Theor Appl Genet (2014) 127:1073–1090
DOI 10.1007/s00122-014-2281-3

ORIGINAL PAPER

# Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple

**M. C. A. M. Bink · J. Jansen · M. Madduri · R. E. Voorrips · C.-E. Durel · A. B. Kouassi · F. Laurens · F. Mathis · C. Gessler · D. Gobbin · F. Rezzonico · A. Patocchi · M. Kellerhals · A. Boudichevskaia · F. Dunemann · A. Peil · A. Nowicka · B. Lata · M. Stankiewicz-Kosyl · K. Jeziorek · E. Pitera · A. Soska · K. Tomala · K. M. Evans · F. Fernández-Fernández · W. Guerra · M. Korbin · S. Keller · M. Lewandowski · W. Plocharski · K. Rutkowski · E. Zurawicz · F. Costa · S. Sansavini · S. Tartarini · M. Komjanc · D. Mott · A. Antofie · M. Lateur · A. Rondia · L. Gianfranceschi · W. E. van de Weg**

Received: 26 April 2013 / Accepted: 31 January 2014 / Published online: 25 February 2014


**Abstract**
***Key message*** **Proof of concept of Bayesian integrated QTL analyses across pedigree-related families from breeding programs of an outbreeding species. Results include QTL confidence intervals, individuals' genotype probabilities and genomic breeding values.**
*Abstract* Bayesian QTL linkage mapping approaches offer the flexibility to study multiple full sib families with known pedigrees simultaneously. Such a joint analysis increases the probability of detecting these quantitative trait loci (QTL) and provide insight of the magnitude of QTL across different genetic backgrounds. Here, we present an improved Bayesian multi-QTL pedigree-based approach on an outcrossing species using progenies with different (complex) genetic relationships. Different modeling assumptions were studied in the QTL analyses, i.e., the a priori expected number of QTL varied and polygenic effects were considered. The inferences include number of QTL, additive QTL effect sizes and supporting credible intervals, posterior probabilities of QTL genotypes for all individuals in the dataset, and QTL-based as well as genome-wide breeding values. All these features have been implemented in the FlexQTL™ software. We analyzed fruit firmness in a large apple dataset that comprised 1,347 individuals forming 27 full sib families and their known ancestral pedigrees,

Communicated by M. J. Sillanpaa.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00122-014-2281-3) contains supplementary material, which is available to authorized users.

M. C. A. M. Bink (✉) · J. Jansen
Biometris, Wageningen University and Research Centre, Droevendaalsesteeg 1, P.O. Box 16, 6700 AA Wageningen, The Netherlands
e-mail: marco.bink@wur.nl

M. Madduri · R. E. Voorrips · W. E. van de Weg
Plant Breeding, Wageningen UR, Droevendaalsesteeg 1, P.O. Box 16, 6700 AA Wageningen, The Netherlands

C.-E. Durel · A. B. Kouassi · F. Laurens · F. Mathis
INRA, UMR1345 Institut de Recherche en Horticulture et Semences, SFR 4207 Quasav, Pres L'UNAM, 49071 Beaucouzé, France

C.-E. Durel · A. B. Kouassi · F. Laurens
UMR1345 Institut de Recherche en Horticulture et Semences, Université d'Angers, 49045 Angers, France

C.-E. Durel · A. B. Kouassi · F. Laurens
UMR1345 Institut de Recherche en Horticulture et Semences, AgroCampus-Ouest, 49045 Angers, France

*Present Address:*
A. B. Kouassi
Université Félix Houphhoët-Boigny, Unité de Formation et de Recherche (UFR) 'Biosciences', Laboratoire de Génétique, 22BP 582 Abidjan 22, Abidjan, Côte d'Ivoire

*Present Address:*
F. Mathis
Fabienne Mathis, VEGEPOLYS, Pôle de compétitivité, 7 rue Dixmeras, 49044 Angers Cedex 01, France

C. Gessler · D. Gobbin · F. Rezzonico · A. Patocchi
Plant Pathology, Institute of Integrative Biology (IBZ), ETH Zurich, 8092 Zurich, Switzerland

*Present Address:*
D. Gobbin
Tecan Group Ltd., 8708 Männedorf, Switzerland

F. Rezzonico · A. Patocchi · M. Kellerhals
Research Station Agroscope, Schloss 1, 8820 Wädenswil, Switzerland

with genotypes for 87 SSR markers on 17 chromosomes. We report strong or positive evidence for 14 QTL for fruit firmness on eight chromosomes, validating our approach as several of these QTL were reported previously, though dispersed over a series of studies based on single mapping populations. Interpretation of linked QTL was possible via individuals' QTL genotypes. The correlation between the genomic breeding values and phenotypes was on average 90 %, but varied with the number of detected QTL in a family. The detailed posterior knowledge on QTL of potential parents is critical for the efficiency of marker-assisted breeding.

## Introduction

The ongoing quantitative trait loci (QTL) analyses of complex traits in outcrossing plants and animals contributed to the understanding of quantitative trait genetics through the discovery of many QTL. However, few of these QTL have been adopted by breeders for marker-assisted breeding (MAB) due to various reasons including the following:

- The majority of QTL discoveries have been based on germplasm with a narrow genetic basis—often just a single progeny (King et al. 2000; Maliepaard et al. 2001; Quilot et al. 2004; Fanizza et al. 2005; Kenis et al. 2008; Costa et al. 2010; Pinto et al. 2010; Zhang et al. 2010; Lerceteau-Köhler et al. 2012)—and probably only a small proportion of the total number of relevant QTL has been detected which may explain only a limited fraction of the total genetic variance present in a breeding program.
- Many useful alleles are missed as these are not present or do not segregate into specific single mapping families; application in MAB would thus lead to genetic erosion.
- For most QTL little is known of their mode of action and their robustness in different genetic backgrounds, i.e., the estimated magnitude of the QTL may be different for families derived from other parents.
- The application of MAB becomes redundant if the favorable QTL allele is already present in high frequency in the breeding population. In the latter case, MAB may still be applicable when crosses with new unrelated germplasm are considered.
- The transferability of linkage phase between QTL and marker alleles over genetic backgrounds is unclear when marker densities are moderate to low. Without confirmation in relevant material, MAB approaches based on such limited information risk being inefficient or even counter-productive. Besides, estimated confi-

*Present Address:*
F. Rezzonico
Research Group Environmental Genomics and Systems Biology, Institute of Natural Resource Sciences, Zürich University of Applied Sciences ZHAW, Grüental, 8820 Wädenswil, Switzerland

A. Boudichevskaia · F. Dunemann · A. Peil
Institute for Breeding Research on Horticultural Crops, Julius Kühn-Institut, Pillnitzer Platz 3a, 01326 Dresden, Germany

*Present Address:*
A. Boudichevskaia
Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Corrensstr. 3, 06466 Gatersleben, Germany

*Present Address:*
F. Dunemann
Julius Kühn-Institut, Institute for Breeding Research on Horticultural Crops, Erwin Baur Str. 27, 06484 Quedlinburg, Germany

A. Nowicka
Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences, SGGW, 02-776 Warsaw, Poland

B. Lata · M. Stankiewicz-Kosyl
Laboratory of Basic Research in Horticulture, Faculty of Horticulture, Biotechnology, and Landscape Architecture, Warsaw University of Life Sciences SGGW, 02-776 Warsaw, Poland

K. Jeziorek · E. Pitera · A. Soska · K. Tomala
Department of Pomology, Faculty of Horticulture, Biotechnology and Landscape Architecture, Warsaw University of Life Sciences, SGGW, 02-776 Warsaw, Poland

K. M. Evans · F. Fernández-Fernández
East Malling Research, New Road, East Malling, Kent ME19 6BJ, UK

*Present Address:*
K. M. Evans
Washington State University (WSU-TFREC), 1100 N. Western Avenue, Wenatchee, WA 98801, USA

W. Guerra
Research Centre for Agriculture and Forestry Laimburg, 39040 Vadena, BZ, Italy

M. Korbin · S. Keller · M. Lewandowski · W. Plocharski · K. Rutkowski · E. Zurawicz
Research Institute of Horticulture, 96-100 Skierniewice, Poland

F. Costa · S. Sansavini · S. Tartarini
Department of Fruit and Woody Plant Science, Current Department of Agricultural Sciences, University of Bologna, Via Fanin 46, 40127 Bologna, Italy

dence intervals for QTL positions are usually large, and application would thus result in significant linkage drag.

These issues may be alleviated by QTL mapping in multiple families from ongoing breeding programs, increasing the probability of identifying critical loci and alleles and testing their modes of action in a range of genetic backgrounds and environments that are relevant to breeders, making results more generally applicable. The use of breeding material in genetic research has several additional advantages: a major reduction in experimental costs, since plant materials and part of the phenotypic measurements are already available. Also, continuously increasing numbers of individuals and phenotypic data over time will strengthen the statistical power. Moreover, available pedigree records are used to exploit known genetic structures. The interest in the use of multiple genetically related plant populations in dissecting quantitative trait variation into underlying QTLs has grown rapidly (Blanc et al. 2006; Yu et al. 2008; Huang et al. 2011). In the presence of pedigree structures, the explicit modeling of familial relatedness in QTL and association mapping approaches may significantly improve the power of detection (Bink and Van Arendonk 1999; Yu et al. 2006) To date, the experimental setup of such QTL studies in plants is often restricted to predefined fixed designs such as factorial or diallel to allow standard statistical analyses. To better explore available full sib (FS) families, more flexible statistical procedures are required to utilize complex pedigree relationships. Bayesian approaches to pedigree-based multiple QTL mapping have been proposed and applied in human and animal genetics (Heath 1997; Bink and Van Arendonk 1999; Uimari and Sillanpaa 2001). These approaches exploit the identity by descent (IBD) principle for linking haplotypes over successive generations in known pedigrees (Thompson 2008).

F. Costa · M. Komjanc · D. Mott
Department of Genetics and Biology of Fruit Crops, Research and Innovation Centre, Foundation Edmund Mach, Via Mach 1, 38010 Trento, Italy

A. Antofie · M. Lateur · A. Rondia
Walloon Agricultural Research Centre (CRA-W), Liroux 9, 5030 Gembloux, Belgium

*Present Address:*
A. Antofie
Direction Générale Qualité et Sécurité, Métrologie Légale SPF Economie, PME, Classes Moyennes et Energie, North Gate, Bd du Roi Albert II, 16, 1000 Bruxelles, Belgium

L. Gianfranceschi
Department of Biosciences, University of Milan, Via Celoria 26, 20133 Milan, Italy

The presence of multiple QTL with minor phenotypic effects that usually remain below the detection threshold (Hayes and Goddard 2001) is usually referred to as the polygenic variance component. Accounting for such polygenic effects will likely increase the power and precision to detect and locate real QTL and will also avoid false-positive results (Yu et al. 2006, 2008; Stich et al. 2008).

