A Mixed-Model Approach to Association Mapping Using Pedigree Information With an Illustration of Resistance to *Phytophthora infestans* in Potato

M. Malosetti,[*,1] **C. G. van der Linden,**[†] **B. Vosman**[†] **and F. A. van Eeuwijk**[*,2]

Laboratory of Plant Breeding, Department of Plant Sciences, Wageningen University, 6700 AJ, Wageningen, The Netherlands and †Department of Biodiversity and Breeding, Plant Research International, Wageningen UR, 6700 AA, Wageningen, The Netherlands

ABSTRACT

Association or linkage disequilibrium (LD)-based mapping strategies are receiving increased attention for the identification of quantitative trait loci (QTL) in plants as an alternative to more traditional, purely linkage-based approaches. An attractive property of association approaches is that they do not require specially designed crosses between inbred parents, but can be applied to collections of genotypes with arbitrary and often unknown relationships between the genotypes. A less obvious additional attractive property is that association approaches offer possibilities for QTL identification in crops with hard to model segregation patterns. The availability of candidate genes and targeted marker systems facilitates association approaches, as will appropriate methods of analysis. We propose an association mapping approach based on mixed models with attention to the incorporation of the relationships between genotypes, whether induced by pedigree, population substructure, or otherwise. Furthermore, we emphasize the need to pay attention to the environmental features of the data as well, *i.e.*, adequate representation of the relations among multiple observations on the same genotypes. We illustrate our modeling approach using 25 years of Dutch national variety list data on late blight resistance in the genetically complex crop of potato. As markers, we used nucleotide binding-site markers, a specific type of marker that targets resistance or resistance-analog genes. To assess the consistency of QTL identified by our mixed-model approach, a second independent data set was analyzed. Two markers were identified that are potentially useful in selection for late blight resistance in potato.

OVER the past decade, marker-assisted selection has become a standard tool in breeding programs (YOUNG 1999; ASINS 2002; DEKKERS and HOSPITAL 2002; COLLARD *et al.* 2005). The selection of genetically superior individuals using molecular marker alleles linked to quantitative trait loci (QTL) affecting trait variation instead of using phenotypic trait values can speed up genetic improvement considerably (HOSPITAL *et al.* 1997). As a first step to marker-assisted selection (MAS), marker loci need to be identified that are linked to QTL. In plants, QTL mapping up until now relied principally on the development of segregating populations derived from crosses between inbred lines (F_2, backcross, recombinant inbred lines, etc.). In such populations, observed linkage disequilibrium (LD) between markers and QTL alleles by the nature of the population must necessarily be the result of physical linkage. Although designed segregating populations are easy to create, they come with a number of disadvantages. First,

the amount of segregating genetic variation within the population is limited, because per locus at most two alleles can segregate (in a diploid species), whereas in the absence of allele polymorphisms between the parents no QTL can be identified. Second, the genetic backgrounds within which mapping studies take place are generally not representative of backgrounds used in elite germplasm (JANNINK *et al.* 2001). Third, the relatively low number of generations after maximum LD, where the maximum LD is reached in the F_1, implies a reduced number of sampled meioses within designed populations (typically a few hundred), leading to relatively long stretches of chromosomes being in LD. Consequently, the characteristic size of confidence intervals for QTL locations is between 10 and 20 cM (DARVASI *et al.* 1993). The recognition of these and other limitations has motivated the interest in alternative mapping strategies popular in human and animal genetics, where these techniques are known as LD mapping and association mapping (LYNCH and WALSH 1998; WALSH 2002), designations we use interchangeably.

For association mapping, the population under study consists of a collection of genotypes, and not a highly designed population of genotypes, and the LD required

[1]*Present address:* Biometris, Department of Plant Sciences, Wageningen University, P.O. Box 100, 6700 AC, Wageningen, The Netherlands.

