measured as absorbance at 515 nm (E515) using a spectrophotometric method described by Taylor (1989).

The date of budbreak was assessed over 2 years by visual observations made twice weekly across the population in the field at SCRI. Dates were converted into day number for the year.

A total of 237 AFLP and SSR markers were scored on the parents and offspring. As described in Brennan et al. (2008), analysis of these markers showed that the offspring formed two distinct clusters, containing 82 and 43 individuals. The larger group had segregation ratios consistent with a full-sib population from the intended parents, while the smaller group appeared to have been derived from selfing. Additional SSR and SNP markers subsequent to the initial publication of the linkage map were added to the map following the approach of Brennan et al. (2008).

Statistical model

Here we propose a mixed model to combine the crossed and selfed offspring, and to model correlations across the 3 years, and apply it to re-analyse phenotypic data on anthocyanin content and budbreak.

The general form of the mixed model equation is

$$y = X\beta + Zu$$

where y is a vector of length IJ containing the stacked trait values for genotype i in year j , $i = 1, \dots, I$, $j = 1, \dots, J$, β is a vector of fixed effects, X is the design matrix that associates the trait values with the fixed effects, u is a vector of random genetic by environmental (i.e. year) effects and Z is the design matrix that associates trait values with the random effects. The effects u are assumed to have a $N(0, G)$ distribution. Different forms of the variance-covariance matrix G can be compared to establish which model for genetic variances and correlations among years represents the data best.

The simplest model is $G = \sigma_G^2 I$, where I is an $IJ \times IJ$ identity matrix. In this case, all genetic effects are independent and identically distributed (iid). This is unlikely to be realistic, as the performance of genotypes in different years is expected to be correlated. The most complicated model is to fit a completely unstructured model for G , with any correlation possible between any pairs of observations. However, there are too many parameters to estimate in this case. Instead, G can be resolved as a direct product of two factors, genotype and year: $G = G_g \otimes G_y$. We compare four direct product models (Table 1). In the first three, the genotype effects are assumed to be iid. In model I, the year effects are also assumed to be iid, so this is equivalent to the simplest model above. In model II, the year effects are modelled using a diagonal variance–covariance matrix, so that different years can have different variances, but the effects are uncorrelated. In model III, the year effects are modelled by an unstructured matrix to allow for correlation among the years. In model IV, the year effects are again modelled by an unstructured matrix but the genotypes are modelled as separate terms for the crossed and selfed offspring, so that these can have different variances.

Initially, the fixed terms in the model are year, type (i.e. crossed or selfed) and the interaction between these. This model was used to choose the best structure for G , comparing the deviances by a likelihood ratio test. Once a suitable model for G was identified, each molecular marker was included in the model in turn as a factor. The markers were tested for a significant main effect, or a significant interaction with year or type, using an approximate F test. To avoid issues of multiple testing, only marker effects with significance $p < 0.001$ are reported.

All mixed models were fitted using REML in Genstat 12 for Windows (Genstat 2009).

Table 1 Comparison of genetic models for anthocyanin concentration E515 and budbreak BB

Trait	Model	G_g	G_y	Deviance	LRS	df	Sig.
E515	I	id	id	−778.04			
	II	id	diag	−791.27	13.23	2	<0.001
	III	id	unstr	−813.56	22.29	3	<0.001
	IV	Cross + self	unstr	−824.26	10.70	6	0.098
BB	I	id	id	821.47			
	II	id	diag	816.39	5.08	1	0.024
	III	id	unstr	802.50	13.89	1	<0.001
	IV	Cross + self	unstr	800.00	2.50	3	0.475

G_g and G_y are the variance–covariance matrices for the genotype and year effects, LRS is the likelihood ratio test statistic (equal to the change in deviance), df degrees of freedom and $Sig.$ significance. *id*, *diag* and *unstr* indicate identity, diagonal and unstructured matrices

Results

Linkage map

Figure 1 shows the updated linkage map, derived for each parent separately. Markers that segregate in both parents, linking the maps, are underlined. The order of the linking markers is not always consistent between the two parents, probably due to the fairly small size of the population. Because of this, marker genotypes were used as explanatory factors in the mixed model rather than infer QTL genotypes along a grid as Malosetti et al. (2004) and others have done.