The European project HiDRAS ('High-quality Disease Resistant Apples for a Sustainable agriculture') (Gianfranceschi and Soglio 2004; Patocchi et al. 2009) was initiated in 2003 to deliver proof of concept on the use of integrated QTL analyses over multiple pedigreed FS families of an outbreeding species. The project included the further development of the critical statistical tools (Bink et al. 2008a, b; Jansen et al. 2009) and molecular marker infrastructure (Silfverberg-Dilworth et al. 2006) as well as the SSR-genotyping procedures (Patocchi et al. 2009), validation of pedigrees (Evans et al. 2011) and phenotyping for a series of fruit quality traits (Kouassi et al. 2009). These data have been stored in a dedicated private AppleBreed database (Antofie et al. 2007) to facilitate easy access by breeders and geneticists. Moreover, software has been developed to visualize phenotypic and genotypic data for related individuals (Voorrips et al. 2012). The experimental design comprised 350 cultivars and breeding lines and 27 FS families interconnected in a complex pedigree that are part of ongoing breeding programs in four European countries.

The main objective of the current paper is to present the feasibility and utilization of the integrated QTL analyses of complex traits over multiple FS families of an outcrossing plant species when dealing with complex datasets comprising diverse pedigree structures. Here, we (1) present the flexible Bayesian framework for QTL analysis as implemented in the FlexQTL™ software (www.flexqtl.nl) to study genetic models with additive QTL and polygenic effects, (2) perform QTL analyses of a complex trait in 27 related and pedigreed FS families of apple and (3) illustrate how breeders can strengthen their breeding decisions by making use of the identified QTL, the individuals' QTL genotypes and their genomic breeding value (GBV) estimates. The analyses are performed for the trait fruit firmness as assessed after 2 months of cold storage, which is a major fruit quality trait in apple. The mapped QTL are compared to previously reported QTL.

## Materials and methodology

### HiDRAS data

All marker and phenotypic data have been generated and pedigrees validated in the EU project HiDRAS (www.hidras.unimi.it) (Gianfranceschi and Soglio 2004)

and retrieved from the dedicated private HiDRAS AppleBreed database (Antofie et al. 2007). The addition of marker data and consistency checking with pedigree information are still ongoing; for the current study we have taken the data as available on 01 June 2012 (Online Resource 1).

*Germplasm*

The plant material used in our study consisted primarily of 27 full sib (FS) families (mapping populations), with a total of 1,349 individuals. These FS families were created by crosses among 33 parents and originated from five different breeding programs from four European countries (INRA-France; JKI-Germany; RCL-Italy; RIPF-Poland; and SSGW-Poland) (Fig. 1). Their pedigree relationships are presented in Online Resource 2. The FS families varied in size from 26 to 96 genotyped individuals, but most families comprised about 50 individuals. The number of individuals is slightly lower and the range of family sizes is slightly smaller than in Patocchi et al. (2009), due to exclusion of individuals with erroneous parentage (as revealed by the marker data) and of individuals for which phenotypic data were lacking. The pedigree records of the FS families traced back several generations to 40 founder individuals, i.e., individuals with both parents unknown. These 40 founders and the intermediate individuals were also included in pedigree data and were genotyped when DNA was available.

*Phenotypic data*

Fruit firmness is a key fruit quality trait of apple (Wei et al. 2010; Costa et al. 2012). Firmness after 2 months of cold storage is a good indicator for the storability of apple (Kouassi et al. 2009). Firmness was instrumentally measured in three successive years, i.e., 2003, 2004 and 2005, and at five different sites (see before) throughout Europe. The trait values are the means of a total of 20 assessments per individual/year at two opposite sides of ten fruits, using a penetrometer, the type of which varied among partners. Scores correspond to the maximal force required for a cylindrical probe of 2 cm long and 1 cm wide to penetrate into the peeled fruit up to a depth of 7 mm. The 27 FS families were grown and phenotyped at one of the five different locations and in several cases not recorded for all 3 years due to bi-annual fruit bearing. A reference set of 30 standard apple cultivars was present at each of the five locations and these individuals were used to pre-adjust the phenotypic data for location (including type of technical instruments) and year effects. Each observation was modeled as a linear function of a grand mean, year, location and genotype. We used GenStat software (Genstat Committee 2004) to fit a linear model to obtain the best linear unbiased prediction (BLUP) values for all individuals with phenotypes. These BLUP values are taken as the trait phenotypes in our QTL analysis (available in Online Resource 1).

**Fig. 1** Phenotype histograms for fruit firmness after 2 months of cold storage for the 27 full sib families with size ranging from 24 to 83 (Fig. 1). The names of the parents, the number of progeny with phenotypes and the breeding program are given for each family. Note that several parents were used multiple times (both as father and/or mother)

*Phenotypic distributions*

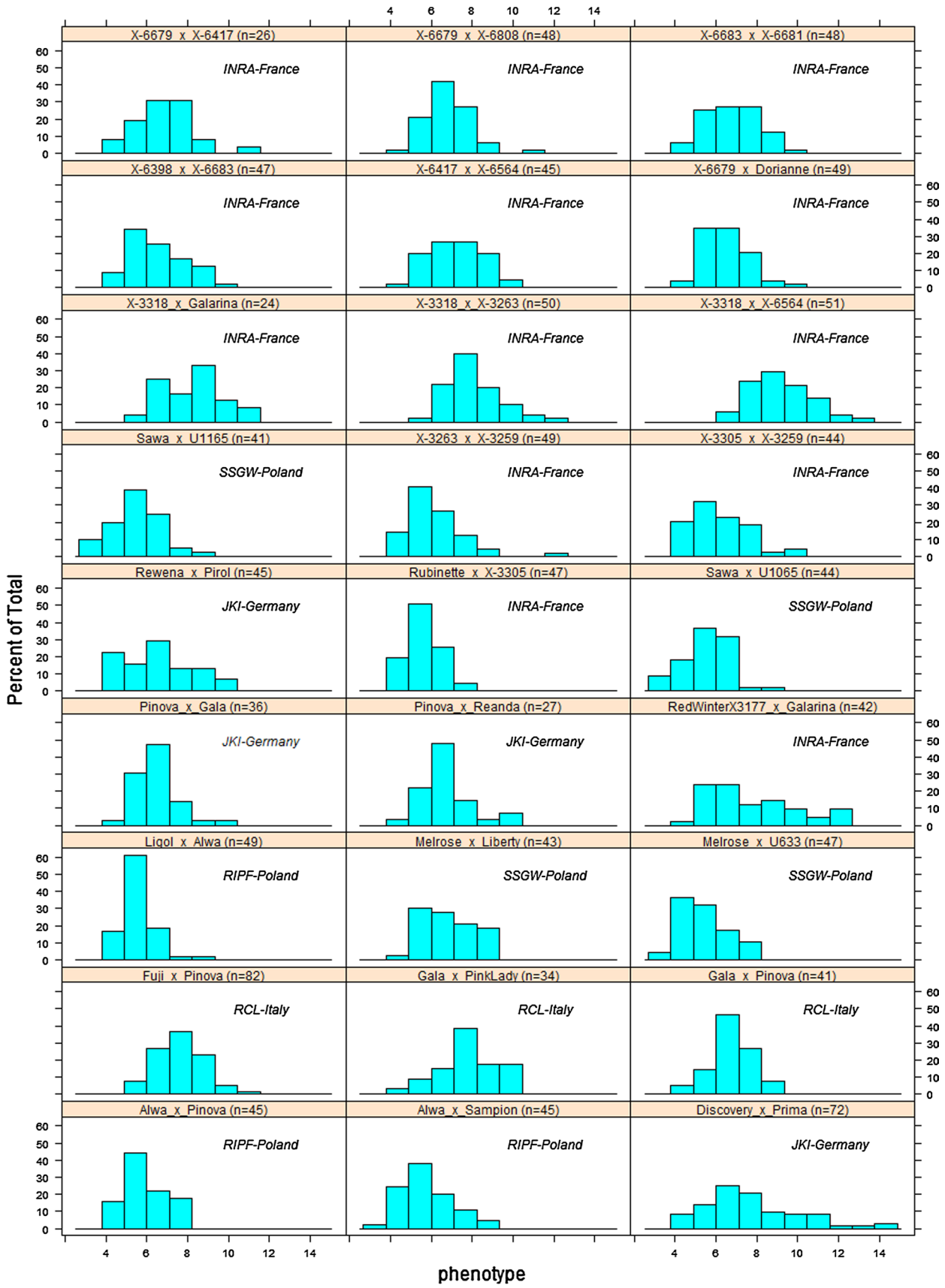
The distributions of phenotypes across the 27 FS families show considerable variation with the largest and smallest range (and variance) for the FS families derived from 'Discovery' × 'Prima' and 'Ligol' × 'Alwa', respectively (Fig. 1). No outliers were present that could reduce the overall quality of the data.

*Marker data*

A set of 87 simple sequence repeat (SSR) loci was examined covering 17 chromosomes and spanning about 11 Morgan (Patocchi et al. 2009). The average distance between neighboring markers was 13 centiMorgan (cM); however, gaps up to 40 cM occurred on chromosomes 3, 6 and 15 (Online Resource 3). Some chromosomal regions were not covered, e.g., the lower and upper parts of chromosome 7 due to absence of suitable SSR markers at the time of genotyping in the HiDRAS project. The order and distances of markers on the linkage map were primarily based on the reference population 'Fiesta' × 'Discovery' (Silfverberg-Dilworth et al. 2006). The length of the 'Fiesta' × 'Discovery' map was over 1,500 cM and only 73 % thereof was covered in this study. Details on the treatment of null alleles and the check of consistency of marker data between parents and offspring and inheritance patterns are given in Online Resource 3.

Bayesian modeling for QTL mapping

The dissection of quantitative traits into genetic components was explored via a Bayesian approach (Gelman et al. 2004) as implemented in the FlexQTL™ software (Bink et al. 2002, 2008b, 2012). A major feature of this Bayesian approach was the implicit exploration of competing models with respect to different numbers of QTL explaining the phenotypic trait variation. In statistical terms, the number of QTL is treated as a random variable and the posterior distribution is estimated.


X-6679 x X-6417 (n=26)
INRA-France
X-6679 x X-6808 (n=48)
INRA-France
X-6683 x X-6681 (n=48)
INRA-France
X-6398 x X-6683 (n=47)
INRA-France
X-6417 x X-6564 (n=45)
INRA-France
X-6679 x Dorianne (n=49)
INRA-France
X-3318_x_Galarina (n=24)
INRA-France
X-3318_x_X-3263 (n=50)
INRA-France
X-3318_x_X-6564 (n=51)
INRA-France
Sawa x U1165 (n=41)
SSGW-Poland
X-3263 x X-3259 (n=49)
INRA-France
X-3305 x X-3259 (n=44)
INRA-France
Rewena x Pirol (n=45)
JKI-Germany
Rubinette x X-3305 (n=47)
INRA-France
Sawa x U1065 (n=44)
SSGW-Poland
Pinova_x_Gala (n=36)
JKI-Germany
Pinova_x_Reanda (n=27)
JKI-Germany
RedWinterX3177_x_Galarina (n=42)
INRA-France
Ligol x Alwa (n=49)
RIPF-Poland
Melrose x Liberty (n=43)
SSGW-Poland
Melrose x U633 (n=47)
SSGW-Poland
Fuji x Pinova (n=82)
RCL-Italy
Gala x PinkLady (n=34)
RCL-Italy
Gala x Pinova (n=41)
RCL-Italy
Alwa_x_Pinova (n=45)
RIPF-Poland
Alwa_x_Sampion (n=45)
RIPF-Poland
Discovery_x_Prima (n=72)
JKI-Germany
Percent of Total
phenotype

*Quantitative trait loci (QTL)*

The Bayesian model takes each QTL to be biallelic, allowing three genotypes to be distinguished, i.e., *QQ*, *Qq*, and *qq*, having genotypic values equal to $+a$, $d$ and $-a$, respectively. The variables $a$ and $d$ represent the additive and dominance effects of a single locus. In this study, all models excluded dominance, i.e., $d = 0$, although including dominance effects would be straightforward. The positions for the putative QTL were denoted by $\lambda$ and were specified in centiMorgan. The QTL genotypes of individuals were a priori unknown and modeling was based on the independent assignment of alleles $Q$ and $q$ to founders (=individuals without known parents) and segregation indicators to trace transmission from parents to offspring (Thompson 2008). Note that the software implementation requires that either both or none of the parents are known and dummy founders were introduced for a small set of single known parents.