[2]*Corresponding author:* Biometris, Department of Plant Sciences, Wageningen University, P.O. Box 100, 6700 AC, Wageningen, The Netherlands. E-mail: fred.vaneeuwijk@wur.nl

for finding marker–trait associations comes from un-recorded sources of LD in that population (JANNINK *et al.* 2001). Good examples of populations for use in association mapping in plant genetics are breeding and gene bank collections of cultivars, breeding lines, germ-plasm accessions, etc. Association mapping strategies offer larger flexibility in the choice of genotypes for mapping, which is advantageous from a plant breeding perspective. The genetic variation under study can increase quantitatively, because more than two alleles per locus can be present, and qualitatively, because the QTL alleles can be mapped in germplasm that is directly relevant to breeding programs. Estimates for QTL ef-fects will be more realistic since they are estimated within the relevant genetic background for plant breeders. Moreover, marker–trait associations will result from tight linkage, tighter than in most designed populations, making it possible to map QTL with higher precision. An important asset of association mapping strategies is the straightforward utilization of large amounts of historical phenotypic data that are available for mapping efforts at no or little extra costs. For example, breeders are rou-tinely evaluating their germplasm in the relevant target population of environments, thereby producing a wealth of valuable phenotypic information readily available for QTL detection by association mapping efforts, a pro-cedure referred to as "*in silico* mapping" by PARISSEAUX and BERNARDO (2004).

In association mapping, linkage is not the only cause of LD. When the set of genotypes for inclusion in the mapping study is not homogeneous in the sense of, for example, not being an F_2, recombinant inbred line, or double-haploid population, observed LD could be the consequence of mechanisms other than linkage, such as population admixture, specific mating systems, genetic drift, and selection (JANNINK and WALSH 2002; FLINT-GARCIA *et al.* 2003). Although possibly confusing, the term LD refers strictly to the dependence of alleles at different loci within gametes, which may be a result of physical linkage between the loci, but not necessarily so (JANNINK and WALSH 2002). The success of association mapping efforts depends on the possibilities of separat-ing LD due to linkage from LD due to other causes. Population structure is seen as a second major cause of marker–trait associations in addition to linkage (PRITCHARD *et al.* 2000b; JANNINK *et al.* 2001; FLINT-GARCIA *et al.* 2003). Most association–mapping strate-gies therefore start by first inspecting the population to assess whether groups can be discerned within the population and then testing for marker–trait associa-tion after correction for group effects, *i.e.*, testing for marker–trait association within groups. The principle is that only associations caused by linkage will survive after removing the genetic correlations due to group ef-fects. A popular method to detect population structure is that proposed by PRITCHARD *et al.* (2000b). An imple-mentation of this method is available in the software STRUCTURE (http://pritch.bsd.uchicago.edu/software/structure2_1.html) in which, within a Bayesian frame-work, molecular marker information is used to assign group membership probabilities to the genotypes. The estimated probabilities for group membership can then be used to assign genotypes to groups within which marker–trait associations are tested (REMINGTON *et al.* 2001; SIMKO *et al.* 2004b; SKØT *et al.* 2005). Alternatively, the groups can be integrated as an extra factor or a set of covariables in a statistical model relating phenotype to genotype (THORNSBERRY *et al.* 2001; WILSON *et al.* 2004).

As a valid alternative to the use of STRUCTURE, classical multivariate analysis methods can be used to classify genotypes. In that case a matrix of genetic/genotypic distances is calculated from molecular marker information and used as input for clustering and/or scaling techniques (IVANDIC *et al.* 2002; KRAAKMAN *et al.* 2004). For collections of varieties and breeding lines, genotypic relationships as obtained from the pedigree or from similarities in neutral marker profiles (YU *et al.* 2006) can be translated to distances that are subsequently analyzed by cluster analysis. The groups detected by such a cluster analysis can be interpreted as representing population structure and form an approx-imation to the original relationships between the geno-types as present before the grouping. The identified groups can be used as a kind of correction factor in as-sociation analyses. An example of this approach was given by SIMKO *et al.* (2004a), where pedigree data produced a relationship matrix that was used as input to a cluster analysis with the aim of classifying potato cultivars. When pedigree or neutral marker information is present for a set of genotypes, this information can also be incorporated directly in a QTL mapping analysis in the form of a genotypic relationship matrix that structures the variance–covariance matrix between gen-otypes, without making the detour of using the same matrix for finding groups. For maize, PARISSEAUX and BERNARDO (2004) show how to integrate the pedigree-based relationship matrix into a QTL mapping analysis within a mixed-model framework. In YU *et al.* (2006), published shortly after submission of this article, both a marker-based relationship matrix and a factor repre-senting population structure are included in a mixed model for association analysis of a single trait in a single environment. In this article, we follow PARISSEAUX and BERNARDO (2004) for the direct incorporation of the pedigree relationships in an association analysis for mul-tienvironment data. However, like SIMKO *et al.* (2004a,b) and GEBHARDT *et al.* (2004), we focus our modeling ef-forts on potato.