Analysis of anthocyanin concentration

Table 1 compares the different models for the genetic variance for anthocyanin content, assessed as E515. A diagonal matrix of year effects (allowing different variances in different years) was significantly better than an iid model ($p < 0.001$), and an unstructured model allowing correlations between years was significantly better than the diagonal model ($p < 0.001$). The separation of the selfed and crossed offspring was not significant. We therefore used model III for analysis including markers. In model III there was a significant interaction between the year and the type ($p = 0.002$). The mean anthocyanin content of the crossed lines was significantly higher than that for the selfed lines, especially in the third year, with mean differences of 0.062, 0.069 and 0.186 (average standard error of difference 0.031), respectively.

No markers were identified as significantly associated at a level $p < 0.001$ with this trait using the approach of Brennan et al. (2008). Using the mixed model analysis, the most significant marker was E40M39-125 on chromosome 6, from the female parent, which was significant with $p < 0.001$. There were no significant interactions between this marker and year or type. Offspring with this marker had a mean anthocyanin concentration 0.073 (standard error (SE) 0.024) lower than offspring without the marker. Marker E45M58-334 on chromosome 3, a 3:1 marker segregating in both parents, also had a significant main effect ($p = 0.001$), with no interactions with year or type. Offspring with this marker had a mean anthocyanin concentration 0.075 (SE 0.027) higher than offspring without the marker. A 3:1 marker on chromosome 4, E45M40-97, showed a significant interaction ($p < 0.001$) with type; selfed offspring with this marker had a mean anthocyanin concentration 0.082 (SE 0.042) higher than those without it, while crossed offspring with the marker had a mean anthocyanin concentration 0.111 (SE 0.034) lower than those without it. Linked markers from the female parent showed a similar, but less significant, relationship.

The analysis of markers from the male parent was based on the crossed offspring only, as these do not segregate in the selfed offspring. The most significant marker was E40M39-149 on chromosome 5. For this marker the main effect was significant with $p < 0.001$, but there was a significant interaction with year ($p = 0.014$): the presence of the marker was associated with an increase in mean anthocyanin content in each year but the size of the effect differed, being 0.105 (SE 0.032), 0.199 (SE 0.040) and 0.053 (SE 0.041) in years 1, 2 and 3, respectively.

The regions with significant associations with anthocyanin content are shown on the linkage map in Fig. 1.

Analysis of budbreak

The analysis of budbreak was similar, but based on 2 years data. Again, the unstructured model for the year effects was significantly better than simpler models ($p < 0.001$) but there was no significant improvement by a separation of the selfed and crossed genotypes. In model III there was a significant interaction between the year and the type ($p = 0.002$). Budbreak was significantly earlier in 2003 than in 2002 for both selfed and crossed individuals ($p < 0.001$) due to differences in environmental conditions. Although the total overall chilling units (hours below 7°C) between October and the end of March were fairly similar in both years, chilling units in the early part of the winter were significantly higher in 2002–2003 than 2001–2002. In 2002, the crossed plants flowered significantly later than the selfs (mean difference 3.1 days, SE 1.00) but in 2003 the types were not significantly different.

The analysis of Brennan et al. (2008), based on the mean over years, found that marker E40M43-220 on chromosome 3 was most significantly associated with budbreak in the crossed individuals. It also found an association with markers from chromosome 8. This analysis confirmed these associations as significant ($p < 0.001$), and showed that there were no significant interactions between these markers and year or type. Offspring with the E40M43-220 marker showed budbreak with a mean date 3.6 days (SE 0.93) earlier than offspring without this marker. Offspring with the marker E40M60-157 on chromosome 8 showed budbreak with a mean date 2.1 days (SE 0.84) earlier than those without. An additional significant marker was also detected, E35M55-128 on chromosome 4 ($p < 0.001$). Again this showed no interactions with year or type. Offspring with this marker showed budbreak with a mean date 2.1 days (SE 0.89) later than offspring without it. All of these markers are segregating in the female parent only.