*Probability model*

The full probability model for the vector of phenotypes (**y**) was defined by a linear QTL model with several factors that might affect our trait of interest,

$$\mathbf{y} \sim N\left(\mathbf{1}\mu + \mathbf{Xb} + \mathbf{Wa}, \sigma_e^2\right) \quad (1)$$

where $\mu$ is an intercept (overall mean); **b** are environmental effects (if present), and **a** a vector of regression coefficients on the QTL covariates. In addition, **X** is the design matrix for environmental effects and **W** a design matrix that links the QTL effects to the observed phenotypes. Note that the total number of columns in **W** is proportional to the number of QTL ($N_{QTL}$) and thus its dimension varies with the number of QTL in the model along the Markov chain simulation process.

*Prior distributions*

The Bayesian modeling assigned normal priors to the vectors **a** and **e** in Eq. (1), i.e., $\mathbf{a} \sim N(\mathbf{0}, \mathbf{I}\sigma_a^2)$, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$, $\sigma_a^2$ is the per QTL explained variance (cf. Bink et al. (2008b)) and $\sigma_e^2$ is the residual variance. The variances were estimated in the model using inverse Gamma distributions as priors (Bink et al. 2008b). The variables $\mu$ and **b** were assigned uniform prior distributions ("fixed effects" in a non-Bayesian model). The QTL allele frequency ($f_a$) took any value between 0 and 1 with equal prior probability. The QTL positions ($\lambda$) were assigned a uniform distribution along those genome regions covered by markers. A Poisson distribution was taken as prior for the number of QTL in the model. Different mean values for the Poisson distribution, i.e., $E(N_{QTL}) = (1, 5, 10)$, were evaluated to assess sensitivity of posterior inference to the prior assumptions. It should be noted that the priors for the number of QTL were specified after several (short) preliminary MCMC simulation runs. Also, different values may suit other datasets as appropriate values will differ from case to case. For example, high values may be considered in case of statistically powerful datasets comprising large numbers of individuals and high marker density and quality.

*The Infinitesimal model (TIM) for polygenic effects*

Hayes and Goddard (2001) reported on the distribution of the effects of loci affecting quantitative traits and concluded that many loci having small contributions will be missed in QTL mapping experiments. Furthermore, the genetic map in this study contained several chromosomal regions with poor or absent marker coverage, and QTL residing in these 'unmarked' regions may also remain undetected in the default QTL model. The joint contribution of the group of loci with small contributions and the group of 'unmarked' QTL could be modeled via the inclusion of a polygenic component (Lynch and Walsh 1998) into the QTL model (1). Based on the known genealogy, i.e., each offspring had its two parents specified in the dataset, we derived the additive genetic relationship matrix **A** with its entries equal to twice the coefficients of coancestry between individuals (Lynch and Walsh 1998). The full probability model of Eq. (1) was extended to include polygenic effects, i.e., $\mathbf{y} \sim N(\mathbf{1}\mu + \mathbf{Xb} + \mathbf{Wa} + \mathbf{Zu}, \sigma_e^2)$, where **u** is a vector of polygenic effects assuming $\mathbf{u} \sim N\left(\mathbf{0}, \mathbf{A}\sigma_u^2\right)$ in which $\sigma_u^2$ is the polygenic variance and **Z** the design matrix linking the polygenic effects to the observed phenotypes. For reasons of comparison, we also fitted the full probability model with polygenic effects and without QTL effects, i.e., $\mathbf{y} \sim N(\mathbf{1}\mu + \mathbf{Xb} + \mathbf{Zu}, \sigma_e^2)$.

*Posterior sampling by simulation*

Markov chain Monte Carlo (MCMC) simulation (Gilks et al. 1996) as implemented in the software FlexQTL™ (Bink et al. 2002, 2008b) was applied to obtain samples from the joint posterior distribution of the variables in the probability model (1).

$$f(\mu, \mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2, f_a, \lambda, N_{QTL}|\mathbf{y}) \quad (2)$$

The Monte Carlo accuracy was monitored and the length of simulation chains was required to be equivalent to at least 100 effective chain samples (Gelman and Rubin 1992; Sorensen and Gianola 2002). Assessment of convergence was also performed via monitoring the mixing between competing models with different numbers of QTL (Brooks et al. 2003) and confirmed the results for effective samples. The FlexQTL™ software produced trace plots similar to

the R/coda package (Plummer et al. 2006) to inspect convergence visually. In all analyses, a simulation length of 500,000 iterations was sufficient. To save computer storage space, we applied a thinning of 500 along subsequent samples. Thus, 1,000 stored samples were available for further posterior inferences.

*Posterior sampling by simulation*

The narrow sense heritability was estimated by

$$h^2 = \left(\sigma_p^2 - \sigma_e^2\right) \Big/ \left(\sigma_p^2\right), \tag{3}$$

with the percentage of phenotypic variance explained by all additive genetic factors in the model jointly and $\sigma_p^2 = \sigma_g^2 + \sigma_e^2$.The genetic variance ($\sigma_g^2$) comprises potentially two components: $\sigma_g^2 = \sigma_{QTL}^2 + \sigma_u^2$, i.e., the QTL variance and the polygenic variance and some of the studied models omit one of the variances. The genetic variance explained by all QTL jointly may be estimated from

$$\sigma_{QTL}^2 = \sum_j^{N_{QTL}} \left(2 f_{a_j}(1 - f_{a_j})\left[a_j\right]^2\right), \tag{4}$$

where Hardy–Weinberg equilibrium and linkage equilibrium among QTL were assumed in the initial founder population (Falconer 1989). Note that the founder allele frequency may be somewhat different from the allele frequency in the dataset when there is unequal representation of founder alleles in the FS families.

*QTL model selection via Bayes factors*

The use of posterior probabilities on the number of QTL is most straightforward, but can be severely affected by prior assumptions when there is limited evidence coming from the data. Alternatively, for each chromosome the inference on the number of QTL was based on a pairwise comparison of models differing by one QTL from each other via the use of Bayes factors (BF) (Kass and Raftery 1995). The genome-wide total number of QTLs is then obtained by the summation of the numbers of QTLs for the individual chromosomes. Taking two times the natural log of Bayes factors, denoted 2lnBF, allowed a more easy interpretation as this transformed statistic has a similar scale to the likelihood ratio test. A value for 2lnBF that was greater than 2, 5 and 10, indicating positive, strong, and decisive evidence, respectively, in favor of a second QTL model (Kass and Raftery 1995). For each of the 17 chromosomes, we calculated the $2\text{lnBF}_{10}$ values indicating the evidence for a 1-QTL model over a 0-QTL model. Note that negative values for $2\text{lnBF}_{10}$ indicate the favoring of a 0-QTL model. Similarly, we calculated the $2\text{lnBF}_{21}$ and $2\text{lnBF}_{32}$ values for chromosomes with evidence for multiple QTL.

*QTL mapping: bin-wise evidence*

The inferences on plausible QTL positions are based on posterior QTL probability (or intensity) estimates (Sillanpaa and Arjas 1998) on a 2 cM binning of the genome. The binning size of 2 cM was chosen to obtain a smoothed profile and also because of the moderately low marker density along the genome. Also, we used Bayes factors to identify chromosomal regions with positive QTL evidence ($2\text{lnBF} \geq 2$). The Bayes factor for a chromosomal bin was defined as the ratio of the posterior odds to the prior odds for inclusion versus exclusion of the bin (Kass and Raftery 1995). For a total genome length of 1,133 cM and $\text{E}(N_{QTL}) = (1, 5, 10)$, the prior probability for a 2 cM bin was equal to $1.77 \times 10^{-3}$, $0.88 \times 10^{-3}$, $1.77 \times 10^{-2}$, respectively. The threshold of $2\text{lnBF} > 2$ (positive evidence) yielded the posterior inclusion probability thresholds equal to $0.48 \times 10^{-2}$, $2.36 \times 10^{-2}$, $4.86 \times 10^{-2}$, for $\text{E}(N_{QTL}) = (1, 5, 10)$, respectively. For example, a particular chromosomal bin with a posterior probability of 0.04 would be included for the models with $\text{E}(N_{QTL}) = (1, 5)$, but not for the $\text{E}(N_{QTL}) = (10)$.

*QTL effects*

For each 2 cM bin that passed the posterior inclusion threshold (previous section), we estimated the posterior mean and the 90 % credible (or confidence) region that was bounded by the 5 and 95 % quantiles. These credible regions reflected the remaining uncertainty on the effect sizes and were expected to be smaller for QTL that segregated in more FS families.

*QTL genotype assignment*

For each individual the stored samples of QTL genotypes ($QQ$, $Qq$, and $qq$) were used to estimate posterior probabilities of QTL genotypes for those 2 cM bins that exceeded the posterior inclusion probability. To examine certainty on the posterior genotype probabilities, denoted as $p(gtp|\mathbf{y})$, we used,

$$\frac{p(gtp|\mathbf{y})}{1 - p(gtp|\mathbf{y})} = \text{BF} \times \frac{p(gtp)}{1 - p(gtp)} \tag{5}$$

where $p(gtp)$ denoted the prior genotype probability. Again, we applied a Bayes factor threshold of $2\text{lnBF} = 2$ (positive evidence). Starting from the prior QTL allele frequency ($f_a$) equal to 0.5, the prior QTL genotype probability distribution of $p(gtp = QQ, Qq, qq) = (0.25, 0.50, 0.25)$ yielded posterior genotype probability threshold

values of $p(gtp = QQ, Qq, qq \,|\mathbf{y}) = (0.48, 0.73, 0.48)$. To avoid problems of unique assignment of (homozygous) genotypes, we slightly increased the Bayes factor threshold up to 2.2 that gave posterior genotype probability threshold values of $p(gtp = QQ, Qq, qq \,|\mathbf{y}) = (0.501, 0.75, 0.501)$. The colors red ($QQ$), green ($Qq$) and blue ($qq$) were used to identify individuals and chromosome regions that exceed these thresholds. Note that the information from QTL genotype probabilities could be used to identify individuals of interest for breeding, e.g., passing on favorable alleles to offspring, or further experimentation such as QTL validation studies, e.g., segregating individuals at interesting QTL to be used as parents of new mapping populations.