Potato (*Solanum tuberosum*) is an autotetraploid spe-cies ($2n = 4x = 48$), but the simpler genetics at the diploid level combined with the fact that many resis-tance sources are introgressed from wild relatives have motivated the use of diploid interspecific crosses for

most QTL studies in resistance breeding (COLLINS *et al.* 1999; OBERHAGEMANN *et al.* 1999; EWING *et al.* 2000; VISKER *et al.* 2003). Nevertheless, examples of mapping of resistance QTL on the basis of tetraploid crosses have also been reported in the literature (MEYER *et al.* 1998; BORMANN *et al.* 2004). Although classical QTL mapping approaches at the diploid and the tetraploid level have been proven successful, the use of tetraploid potato cultivars in association mapping has advantages both of testing for QTL at the original ploidy level and of testing within a more representative genetic background for the crop. The first examples of successful association mapping studies in tetraploid potato were presented by SIMKO *et al.* (2004a,b) and GEBHARDT *et al.* (2004). These applications involved disease resistances for which candidate genes were defined. The statistical methodology for assessing marker–trait associations consisted principally of the comparison of trait means for two allele states by some form of *t*-test, either the classical *t*-test or a robust (permutation-based) or nonparametric alternative (Wilcoxon/Mann–Whitney *U*-test). When population structure was detected, the tests were performed within identified groups (SIMKO *et al.* 2004a).

The use of *t*-tests for establishing marker–trait associations imposes restrictions on the types of analyses and inference. In the context of association mapping, it is attractive to use historical collections of phenotypic data, with the archetypical example of such a collection consisting of the series of trials performed to test candidate varieties on value for cultivation and use (VCU), with the ultimate goal of compiling national and recommended lists of varieties. Such historical data sets are highly unbalanced across years as new genotypes are added each year, while others are discarded. A good method for association analysis should be able to combine information across multiple years and to cope with unbalancedness. Similarly, it should be possible to incorporate pedigree information, as for major crops pedigree information is available on most if not all genotypes submitted to national testing authorities investigating VCU.

In this article, we show how powerful and flexible a mixed-model framework can be for association mapping in plant species. The use of mixed models for genetic studies is more or less the standard in animal breeding applications, but it is far less common in plant studies. We illustrate the use of mixed models by analyzing two independent data sets on resistance to late blight, the most important potato disease, caused by the oomycete *Phytophthora infestans*, where the second data set served as an empirical check on the QTL identified in the first set. The objective was to identify markers linked to disease resistance, where we used a special type of marker that targets resistance-related regions in the genome (VAN DER LINDEN *et al.* 2004; CALENGE *et al.* 2005). Since the structure of the data in this article is common to many breeding programs and

evaluation trials, the described association approach represents an example that can be adapted to specific programs, in potato and other species. The approach does not require the use of special-purpose software and can be implemented in any statistical package with extended facilities for mixed-model analyses, like SAS (SAS INSTITUTE 1999), GENSTAT (PAYNE and ARNOLD 2002), and ASREML (GILMOUR *et al.* 1998).

MATERIALS AND METHODS

Phenotypic and pedigree data: A first, exploratory association analysis was performed on phenotypic data stemming from VCU trials supporting the Dutch Potato Variety List. The data included evaluations for resistance against leaf blight (*P. infestans*) for 123 cultivars collected in the period 1975–2002. The number of years in which a particular cultivar was evaluated varied from a few years, 68 cultivars were evaluated from 1 to 3 years, to up to 25 years for 2 cultivars. Since in each year some cultivars were removed from the variety list while others were introduced, the phenotypic data set was highly unbalanced. The VCU trials were always inoculated with the same *P. infestans* isolate and cultivars were scored for resistance to leaf late blight on a scale from 1 (highly susceptible) to 9 (highly resistant). Although the original trials contained replicates, we worked with the trial means across replicates, as these were the only data available to us. We performed our statistical analyses on the original, ordinal measurements (1–9), as preliminary analyses showed that diagnostic plots of residuals showed no violation of the classical assumptions for analysis of variance and mixed models.

A second phenotypic data set for another 81 cultivars was constructed to corroborate significant marker–trait associations found in the study on the 123 above-mentioned cultivars. The 81 cultivars for our confirmatory study were a subset from the set of cultivars used by GEBHARDT *et al.* (2004). Cultivars in the confirmatory study were chosen differently from the cultivars in the exploratory study. Phenotypic data for the cultivars in the confirmatory study were passport data in which genotypes (accessions) were scored for resistance to late blight again on a scale from 1 (susceptible) to 9 (resistant). Because the data set was composed of evaluations performed in different countries, it is likely that not all the infections were produced by the same isolate (GEBHARDT *et al.* 2004).