The regions with significant associations with budbreak are shown on the linkage map in Fig. 1.

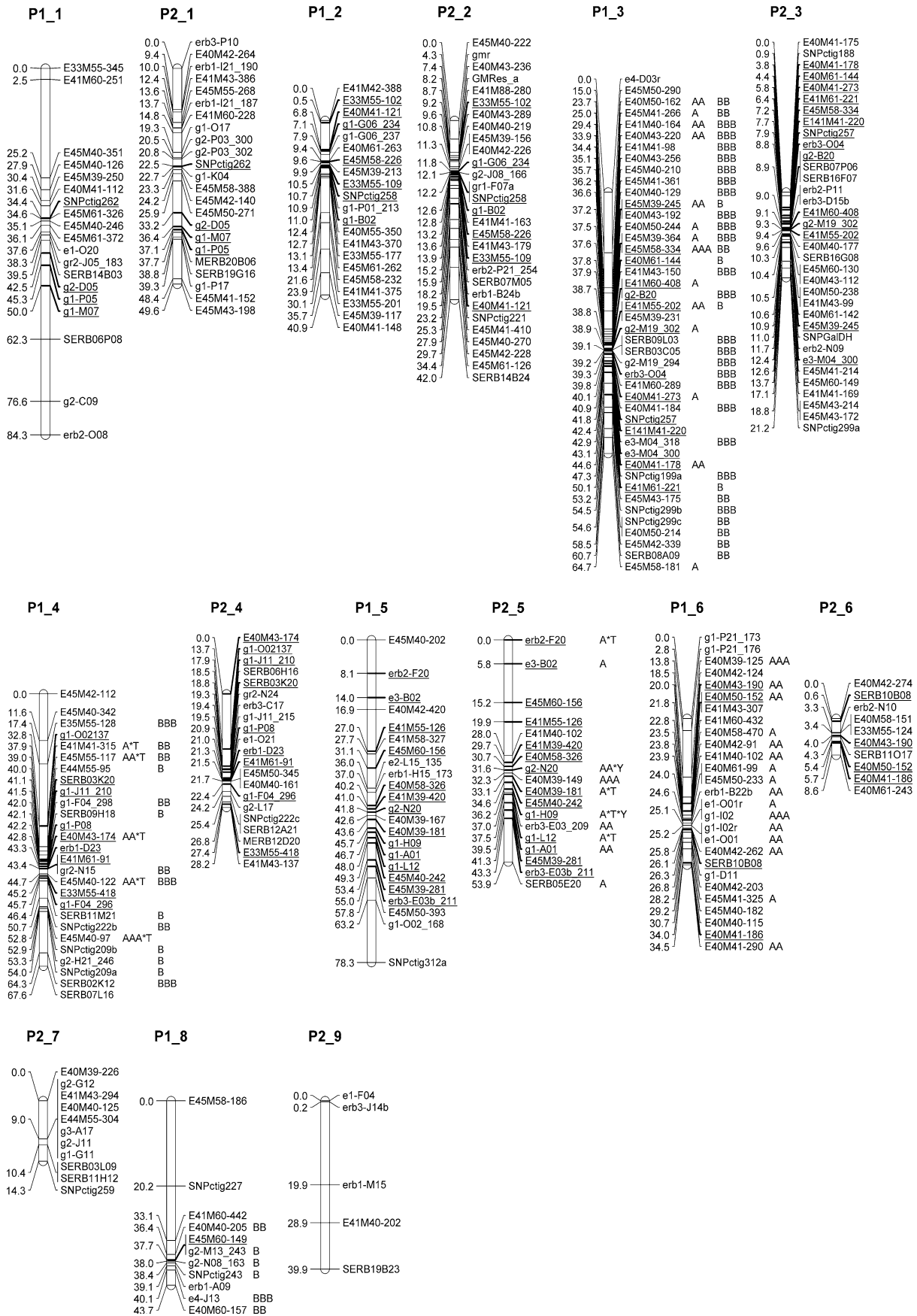
Fig. 1 Linkage map for the two parents separately. P1 is the seed parent SCRI 36/1/100 and P2 is the pollen parent EMRS B1834. *Underlined markers* segregate in both parents and provide a link between the two maps. *A, AA, AAA* show significant main effects for anthocyanin concentration, with $p < 0.05$, < 0.01 and < 0.001 , respectively, and *B, BB, BBB* show significant effects for budbreak. Significant interactions with type (cross or self) are shown as A*T, etc. and significant interactions with year are shown as A*Y, etc