*Genomic breeding values*

A new feature to the Bayesian multiple QTL analysis is the calculation of breeding values for which all information is implicitly available from the MCMC simulation. We will first calculate breeding values per chromosomal bin (=individual QTL) and then calculate the genomic breeding values by summing over chromosomal bins. The breeding values for individual QTL were calculated from the QTL genotype ($QQ$, $Qq$, $qq$) probability and the additive QTL effect for each chromosomal bin where a QTL was present at iteration $t$, denoted as $I(\lambda_q)^{(t)}$, along the Markov chain Monte Carlo simulation with $N_{\text{smpl}}$ stored samples. The aggregate genomic breeding value for an individual $i$ was calculated by summation of the weighted breeding values over all chromosomal bins, i.e.,

$$gebv_i = \sum_{\text{bin}=1}^{N_{\text{bin}}} \left( {}^{1}/_{N_{\text{smpl}}} \sum_{t=1}^{N_{\text{smpl}}} \sum_{q=1}^{N_{\text{QTL}}} I(\lambda_q)^{(t)} w_{i,q}^{(t)} a_q^{(t)} \right) \qquad (6)$$

with $w_{i,q}$ referring to the $(i,q)$ element of design matrix $\mathbf{W}$ in Eq. (1). The calculation of the genomic breeding values in (6) was identical to Bayesian methods for genomic selection (Meuwissen et al. 2001), i.e., the breeding value includes all (QTL) effects throughout the genome irrespective of their significance. The estimation in (6) was extended to calculate genomic breeding values for specific chromosome segments, i.e., those bins with increased QTL probability

$$gebv_i^{\text{BF}} = \sum_{\text{bin}=1}^{N_{\text{bin}}} \left( {}^{1}/_{N_{\text{smpl}}} \sum_{t=1}^{N_{\text{smpl}}} \sum_{q=1}^{N_{\text{QTL}}} \text{BF}(\lambda_q) I(\lambda_q)^{(t)} w_{i,q}^{(t)} a_q^{(t)} \right) \qquad (7)$$

where $\boldsymbol{BF}(\lambda_q)$ is an indicator variable with value 1 if the positional bin $\lambda_q$ has a Bayes factor exceeding a threshold for positive (>2) or strong (>5) evidence and zero otherwise. Note that the BF indicator variable is a posteriori calculated. Excluding the summation along the genome in (7) allowed the calculation of bin-wise breeding values for those bins with positive QTL evidence

$$gebv_{i,\lambda_q}^{\text{BF}} = {}^{1}/_{N_{\text{smpl}}} \sum_{t=1}^{N_{\text{smpl}}} \text{BF}(\lambda_q) I(\lambda_q)^{(t)} w_{i,q}^{(t)} a_q^{(t)} \qquad (8)$$

These bin-wise breeding values were used to track the most important regions contributing to the individuals' genomic breeding values. Note that the $BF(\lambda)$ imposed an inclusion threshold, while the $I(\lambda)^{(t)}$ invokes a weighting of the bin-wise breeding values contributing to the total genomic breeding values.

The accuracy of prediction was calculated as the correlation between these genomic breeding values and the observed trait values. Daetwyler et al. (2008) showed that this accuracy could be seen as a function of the product of the observed heritability and the ratio of the number of phenotypes to the number of loci involved; the accuracy will exceed the heritability when the number of phenotypes is relatively large. Note that no cross-validation was pursued to thoroughly evaluate the accuracy of prediction as the primary aim of this study is QTL analysis.

## Results

### Heritability estimates

The heritability estimates for the QTL models (1) ranged from 0.62 up to 0.67 with increasing values for higher values of $E(N_{\text{QTL}})$ (Table 1) where the infinitesimal polygenic models (*TIM_R0* and *TIM_R1*—used for benchmarking) yielded a similar level of estimated heritability, i.e., 0.64. The highest estimates for heritability (0.72) were realized from the models with both QTL and polygenes. In these later models, there was an increasing trend in heritability estimates due to QTL with increasing values for higher values of $E(N_{\text{QTL}})$, and a simultaneous decreasing contribution of polygenic effects. The estimated posterior standard deviations were always relatively small for residual variance ($\leq$0.10) and larger for QTL variance ($\geq$0.25), especially for models that included a polygenic component.

### Number of QTL

For the inference on the number of QTL, we distinguish QTL that have strong to decisive evidence across all models (2lnBF $\geq$ 5), and QTL that have indicative evidence (2lnBF between 2 and 5). The chromosome-wide $2\ln\text{BF}_{10}$ values (evidence for one QTL versus none) showed mostly consistent patterns across the genetic models (Table 2) with clear examples of decisive QTL evidence on chromosomes 1 and 10 with very high values (around 30). Also for chromosome

**Table 1** Estimated posterior mean (p.m.) and standard deviation (p.s.d.) of variance components due to error ($\sigma^2_e$), QTL($\sigma^2_{QTL}$), polygenes ($\sigma^2_u$), heritability ($h^2$) and prediction accuracy for three additive genetic models (QTL [Q], QTL + TIM [TIMQ] and TIM) with alternative a priori Poisson distributions for the number of QTL, i.e., E($N_{QTL}$) = {1, 5, 10}

Each combination had two replicates indicated by R0 and R1 (differing in starting seeds for the MCMC simulation). The phenotypic variance was equal to 2.80

| | $\sigma^2_{QTL}$ | | $\sigma^2_u$ | | $\sigma^2_e$ | | $h^2$ | |
|---|---|---|---|---|---|---|---|---|
| | p.m. | p.s.d. | p.m. | p.s.d. | p.m. | p.s.d. | Model | QTL |
| Q1_R0 | 2.37 | 0.34 | | | 1.06 | 0.07 | 0.62 | 0.62 |
| Q1_R1 | 1.97 | 0.25 | | | 1.03 | 0.06 | 0.63 | 0.63 |
| Q5_R0 | 2.16 | 0.29 | | | 0.97 | 0.07 | 0.65 | 0.65 |
| Q5_R1 | 2.08 | 0.29 | | | 0.96 | 0.07 | 0.66 | 0.66 |
| Q10_R0 | 2.14 | 0.28 | | | 0.97 | 0.07 | 0.66 | 0.66 |
| Q10_R1 | 2.27 | 0.30 | | | 0.92 | 0.07 | 0.67 | 0.67 |
| Mean | 2.16 | 0.29 | | | 0.99 | 0.07 | 0.65 | 0.65 |
| TIMQ1_R0 | 2.00 | 0.35 | 0.73 | 0.25 | 0.82 | 0.10 | 0.71 | 0.45 |
| TIMQ1_R1 | 1.77 | 0.30 | 0.72 | 0.24 | 0.81 | 0.10 | 0.71 | 0.46 |
| TIMQ5_R0 | 2.01 | 0.30 | 0.48 | 0.21 | 0.82 | 0.10 | 0.71 | 0.54 |
| TIMQ5_R1 | 2.09 | 0.32 | 0.56 | 0.21 | 0.80 | 0.09 | 0.72 | 0.52 |
| TIMQ10_R0 | 2.13 | 0.32 | 0.54 | 0.22 | 0.81 | 0.10 | 0.71 | 0.52 |
| TIMQ10_R1 | 2.15 | 0.31 | 0.42 | 0.19 | 0.78 | 0.09 | 0.72 | 0.57 |
| Mean | 2.03 | 0.32 | 0.57 | 0.22 | 0.81 | 0.10 | 0.71 | 0.51 |
| TIM_R0 | | | 1.68 | 0.25 | 1.01 | 0.13 | 0.64 | |
| TIM_R1 | | | 1.67 | 0.25 | 1.02 | 0.13 | 0.64 | |

**Table 2** Chromosome-wise 2lnBF values for the $N_{QTL}$ model versus $N_{QTL}$-1 model ($N_{QTL}$ = 1, 2, or 3) for those chromosomes with increased posterior QTL evidence

| | 1 QTL chromosome | | | | | | | | 2 QTL chromosome | | | | | 3 QTL chromosome |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 3 | 6 | 8 | 10 | 14 | 15 | 16 | 1 | 3 | 6 | 10 | 16 | 10 |
| Q1_R0 | **34** | **8** | **33** | 3 | **32** | 4 | **13** | **7** | 2 | **6** | 1 | **5** | 4 | 2 |
| Q1_R1 | **34** | **31** | **33** | 2 | **32** | **5** | **12** | **8** | 3 | **9** | 2 | 4 | 3 | **5** |
| Q5_R0 | **30** | **8** | **12** | 4 | **26** | 4 | **11** | **6** | 3 | **5** | 4 | **5** | 2 | 2 |
| Q5_R1 | **30** | **13** | **29** | 3 | **24** | 4 | **15** | **5** | 3 | **6** | 2 | **7** | 1 | 3 |
| Q10_R0 | **30** | **11** | **28** | 4 | **26** | 4 | **9** | **6** | 3 | **5** | 4 | **5** | 2 | 3 |
| Q10_R1 | **28** | **8** | **26** | 4 | **23** | 4 | **11** | **5** | 3 | 4 | 4 | **5** | 2 | 3 |
| Mean | **31** | **13** | **27** | 3 | **27** | 4 | **12** | **6** | 3 | **6** | 3 | **5** | 2 | 3 |
| TIMQ1_R0 | **33** | 2 | **6** | 0 | **32** | −1 | **17** | **32** | **5** | −4 | −1 | 4 | **7** | 0 |
| TIMQ1_R1 | **33** | **7** | **12** | 3 | **31** | 4 | **14** | 4 | **5** | 3 | 5 | **7** | 2 | 4 |
| TIMQ5_R0 | **30** | **5** | **8** | 3 | **27** | 3 | **10** | **6** | 3 | 3 | 2 | **5** | 3 | 2 |
| TIMQ5_R1 | **30** | 4 | **7** | 2 | **27** | 2 | **13** | **6** | 2 | 1 | 3 | 4 | **5** | 2 |
| TIMQ10_R0 | **30** | 4 | **8** | 2 | **28** | 1 | **15** | **12** | 2 | 1 | 1 | 2 | **7** | 1 |
| TIMQ10_R1 | **28** | **6** | **9** | 3 | **24** | 3 | **13** | **5** | 2 | 3 | 2 | **5** | 3 | 2 |
| Mean | **31** | 4 | **8** | 2 | **28** | 2 | **14** | **11** | 3 | 1 | 2 | 4 | 4 | 2 |

The 2lnBF values may be interpreted as non-significant (0–2); positive (2–5); strong (5–10); decisive (>10) (cf. Kass and Raftery 1995). Note that the negative values indicate the favoring of the sparser QTL model. Values ≥5 are printed in bold

Each combination had two replicates indicated by R0 and R1 (differing in starting seeds for the MCMC simulation)

Values are given for two additive genetic models (QTL [Q] and QTL + TIM [TIMQ]) and alternative a priori Poisson distributions for the number of QTL, i.e., E($N_{QTL}$) = (1, 5, 10)

15, we found consistent decisive evidence (around 13) for harboring 1 QTL. Three chromosomes (3, 6, and 16) showed a clear bimodality of the $2\ln BF_{10}$ values for the models excluding and including a polygenic component. Chromosome 6 had only three markers at fairly large distances which could be the cause of the limited power to distinguish the QTL signal from polygenic effects. A similar phenomenon was observed for chromosome 3, although the variation among the models excluding a polygenic component was larger. Also this chromosome suffered from limited marker information, as only