Pedigree information for the 123 cultivars in the exploratory study and the 81 cultivars in the confirmatory study was retrieved from the Online Pedigree Database of the Laboratory of Plant Breeding of Wageningen University (HUTTEN and VAN BERLOO 2001).

Molecular marker data and LD between markers: The 123 cultivars in the exploratory study and the 81 in the confirmatory study have been genotyped with so-called nucleotide-binding-site (NBS) markers. These markers were developed upon the observation that a number of resistance genes for *P. infestans* encoded for proteins with a nucleotide-binding domain (NBS) and a leucine-rich repeat (LRR) domain (BALLVORA *et al.* 2002; VAN DER VOSSEN *et al.* 2003; BORMANN *et al.* 2004; HUANG *et al.* 2004). The production of NBS markers, NBS profiling, involves a PCR-based system that efficiently and reproducibly targets resistance genes (R genes) and resistance gene analogs (VAN DER LINDEN *et al.* 2004; CALENGE *et al.* 2005). For our genotypes, 49 NBS markers were dominantly scored as present (1) or absent (0). None of the 49 markers has been placed on a potato map, so their position in the genome is not yet known.

For the 81 cultivars in the confirmatory study, we had information about an additional marker, R1-1400, corresponding to a major resistance gene to late blight located on chromosome V (BALLVORA *et al.* 2002). After checking independence (nonassociation) of R1-1400 with the full set of NBS markers, we used R1-1400 as a cofactor in the model screening for marker–trait associations in the confirmatory data set.

Before looking at marker–trait associations, we calculated as descriptive statistics pairwise LD between markers using the square of the coefficient of correlation, r^2 (PRITCHARD and PRZEWORSKI 2001).

Models for the identification of marker–trait associations: In association mapping the objective is to introduce molecular marker information tagging specific DNA regions to test whether those regions have a significant association with or effect on the trait. Within our mixed-model framework, we partition the total genetic variation, *i.e.*, the variation between genotypes, into a fixed part associated with a particular QTL/ marker and a random part associated with the genetic background consisting of QTL elsewhere in the genome. A simple, but probably unrealistic (see below), starting model (model E1) to test for marker–trait associations in our exploratory late blight variety trial data is

$$\underline{P}_{ij} = \mu + \alpha X_i + \underline{g}_i + \underline{y}_j + \underline{u}_{ij} \qquad (1)$$

(random terms underlined), with $\underline{P}_{ij}$ the late blight resistance score for genotype i in year j, μ an intercept term, α the fixed (QTL) effect of the chromosomal region tagged by a particular marker, X_i the state of the marker in individual i (for a biallelic marker, $X_i = 0$ for marker band absence and $X_i = 1$ for marker band presence), $\underline{g}_i$ a residual random genotypic effect that results from not further identified background QTL, $\underline{y}_j$ the random year effect, and $\underline{u}_{ij}$ a residual term. Since the data consisted of averages over plots within a year, the residual term contains both genotype-by-year interaction and plot residual. The random terms are assumed to be normally and independently distributed with zero mean and a proper variance: $\underline{g}_i \sim N(0,\sigma_g^2)$, $\underline{y}_j \sim N(0,\sigma_y^2)$, and $\underline{u}_{ij} \sim N(0,\sigma^2)$. The mixed model (model E1, Equation 1) can be fitted by restricted maximum likelihood (REML), and marker–trait association assessed by testing for the H_0: $\alpha = 0$ by means of a Wald test (VERBEKE and MOLENBERGHS 2000). For ease of presentation, for all fitted models used in screens for marker–trait associations, the associations were classified according to the *P*-value of the Wald test statistic as: strong ($P < 0.001$), moderate ($0.001 < P < 0.01$), weak ($0.01 < P < 0.05$), and not significant ($P > 0.05$).