Discussion

The use of markers in woody plant breeding programmes is of increasing interest and importance, since the timescales involved in the phenotyping of many key traits are sufficiently long that the use of markers can have a major impact on the programme's efficiency and cost. In blackcurrant, a PCR-based marker linked to a resistance gene effective against gall mite (*Cecidophyopsis ribis*) is already routinely used in the blackcurrant breeding programme at SCRI (Brennan et al. 2009) to identify resistant germplasm, but there is a growing need for markers linked to more complex traits, particularly those relating to fruit quality and sustainability of production. The present study offers the opportunity to assess large numbers of seedlings within an active breeding programme for fruit quality and developmental traits that are crucial to their eventual potential as commercially viable cultivars, through the use of linked markers. If the markers are sufficiently robust, additional benefits to this approach will be a reduction in selection time by at least 2 years.

The analysis using a mixed model identified several markers that were significantly associated with anthocyanin content and budbreak, most of which were not significant using a simpler analysis. The mixed model also tests these markers for interactions with the environment. Generally there was no significant interaction with the environment, suggesting that these markers are good candidates for use in marker-assisted selection for these traits, which are both known to be affected by environment (Jones and Brennan 2009; Jenkins 2008).

The mixed model structure is extremely flexible for representing correlations among different environments. It also permits more complicated population structures than the well-established biparental populations for QTL mapping, such as the combination of crossed and selfed offspring in this experiment. Mixed models have also been used in association mapping populations, using random genotype effects to model the degree of relatedness based on either pedigree data (for example, Malosetti et al. 2007) or marker data (for example, Yu et al. 2006; Stich et al. 2008). Another possible population structure is that of multiple related biparental populations. Bink et al. (2002) used a Bayesian model incorporating pedigree information to analyse six related crosses in potato with 46–80



individuals, while Arbelbide and Bernardo (2006) used a mixed model to study 80 parentals and 373 inbreds from 158 F₂ populations derived from these parents. These tools allow a much broader approach to QTL mapping and a more detailed exploration of plant resources.

Acknowledgments We are grateful to the Scottish Government Rural and Environment Research and Analysis Directorate (RERAD) for funding this work. Meteorological data was kindly provided by Mark Young (Environment-Plant Interactions programme, SCRI).