**Table 3** Comparison of QTL for fruit firmness after 2 months of storage with strong or positive evidence based on the mean values of test statistic 2lnBF in Table 2 and the QTL intensity profiles of Fig. 2, to previously reported QTL on fruit firmness

| Study | Mapping population | Size[a] | $n^{b}_{years}$ | Strong evidence (2lnBF > 5) | | | | | | Positive evidence (2lnBF > 2) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Chromosome | | | | | | Chromosome | | | | | | | |
| | | | | 1b | 3b | 6a | 10b | 15 | 16a | 1a | 3a | 6b | 8 | 10a | 10c | 14 | 16b |
| *Chromosome position (cM)* | | | | *56* | *66* | *32* | *48* | *19* | *19* | *32* | *26* | *54* | *25* | *28* | *70* | *15* | *63* |
| *Cumulative genome position (cM)* | | | | *31* | *169* | *379* | *565* | *895* | *1005* | *7* | *129* | *401* | *443* | *545* | *587* | *855* | *1049* |
| HiDRAS | 27 families | | 3 | 18 | 5 | 7 | 14 | 4 | 6 | 1 | 7 | 6 | 3 | 3 | 4 | 7 | 1 |
| King et al. (2000)[b] | Prima × Fiesta | 152 | 2 | P | | | | P | | | | | p | | P | | |
| Liebhard et al. (2003)[c] | Fiesta × Discovery | 300 | 2 | | D | f, d | | | | | | | | | | f | |
| Kenis et al. (2008) | Telamon × Braeburn | 165 | 2 | | | | T, B | | t, b | | | | | | | t | |
| Costa et al. (2010) | Fuji × Gala | 176 | 2 | | | | G | | | | | | | | | | |

The most probable QTL positions (in italic) are provided on chromosome scale and cummulative genome scale

For the HiDRAS population, we state the map positions and the number of segregating parents (for model "Q5_R0"). For the previous studies, we state the names of the two parents, the size of the family and the number of years with phenotypic data (when underlined the average is taken) of the mapping population. The initial of the mapping parent that was segregating for the QTL is given in capital (lowercase) when the reported QTL was strongly (weakly) significant

[a] The number of progeny with phenotypic data may be less and differ among years

[b] The same population and QTL was reported by Maliepaard et al. (2001)

[c] Fruit firmness was assessed at harvest

four markers were available with a large marker gap at the end. Chromosome 16 showed variation in $2\ln BF_{10}$ values, but the higher values arose from models including a polygenic component. Chromosomes 8 and 14 had consistently positive, but limited evidence ($2\ln BF_{10} = 2.8$ and 3.1, respectively) for QTL presence. Nine chromosomes (2, 4, 5, 7, 9, 11, 12, 13, and 17) always had low values, i.e., near or below 0 indicating that the data did not provide evidence for segregating QTL on these chromosomes (excluded in Table 2).

The chromosome-wide $2\ln BF_{21}$ values (evidence for 2 QTL versus 1 QTL) indicated consistent positive evidence for two chromosomes (1 and 10) and variable positive evidence for three chromosomes (3, 6, and 16, Table 2). A bimodality in the mean of the $2\ln BF_{21}$ values was observed for the models excluding and including polygenic effects for chromosomes 3 and 6 (less strong)—similar to the bimodality in $2\ln BF_{10}$ values. Nevertheless, the $2\ln BF_{21}$ value for chromosome 6 always passed the threshold for strong evidence. With one exception, the $2\ln BF_{21}$ values indicated positive evidence for 2 QTL for chromosome 16. Chromosome 10 was the only one with positive evidence for a third QTL.

Based on the chromosome-wide Bayes factors, we postulated 14 QTL for fruit firmness on 8 chromosomes (see also Table 3). The next step was to identify the most likely QTL positions on these chromosomes.

## QTL intensity map profiles

A high level of consistency was present among the QTL intensity profiles for the six different models (Fig. 2, showing one replicate per model). That is, the chromosome regions with positive evidence of QTL presence are very similar independent of the inclusion or exclusion of a polygenic component or the mean of prior distribution on $N_{QTL}$. Two exceptions are the QTL peak on chromosome 5 in the model Q1_R0 and the absence of a QTL peak on chromosome 3 in the model TIMQ1_R0. The QTL peak at the end of chromosome 16 was only present in the models including a polygenic component. The regions comprising consecutive bins with posterior intensity exceeding the posterior probability threshold (corresponding with $2\ln BF \geq 2$) were indicated by gray color filling of the profile. Note that the number of peaks on a particular chromosome might not be indicative of the number of QTL (see below). Based on the 2lnBF values of Table 2 and the QTL intensity profiles in Fig. 2, we reported in Table 3 six QTL regions with strong evidence and eight regions with positive evidence. The reported QTL positions were the map positions with the highest probability (=posterior modes) within the map-credible regions in the model Q5_R0. The length of these map-credible regions varied from 12 cM (chromosome 1) to 52 cM (chromosome 3).

## QTL effects

Figure 3 shows the 2 cM bin-wise estimated mean and 90 % credible regions for QTL effects for those chromosomal regions exceeding the posterior thresholds for QTL presence in the models Q5_R0 (other models yielded similar estimates, not shown). The QTL on chromosomes 3 and

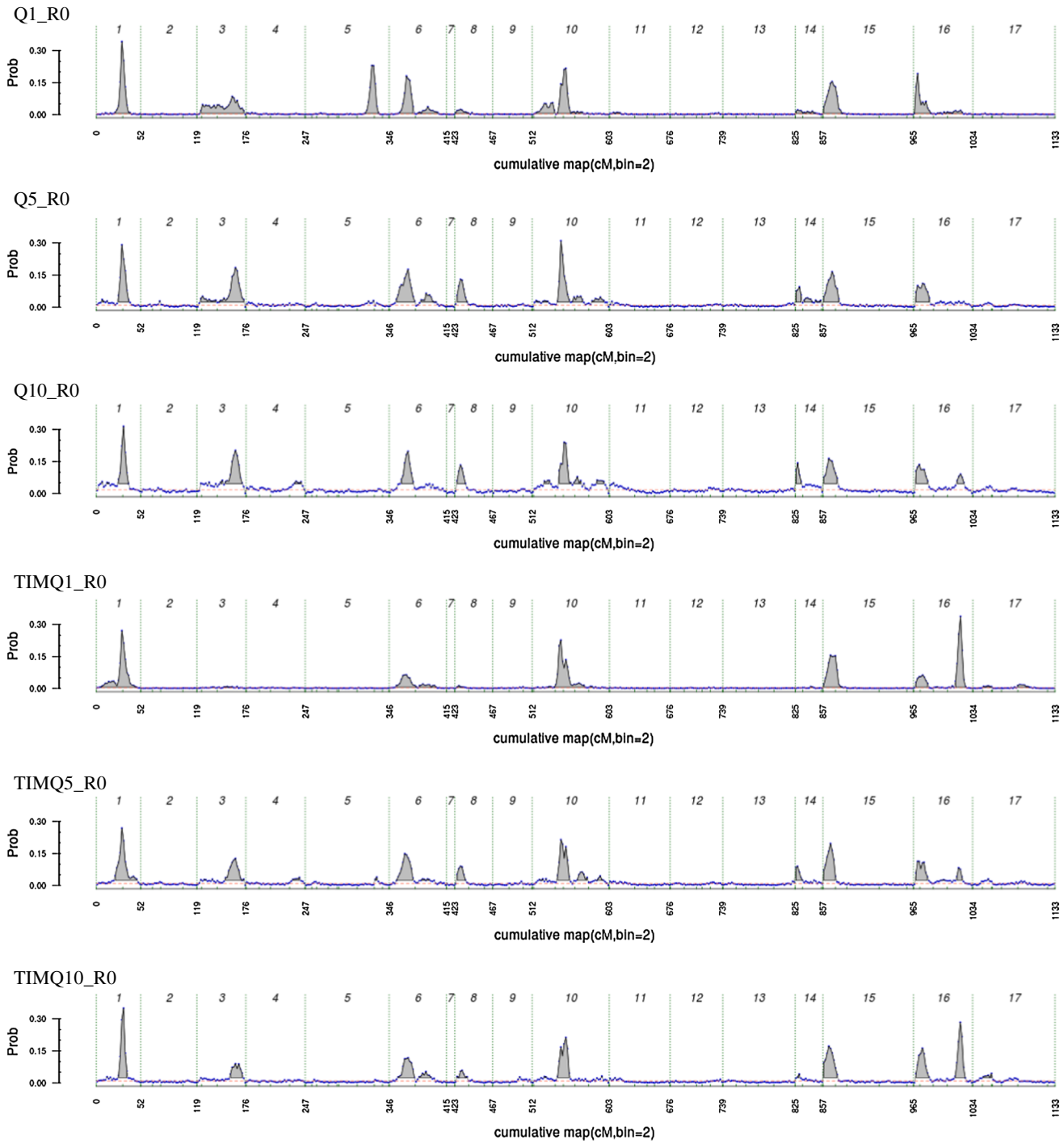



**Fig. 2** Posterior probability of QTL positions (*blue profile line*) along the genome for additive genetic models (QTL and QTL + TIM) with alternative a priori Poisson distributions for the number of QTL [$E(N_{QTL}) = 1,5$, and 10]. The starts and ends of chromosomes are indicated by *dashed vertical lines*, and marker positions are indicated by *inner ticks* at the horizontal axis. The *red dashed horizontal line* indicates the bin-wise prior probability for QTL position (differs with values of [$E(N_{QTL})$]). The *filled gray areas* correspond with regions with positive evidence ($2\ln BF_{10} > 2$) for QTL presence (color figure online)

6 had the highest mean estimates for QTL additive effects, i.e., 1.2 and 1.4, respectively. However, the associated credible regions were also very large, i.e., [0.6, 1.7] and [0.5, 2.2], respectively, which indicated a relatively low accuracy of the mean estimates. A large effect was also observed for the QTL at the end of chromosome 16 which was only present in the models including a polygenic component (not shown). On the other hand, the QTL on chromosomes 1,

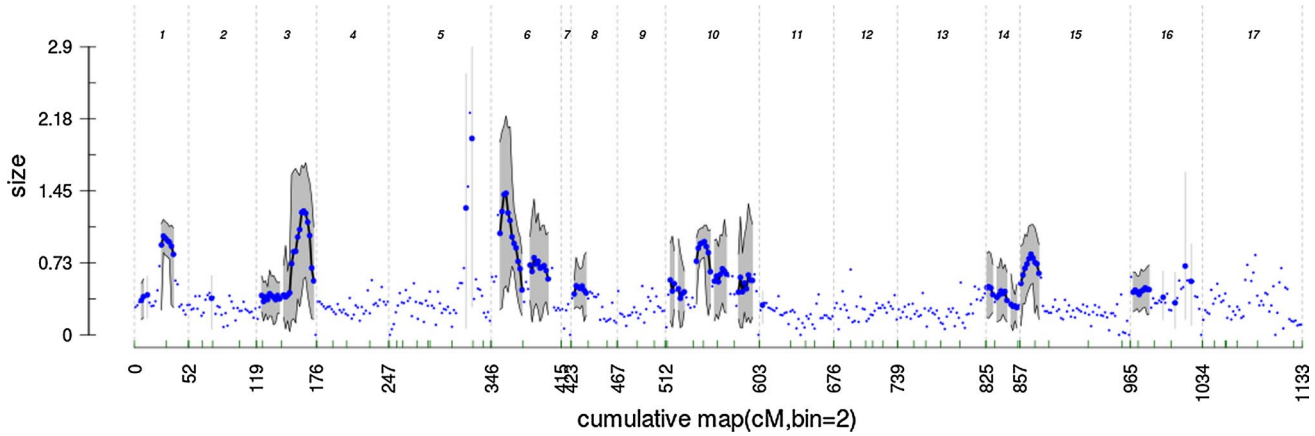



**Fig. 3** Posterior mean (*blue dots*) and 90 % credible region (*gray surfaces*) estimates (per 2 cM bins) of additive QTL effects along the genome for the additive QTL model with the Poisson prior distribution $E(N_{QTL}) = 5$ ("Q5_R0"). The 90 % credible region are bounded by the 5 and 95 % quantiles and are plotted only for those chromosomal 2 cM bins with positive evidence ($2\ln BF_{10} \geq 2$) for QTL presence (as presented in Fig. 2). The starts and ends of chromosomes are indicated by *dashed vertical lines* and marker positions are indicated by *inner ticks* at the *horizontal axis* (color figure online)

10 (middle) and 15 were moderate in size, but the corresponding credible regions were relatively small indicating higher accuracy of the estimated effect size, e.g., the QTL on chromosome 1 had an estimated effect of 1.0 with a credible region equal to [0.8, 1.1].

## QTL genotype probabilities

The data comprised 27 FS families derived from crosses between 33 parents; evidence of QTL presence is obtained from segregation in one or multiple families. Consequently, the inferred QTL genotype probabilities of these 33 parents will most directly show which FS families contributed to the detection, mapping and quantification of segregating QTL for fruit firmness. For all regions with QTL evidence, multiple parents appeared to be heterozygous and generate segregating families as indicated by the green bars in Fig. 4. However, the number of heterozygous parents varied from few, e.g., chromosome 15 and end of chromosome 3, to many, e.g., chromosome 1 and chromosome 10. Also, for several QTL regions, insufficient evidence was present to infer the QTL genotypes, e.g., chromosomes 8 and 16. For the QTL on the bottom of chromosome 10 and the top of chromosome 15, many FS parents were assigned the homozygous *QQ* genotype (red bars in Fig. 4), indicating that the frequency of the allele increasing the phenotypic value is already very high. Conversely, mostly homozygous *qq* genotypes were assigned for the QTL at the end of chromosome 3 and the QTL close to the center of chromosome 10. Along the 'individuals' dimension, large variation was also present. For example, 'Discovery' had heterozygous *Qq* genotypes assigned for many regions (chromosomes 1, 3, 6, 10, and 15), while parent 'Alwa' had no heterozygous *Qq* genotypes assigned at all.

## Genomic breeding values

The estimated genome-wide bin-wise breeding values of the 33 parents indicate that five regions on chromosomes 1, 3, 6, 10 and 15 contributed most to the individuals' genomic breeding values (Fig. 4b). The majority of parents had negative and positive breeding values for chromosomes 3 and 15, respectively. Along the genome, the parents 'Prima' and 'Discovery' had breeding values near zero, partly because these parents were heterozygous for the important QTL. The (total) genomic breeding values of the parents ranged from −2.8 ('Rubinette') to 1.7 ('X-3318').

The total genomic breeding values were also estimated for all other individuals and on a population level these values showed high correlation (0.90) with the observed phenotypes. However, there was substantial variation among the correlations of the 27 FS families for all QTL models (model Q5_R0 given in Fig. 5). Relatively low correlation values pertained to families 'Rubinette' × 'X-3305' (0.57), 'Ligol' × 'Alwa' (0.61), 'Alwa' × 'Pinova' (0.64) and 'Pinova' × 'Gala' (0.65), which may indicate the presence of additional undiscovered QTL or the presence of non-additive effects of the discovered QTL, while high values were obtained for families 'Discovery' × 'Prima' (0.94), 'RedwinterX3177' × 'Galarina' (0.93), 'X-3318' × 'Galarina' (0.92), and 'X-3318' × 'X6564' (0.92). These correlations align well with the number of QTL that were segregating in the parents of the families [Fig. 4, panel (A)], for example, parents 'Discovery' and 'Prima' were heterozygous for at least five QTLs. Also, there was no positive evidence for 'Alwa' and 'Ligol' to be heterozygous at the identified QTL; however, the genotype assignment was undecided for several QTL regions (allowing parents to

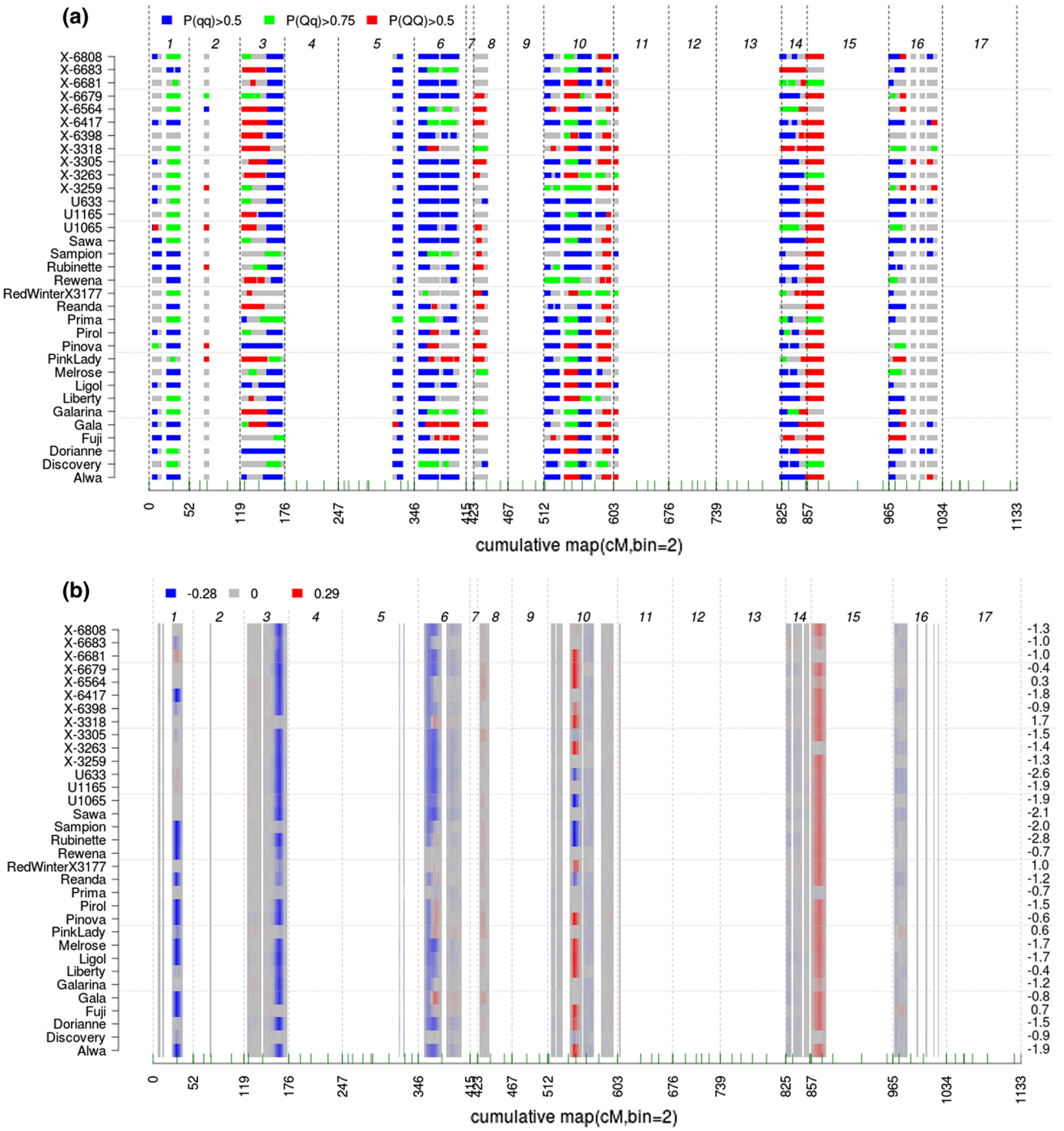
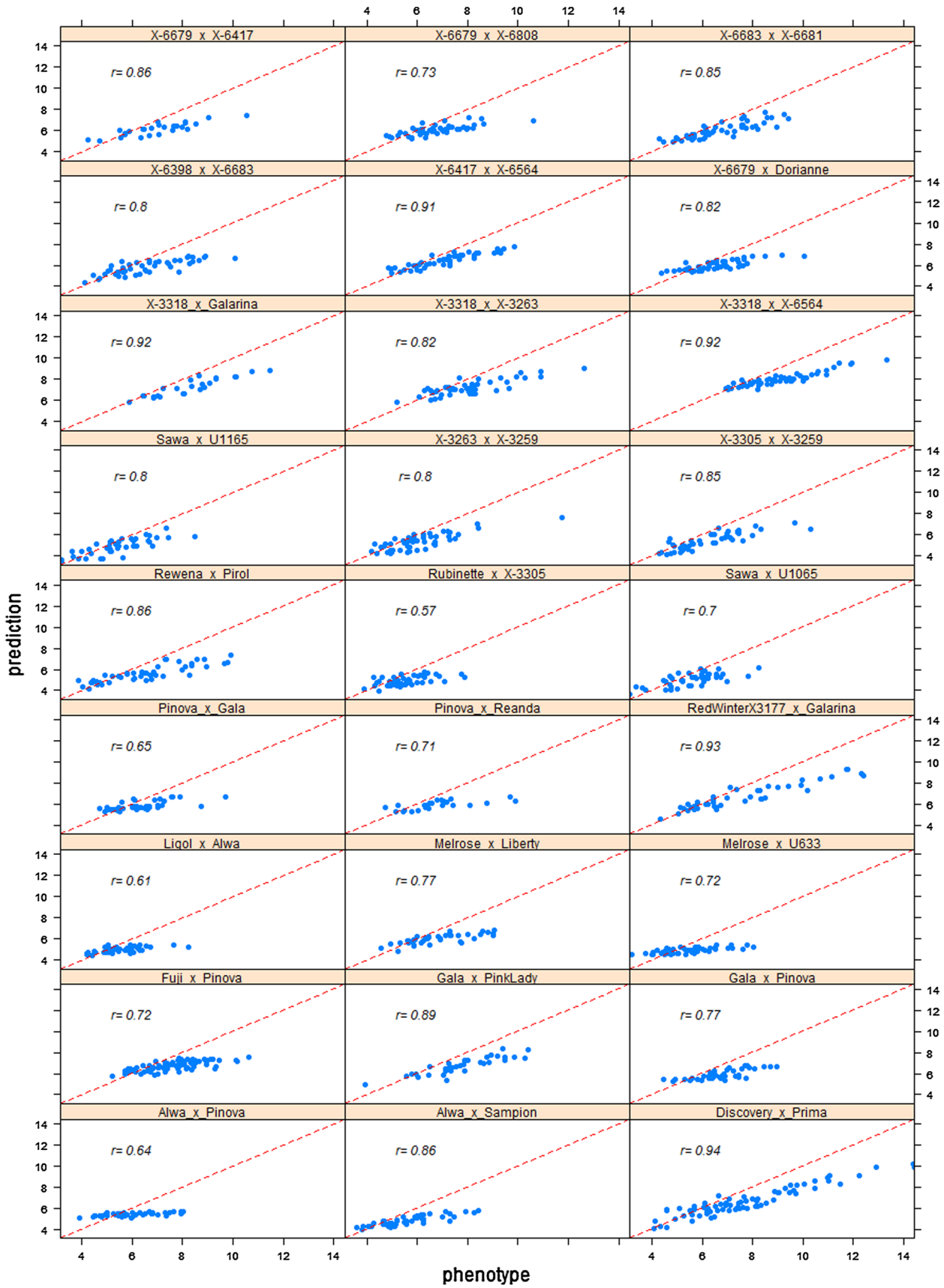