References

- Arbelbide M, Bernardo R (2006) Mixed-model QTL mapping for kernel hardness and dough strength in bread wheat. *Theor Appl Genet* 112:885–890
- Bink MCAM, Uimari P, Sillanpää MJ, Janss LLG, Jansen RC (2002) Multiple QTL mapping in related plant populations via a pedigree-analysis approach. *Theor Appl Genet* 104:751–762
- Bordonaba JG, Terry LA (2008) Biochemical profiling and chemometric analysis of seventeen UK-grown black currant cultivars. *J Agric Food Chem* 56:7422–7430
- Brennan RM (2008) Currants and gooseberries. In: Hancock JF (ed) *Temperate fruit crop breeding*. Springer, Berlin, pp 177–196
- Brennan RM, Graham J (2009) Improving fruit quality in *Ribes* and *Rubus* through breeding. *Funct Plant Sci Biotechnol* 3:22–29
- Brennan RM, Jorgensen L, Hackett C, Woodhead M, Gordon SL, Russell J (2008) The development of a genetic linkage map of blackcurrant (*Ribes nigrum* L.) and the identification of regions associated with key fruit quality and agronomic traits. *Euphytica* 161:19–34
- Brennan R, Jorgensen L, Gordon SL, Loades K, Hackett C, Russell J (2009) The development of a PCR-based marker linked to resistance to the blackcurrant gall mite (*Cecidophyopsis ribis* Acari: Eriophyidae). *Theor Appl Genet* 118:205–212
- Genstat (2009) Genstat for Windows Release 12.1. VSN International Ltd, Hemel Hempstead, Hertfordshire
- Ghosh D, McGhie TK, Fisher DR, Joseph JA (2007) Cryoprotective effects of anthocyanins and other phenolic fractions of Boysenberry and blackcurrant on dopamine and amyloid β -induced oxidative stress in transfected COS-7 cells. *J Sci Food Agric* 87:2061–2067
- Hackett CA, Meyer RC, Thomas WTB (2001) Multi-trait QTL mapping in barley using multivariate regression. *Genet Res* 77:95–106
- Jansen RC, van Ooijen JW, Stam P, Lister C, Dean C (1995) Genotype by environment interaction in genetic mapping of multiple quantitative trait loci. *Theor Appl Genet* 91:33–37
- Jenkins GI (2008) Environmental regulation of flavonoid biosynthesis. In: Givens I, Baxter S, Minihane AM, Shaw E (eds) *Health benefits of organic food: effects of the environment*. CABI, Wallingford, pp 240–262
- Jiang C, Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140:1111–1127
- Jones HG, Brennan RM (2009) Potential impacts of climate change on soft fruit production: the example of winter chill in *Ribes*. *Acta Hort* 838:27–32
- Knott SA, Haley CS (2000) Multitrait least squares for quantitative trait loci detection. *Genetics* 156:899–911
- Malosetti M, Voltas J, Romagosa I, Ullrich SE, van Eeuwijk FA (2004) Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* 137:139–145
- Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics* 175:879–889
- Malosetti M, Ribaut JM, Vargas M, Cross J, van Eeuwijk FA (2008) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress traits in maize (*Zea mays* L.). *Euphytica* 161:241–257
- McDougall GJ, Ross HA, Ikeji M, Stewart D (2008) Berry extracts exert different antiproliferative effects against cervical and colon cancer cells grown in vitro. *J Agric Food Chem* 56:3016–3023
- Piepho HP (2000) A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. *Genetics* 156:2043–2050
- Piepho HP, Büchse A, Emrich K (2003) A hitchhiker's guide to mixed models for randomized experiments. *J Agron Crop Sci* 189:310–322
- Romagosa I, Ullrich SE, Han F, Hayes PM (1996) Use of the additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. *Theor Appl Genet* 93:30–37
- Smith AB, Cullis BR, Thompson R (2005) The analysis of cultivar breeding and evaluation trials: an overview of current mixed model approaches. *J Agric Sci* 143:449–462
- Stich B, Möhring J, Piepho HP, Heckenberger M, Buckler ES, Melchinger AE (2008) Comparison of mixed-model approaches for association mapping. *Genetics* 178:1745–1754
- Taylor J (1989) Colour stability of blackcurrant (*Ribes nigrum* L.) juice. *J Sci Food Agric* 49:487–491
- Tinker NA, Mather DE (1995) Methods for QTL analysis with progeny replicated in multiple environments. *J Agric Genom* 1:1–24
- Verbyla AP, Eckermann PJ, Thompson R, Cullis BR (2003) The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. *Aus J Agric Res* 54:1395–1408
- Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208