**Fig. 4** Posterior estimates of QTL genotype probabilities [*top panel* **(a)**] and genome-wide binned breeding values [*bottom panel* **(b)**] for the 33 parents of the full sib families for the additive QTL model with the Poisson prior distribution $E(N_{QTL}) = 5$ ("Q5_R0"). Estimates are plotted for those chromosome regions with positive evidence ($2\ln BF_{10} \geq 2$) for QTL presence (as presented in Fig. 2). The starts and ends of chromosomes are indicated by *dashed vertical lines*, and marker positions are indicated by *inner ticks* at the *horizontal axis*. In the *top panel*, the *blue*, *green* and *red colors* indicate positive evidence for QTL genotypes $qq(--)$, $Qq(+-)$ and $QQ(++)$, respectively (the *gray color* indicates ignorable evidence for any genotype). In the *bottom panel*, a gradual coloring intensity in *blue*, *gray* and *red colors* indicate the negative, intermediate and positive breeding values, and the accumulated (genomic) breeding values are printed on the right hand side (color figure online)


X-6679 x X-6417
r= 0.86
X-6679 x X-6808
r= 0.73
X-6683 x X-6681
r= 0.85
X-6398 x X-6683
r= 0.8
X-6417 x X-6564
r= 0.91
X-6679 x Dorianne
r= 0.82
X-3318_x_Galarina
r= 0.92
X-3318_x_X-3263
r= 0.82
X-3318_x_X-6564
r= 0.92
Sawa x U1165
r= 0.8
X-3263 x X-3259
r= 0.8
X-3305 x X-3259
r= 0.85
Rewena x Pirol
r= 0.86
Rubinette x X-3305
r= 0.57
Sawa x U1065
r= 0.7
Pinova_x_Gala
r= 0.65
Pinova_x_Reanda
r= 0.71
RedWinterX3177_x_Galarina
r= 0.93
Ligol x Alwa
r= 0.61
Melrose x Liberty
r= 0.77
Melrose x U633
r= 0.72
Fuji x Pinova
r= 0.72
Gala x PinkLady
r= 0.89
Gala x Pinova
r= 0.77
Alwa_x_Pinova
r= 0.64
Alwa_x_Sampion
r= 0.86
Discovery_x_Prima
r= 0.94
prediction
phenotype
4
6
8
10
12
14

◀ **Fig. 5** Correlation ($r$) between observed phenotypes and predicted genomic breeding values for the 27 full sib families for the additive QTL model with the Poisson prior distribution $E(N_{QTL}) = 5$ ("Q5_R0"). The phenotypic values are along the $x$-axis to ease comparison with the phenotypic histograms (Fig. 1)

be segregating with low probabilities). Consequently, the accuracy of the genomic breeding values of their full sib progeny was still moderate.

## Discussion

In this paper we present the statistical methodology and application to the use of multiple pedigreed FS families for the genetic dissection of complex traits of outcrossing species. We describe the Bayesian approach for pedigree-based QTL analysis to discover and characterize multiple QTL, while a more detailed description on the modeling and prior assumptions has been previously presented (Bink et al. 2008b). Here, we provide additional guidelines to make statistical inferences on the number of QTL per chromosome, the QTL effect sizes and the assignment of QTL genotypes to individuals. Finally, we show that the Bayesian method implicitly yields estimates of individuals' genome-wide breeding values.

### Application: proof of concept and statistical methodology

We applied the Bayesian QTL approach to study the complex trait fruit firmness of apple after 2 months of cold storage in the HiDRAS dataset. The estimated Bayes factors for the number of QTL per chromosome combined with the robustness of QTL intensity plots across the studied models indicated positive or strong evidence for a total of 14 QTL (Tables 2, 3). The six QTL with strong evidence have been reported in previous studies on fruit firmness, although evidence was sometimes less significant (Table 3). These previous studies comprised single mapping populations and yielded a total of five strongly significant QTL. Four of these are confirmed in our study (chromosomes 1b, 3b, 10b and 15), while the fifth was less supported here (chromosome 10c). This latter QTL is known to be expressed at a late stage of maturation, following the then diminishing expression of the QTL on chromosome 10b (Costa et al., 2010). Thus, the poor support of the QTL on chromosome 10c in our study may thus derive from having less aged fruits and its detection might improve for instance with more appropriate harvest dates and storage conditions. Our study revealed five QTL that were not reported previously, while none of the previously reported QTL were missed in our study, i.e., no false-negative results. Furthermore, there was complete consistency in the parents of the single mapping populations that were segregating for the QTL. For example, 'Discovery', 'Gala' and 'Prima' were parents in our study and were heterozygous with high probability (Fig. 4) for all QTL that were reported in previous studies.

### Advantages to breeders in the use of multiple families

Using a population of multiple families has several advantages over single mapping populations. Firstly, this increases the chance of having a good representation of available relevant QTL and QTL alleles. Secondly, this combines the power of QTL detection and QTL mapping accuracy, since QTL often segregate in several families. The power to detect QTL is also boosted in our approach as the search is for multiple QTL simultaneously; after detection and accounting for variance explained by the larger QTL, the remaining residual variance is reduced so that smaller QTL can also be detected. All these factors may jointly explain why this study revealed a substantial number of additional QTL for fruit firmness over the previous studies. For example, we were able to detect two QTL on chromosome 6 for which 'Discovery' was segregating, while only one QTL was previously reported on this chromosome (Liebhard et al. 2003). Similarly, in our study, we revealed evidence for segregating QTL in cultivar 'Prima' on chromosomes 1b, 3b, 6a, 10b and 15 (Fig. 4), three of which were additional to the previously reported QTL for this parent (Table 3) (King et al. 2000; Maliepaard et al. 2001). Thirdly, the Bayesian approach explicitly models the pedigree structure among known common ancestors of the mapping populations and trace segregating QTL alleles back to the common ancestor (founder), and to other related genotyped individuals which is highly helpful to select potential parents from candidates in germplasm that was not part of the original mapping populations (Online Resource 4). The latter requires densely spaced markers to obtain IBD probabilities for QTL in the candidates. Fourthly, the use of multiple families provides insight into the QTL contribution across different genetic backgrounds. Knowing the QTL effects and their variances across a wider genetic background sustains the prediction and use of breeding values in breeding programs.

An important factor in the prioritizing of which QTL to develop easy-to-apply markers for will be the frequency of the favorable QTL allele in the breeding germplasm. Suppose the 33 parents in this study are representative of a breeding program, then the QTL on chromosome 15 may be of less interest as the majority of parents are already homozygous for the QTL allele increasing fruit firmness (still, it would be important to have one or both parents of a particular cross carrying the favorable allele). Conversely, the QTL on chromosome 1b or 10b segregate in many

parents and thus we expect marker-assisted selection to be more rewarding.

### Utilization in breeding programs

Our Bayesian QTL approach also yielded genomic estimated breeding values (GEBV) [Figs. 4, 5, using Eqs. (6) and (7)] which may be used in breeding programs similarly to those obtained from genomic selection methods (Kumar et al. 2012). In that case there is no particular interest in the underlying model variables (either QTL or markers), but selection is solely based on the accumulated sum of effects (Meuwissen et al. 2001). Our Bayesian QTL methodology shares a common feature with the Bayesian variable selection methods for genomic prediction, e.g., BayesB (Meuwissen et al. 2001) and BayesC (Habier et al. 2011) since they all assume a mixture of two prior distributions of allelic effects, i.e., one with substantial effects and another that harnesses the effects that are very close or equal to zero. However, a major difference is that our method considers QTL as factors in the model, while the other genomic selection methods take (SNP) markers as explanatory variables. Here, we did not apply cross-validation to assess the prediction accuracy, and the reported accuracy is actually the within the data fit of the QTL-based genomic breeding values. It was not straightforward to apply Bayesian genomic prediction methods as the marker data comprised multi-allelic SSR markers. Instead, we applied genomic prediction (without cross-validation) via G-BLUP breeding values, after obtaining the marker-based relationship matrices from six different relatedness estimators (Bink et al. 2008a). The resulting prediction accuracies ranged from 0.72 up to 0.77, which was clearly lower than those obtained in our Bayesian QTL analyses and was likely due to the very sparse marker density of the dataset.

Alternatively for selection based on GEBV, breeders can capitalize on the discovered QTL through marker-assisted breeding. This requires the inference of the marker haplotypes flanking the favorable QTL alleles, and these marker haplotypes may vary among different FS families. However, having accurate ancestral pedigree and marker data on the FS families will allow favorable QTL alleles to be traced to one or multiple common founders through identity by descent probability as to any genotyped relative (exemplified in Online Resource 4).

### Prospects

In this study the marker coverage and density were limited and this hampered the accurate assignment of QTL genotypes to individuals (i.e., the gray lines in Fig. 4) in the additive models of this study. The restrictions due to marker data and low representation for some QTL sources (founders with few descendants) were reasons to include polygenic effects that may account for QTL in regions with no or limited marker data or of limited representation. Inclusion of polygenic effects did affect the evidence for some QTL, but not for others (Table 2). Remarkably, when fitting polygenic effects the evidence for a QTL at the end of chromosome 16 was increased (Fig. 2). Our Bayesian modeling assumed two alleles per QTL, but the number of alleles could be modeled as a random variable; however, accurate inferences require large and designed datasets [e.g., (Jannink and Wu 2003)]. Further extensions to non-additive effects such as dominance and epistasis or QTL by genetic background interactions (Jansen et al. 2009) will only be meaningful after the genome has been saturated with many more markers using SNP-array genotyping (Chagne et al. 2012; Ganal et al. 2011; Verde et al. 2012; Tung et al. 2010). Increasing the quantity of the marker data may also allow a direct inference of the number of QTL on the genome level, which was indecisive (probabilities and Bayes factors are provided in the Online Resource 5) and the inferences were presented via Bayes factors on a chromosome level. As part of the European FruitBreedomics project (http://fruitbreedomics.com), the germplasm of the current paper will soon be genotyped with a commercially available Infinium 20K SNP array.

## Conclusion

We deliver proof of concept of a Bayesian integrated QTL analysis across pedigree-related families from ongoing outbred breeding programs. The approach supports efficient QTL discovery and characterization in terms of magnitude and prevalence and makes these emerging QTL results immediately applicable in breeding decisions. Its application may considerably accelerate the use of marker-assisted breeding for the improvement of quantitative traits.

**Acknowledgments** This study has been carried out with financial support from the Commission of the European Communities, specific research program "Quality of Life and Management of Living Resource", QLK5-2002-01492 "High-quality Disease Resistant Apples for a Sustainable Agriculture", coordinated by L. Gianfranceschi from the University of Milan (It.). The first author and last author also acknowledge the support from the U.S. Department of Agriculture's National Institute of Food and Agriculture–Specialty Crops Research Initiative project, "RosBREED: Enabling marker-assisted breeding in Rosaceae" (2009-51181-05808) and the European Union-funded project "FruitBreedomics: Integrated approach for increasing breeding efficiency in fruit tree crops" (Grant #FP7- 265582; http://fruitbreedomics.com/). The first author also acknowledges the financial support from Kennisbasis project KB-17-003.01-002 Genomic breeding tools and databases. The INRA experimental unit UE-Horti is also especially thanked for strong support in orchard management and fruit harvest.

**Conflict of interest** The authors declare that they have no conflict of interest.

## References


Antofie A, Lateur M, Oger R, Patocchi A, Durel CE, Van de Weg WE (2007) A new versatile database created for geneticists and breeders to link molecular and phenotypic data in perennial crops: the AppleBreed database. Bioinformatics 23(7):882–891

Bink MCAM, Van Arendonk JA (1999) Detection of quantitative trait loci in outbred populations with incomplete marker data. Genetics 151(1):409–420

Bink MCAM, Uimari P, Sillanpaa J, Janss LLG, Jansen RC (2002) Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor Appl Genet 104(5):751–762

Bink MCAM, Anderson AD, van de Weg WE, Thompson EA (2008a) Comparison of marker-based pairwise relatedness estimators on a pedigreed plant population. Theor Appl Genet. doi:10.1007/s00122-008-0824-1

Bink MCAM, Boer MP, ter Braak CJF, Jansen J, Voorrips RE, Weg WEVD (2008b) Bayesian analysis of complex traits in pedigreed plant populations. Euphytica 161(1–2):85–96

Bink MCAM, Totir LR, ter Braak CJF, Winkler CR, Boer MP, Smith OS (2012) QTL linkage analysis of connected populations using ancestral marker and pedigree information. Theor Appl Genet 124(6):1097–1113. doi:10.1007/s00122-011-1772-8

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theor Appl Genet 113(2):206–224

Brooks SP, Giudici P, Philippe A (2003) Nonparametric convergence assessment for MCMC model selection. J Comput Graph Stat 12(1):1–22

Chagne D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, Lawley C, Vanderzande S, Hellens RP, Kumar S, Cestaro A, Velasco R, Main D, Rees JD, Iezzoni A, Mockler T, Wilhelm L, Van de Weg E, Gardiner SE, Bassil N, Peace C (2012) Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One 7(2):e31745. doi:10.1371/journal.pone.0031745

Costa F, Peace CP, Stella S, Serra S, Musacchi S, Bazzani M, Sansavini S, Van de Weg WE (2010) QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (Malus × domestica Borkh.). J Exp Bot 61(11):3029–3039. doi:10.1093/jxb/erq130

Costa F, Cappellin L, Fontanari M, Longhi S, Guerra W, Magnago P, Gasperi F, Biasioli F (2012) Texture dynamics during postharvest cold storage ripening in apple (Malus × domestica Borkh.). Postharvest Biol Technol 69:54–63. doi:10.1016/j.postharvbio.2012.03.003

Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One 3 (10). doi:10.1371/journal.pone.0003395

Evans KM, Patocchi A, Rezzonico F, Mathis F, Durel CE, Fernández-Fernández F, Boudichevskaia A, Dunemann F, Stankiewicz-Kosyl M, Gianfranceschi L, Komjanc M, Lateur M, Madduri M, Noordijk Y, Weg WE (2011) Genotyping of pedigreed apple breeding material with a genome-covering set of SSRs: trueness-to-type of cultivars and their parentages. Mol Breed 28(4):535–547. doi:10.1007/s11032-010-9502-5

Falconer DS (1989) Introduction to quantitative genetics. Longman, Harlow

Fanizza G, Lamaj F, Costantini L, Chaabane R, Grando MS (2005) QTL analysis for fruit yield components in table grapes (*Vitis vinifera*). Theor Appl Genet 111(4):658–664

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J, Le Paslier M-C, McMullen MD, Montalent P, Rose M, Schön C-C, Sun Q, Walter H, Martin OC, Falque M (2011) A large Maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6(12):e28334

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Stat Sci 7(4):472–475. doi:10.1214/ss/1177011136

Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian data analysis. Texts in statistical science, 2nd edn. Chapman & Hall/CRC, Boca Raton

Genstat Committee (2004) Genstat Release 7.1: the guide to GenStat. VSN International Ltd, Oxford

Gianfranceschi L, Soglio V (2004) The European project HiDRAS: innovative multidisciplinary approaches to breeding high quality disease resistant apples. Acta Horticulturae 663:327–330

Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov chain Monte Carlo in practice. Interdisciplinary statistics. Chapman & Hall, London

Habier D, Fernando R, Kizilkaya K, Garrick D (2011) Extension of the bayesian alphabet for genomic selection. BMC Bioinform 12(1):186

Hayes B, Goddard ME (2001) The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol 33(3):209–229

Heath SC (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am J Hum Genet 61(3):748–760

Huang XQ, Paulo MJ, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA (2011) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proc Natl Acad Sci USA 108(11):4488–4493. doi:10.1073/pnas.1100465108

Jannink JL, Wu XL (2003) Estimating allelic number and identity in state of QTLs in interconnected families. Genet Res 81(2):133–144

Jansen J, Boer MP, Bink MCAM, van de Weg WE (2009) Searching for interacting QTL in related populations of an outbreeding species. Euphytica 166(1):131–144. doi:10.1007/s10681-008-9849-4

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90(430):773–795

Kenis K, Keulemans J, Davey M (2008) Identification and stability of QTLs for fruit quality traits in apple. Tree Genet Genomes 4(4):647–661

King GJ, Maliepaard C, Lynn JR, Alston FH, Durel CE, Evans KM, Griffon B, Laurens F, Manganaris AG, Schrevens T, Tartarini S, Verhaegh J (2000) Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill.). Theor Appl Genet 100(7):1074–1084

Kouassi AB, Durel CE, Costa F, Tartarini S, van de Weg E, Evans K, Fernandez-Fernandez F, Govan C, Boudichevskaja A, Dunemann F, Antofie A, Lateur M, Stankiewicz-Kosyl M, Soska A, Tomala K, Lewandowski M, Rutkovski K, Zurawicz E, Guerra W, Laurens F (2009) Estimation of genetic parameters and prediction of breeding values for apple fruit-quality traits using pedigreed plant material in Europe. Tree Genet Genomes 5(4):659–672. doi:10.1007/s11295-009-0217-x

Kumar S, Bink MCAM, Volz RK, Bus VGM, Chagne D (2012) Towards genomic selection in apple (Malus × domestica Borkh.) breeding programmes: prospects, challenges and strategies. Tree Genet Genomes 8(1):1–14. doi:10.1007/s11295-011-0425-z

Lerceteau-Köhler E, Moing A, Guérin G, Renaud C, Petit A, Rothan C, Denoyes B (2012) Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-QTL in their control. Theor Appl Genet 124(6):1059–1077. doi:10.1007/s00122-011-1769-3

Liebhard R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C (2003) Mapping quantitative physiological traits in apple (Malus × domestica Borkh.). Plant Mol Biol 52(3):511–526

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland

Maliepaard C, Sillanpaa MJ, van Ooijen JW, Jansen RC, Arjas E (2001) Bayesian versus frequentist analysis of multiple quantitative trait loci with an application to an outbred apple cross. Theor Appl Genet 103(8):1243–1253. doi:10.1007/s001220100720

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829

Patocchi A, Fernandez-Fernandez F, Evans K, Gobbin D, Rezzonico F, Boudichevskaia A, Dunemann F, Stankiewicz-Kosyl M, Mathis-Jeanneteau F, Durel CE, Gianfranceschi L, Costa F, Toller C, Cova V, Mott D, Komjanc M, Barbaro E, Kodde L, Rikkerink E, Gessler C, van de Weg WE (2009) Development and test of 21 multiplex PCRs composed of SSRs spanning most of the apple genome. Tree Genet Genomes 5(1):211–223. doi:10.1007/s11295-008-0176-7

Pinto RS, Reynolds MP, Mathews KL, McIntyre CL, Olivares-Villegas JJ, Chapman SC (2010) Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. Theor Appl Genet 121(6):1001–1021. doi:10.1007/s00122-010-1351-4

Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. R News 6:7–11

Quilot B, Wu BH, Kervella J, Genard M, Foulongne M, Moreau K (2004) QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species P-davidiana. Theor Appl Genet 109(4):884–897. doi:10.1007/s00122-004-1703-z

Silfverberg-Dilworth E, Matasci CL, Van de Weg WE, Van Kaauwen MPW, Walser M, Kodde LP, Soglio V, Gianfranceschi L, Durel CE, Costa F, Yamamoto T, Koller B, Gessler C, Patocchi A (2006) Microsatellite markers spanning the apple (Malus × domestica Borkh.) genome. Tree Genet Genomes 2(4):202–224

Sillanpaa MJ, Arjas E (1998) Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics 148(3):1373–1388

Sorensen D, Gianola D (2002) Likelihood, Bayesian, and MCMC methods in quantitative genetics. Statistics for biology and health. Springer-Verlag, New York

Stich B, Mohring J, Piepho H-P, Heckenberger M, Buckler ES, Melchinger AE (2008) Comparison of mixed-model approaches for association mapping. Genetics 178(3):1745–1754. doi:10.1534/genetics.107.079707

Thompson EA (2008) Linkage analysis. In: Balding DJ, Bishop M, Cannings C (eds) Handbook of statistical genetics, 3rd edn. Wiley, Chichester, pp 1141–1167. doi:10.1002/9780470061619.ch33

Tung C-W, Zhao K, Wright M, Ali ML, Jung J, Kimball J, Tyagi W, Thomson M, McNally K, Leung H, Kim H, Ahn S-N, Reynolds A, Scheffler B, Eizenga G, McClung A, Bustamante C, McCouch S (2010) Development of a research platform for dissecting phenotype–genotype associations in rice (*Oryza* spp.) Rice 3 (4):205-217. doi:10.1007/s12284-010-9056-5

Uimari P, Sillanpaa MJ (2001) Bayesian oligogenic analysis of quantitative and qualitative traits in general pedigrees. Genet Epidemiol 21(3):224–242

Verde I, Bassil N, Scalabrin S, Gilmore B, Lawley CT, Gasic K, Micheletti D, Rosyara UR, Cattonaro F, Vendramin E, Main D, Aramini V, Blas AL, Mockler TC, Bryant DW, Wilhelm L, Troggio M, Sosinski B, Aranzana MJ, Arús P, Iezzoni A, Morgante M, Peace C (2012) Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. PLoS One 7(4):e35668

Voorrips RE, Bink MCAM, van de Weg WE (2012) Pedimap: software for the visualization of genetic and phenotypic data in pedigrees. J Hered 103(6):903–907. doi:10.1093/jhered/ess060

Wei JM, Ma FW, Shi SG, Qi XD, Zhu XQ, Yuan JW (2010) Changes and postharvest regulation of activity and gene expression of enzymes related to cell wall degradation in ripening apple fruit. Postharvest Biol Technol 56(2):147–154. doi:10.1016/j.postharvbio.2009.12.003

Yu JM, 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(2):203–208

Yu JM, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178(1):539–551

Zhang GR, Sebolt AM, Sooriyapathirana SS, Wang DC, Bink M, Olmstead JW, Iezzoni AF (2010) Fruit size QTL analysis of an F-1 population derived from a cross between a domesticated sweet cherry cultivar and a wild forest sweet cherry. Tree Genet Genomes 6(1):25–36. doi:10.1007/s11295-009-0225-x