The distributional assumption for the background genetic effect, $\underline{g}_i$, in model E1 (Equation 1) imposes a genetic variance–covariance structure, G, that assumes no relatedness between genotypes, so with N genotypes, $G = I_N \sigma_g^2$, or, equivalently, an $N \times N$ matrix with σ_g^2 on the diagonals and zeroes on the off-diagonals. If the genotypes are barely related this is a reasonable assumption, but usually collections of cultivars will exhibit tighter relations due to shared breeding history. On the basis of an additive genetic model, the covariance between two genotypes i and i' with a coefficient of coancestry $\theta_{ii'}$ is $\text{cov}(i, i') = 2\theta_{ii'}\sigma_g^2$ (LYNCH and WALSH 1998). Therefore, a more realistic model would allow G to take structures other than just diagonal, depending on the degree of relatedness between the genotypes. When A is an $N \times N$ matrix containing the coefficients of coancestry for all pairs of genotypes, calculated from pedigree records, and σ_g^2 is the genetic variance due to undetected QTL elsewhere in the genome, $G = 2A\sigma_g^2$. The choice between alternative models for the G matrix can be based on a model selection criterion

such as the Bayesian information criterion (WOLFINGER 1993). Assumptions about the other random effects can also be relaxed (for example, introducing heterogeneity of residual variance between years) and comparisons between different models for these random effects can be based on the same model selection criterion.

In a second screen of the exploratory data, each of the 49 NBS markers was tested for association with leaf resistance to late blight within a mixed model that now included pedigree information, that is, model E1 (Equation 1) was converted to model E2 (Table 1) by generalizing the distributional assumption for the genotypic random effect to $\underline{g}_i \sim N(0, G)$, with G in this case a genetic variance–covariance matrix equal to $G = 2A\sigma_g^2$, while retaining the distributional assumptions for the year and residual effects: $\underline{y}_j \sim N(0,\sigma_y^2)$, and $\underline{u}_{ij} \sim N(0,\sigma^2)$.

Relations between genotypes are caused by a number of processes. Genotypes are related on the basis of pedigree, but additional relations can be superimposed by selection and drift. Especially the latter process will lead to the formation of subpopulations differing in allele frequencies, which will cause spurious marker–trait associations to occur if this population structure is ignored.

Assuming subpopulations known, within our mixed-model framework, a correlation structure can be imposed on the genotypes by including the subpopulations as the levels of a random factor. As a consequence, all pairs of genotypes will be equally correlated within groups, while being uncorrelated between groups. Such an approach could provide a crude, and sometimes satisfactory, approximation to the "real" genotypic correlations. The corresponding model (E3) becomes

$$\underline{P}_{i(k)j} = \mu + \underline{C}_k + \alpha X_{i(k)} + \underline{g}_{i(k)} + \underline{y}_j + \underline{u}_{i(k)j}, \qquad (2)$$

where $\underline{C}_k$ refers to the kth subpopulation of genotypes, and genotypes are now nested within subpopulation effects, leading to the genotypic, residual, effects $\underline{g}_{i(k)}$ (Table 1). We fitted model E3 (Equation 2) under the assumptions of $\underline{C}_k \sim N(0,\sigma_C^2)$ and $\underline{g}_{i(k)} \sim N(0,\sigma_g^2)$, effectively ignoring the pedigree relationships. We constructed model E4 from model E3 by incorporating the pedigree in the form of $\underline{g}_i \sim N(0, 2A\sigma_g^2)$ (Table 1).

Marker–trait association analyses in plant genetics have come to be to some extent equivalent to first identifying subpopulations and then performing some form of within-subpopulation association analysis (REMINGTON *et al.* 2001; KRAAKMAN *et al.* 2004). The method of choice for classifying genotypes in homogenous groups showing Hardy–Weinberg and linkage equilibrium uses the popular software STRUCTURE (PRITCHARD *et al.* 2000a). Within a Bayesian framework, the method implemented in this package assigns genotypes to a user-defined number of groups on the basis of the information given by molecular markers. Either the model can assume that individuals can be classified into unrelated groups (no admixture model) or that assumption can be relaxed to allow individuals to be the result of admixtures between groups (admixture model).

To assess subpopulations in our exploratory data, we used the 49 biallelic NBS markers, although we admit that these marker data look already *a priori* not very powerful and appropriate for this purpose. To identify subpopulations that subsequently define the levels of a random genotype-grouping factor in our models E3 and E4, we chose in STRUCTURE the admixture model and assumed varying numbers of groups, K, with $K = 2$, 3, and 4. For each assumed population structure, we ran the program discarding the first 10,000 iterations as burn-in period and using the following 100,000 iterations to produce the results on group membership probabilities for individual genotypes.

TABLE 1

Model descriptions

Data	Model code	Model	VCOV pedigree structured	Structured association
Exploratory	E1	$\underline{P}_{ij} = \mu + \underline{y}_j + \alpha X_i + \underline{g}_i + \underline{u}_{ij}$	No	No
	E2	$\underline{P}_{ij} = \mu + \underline{y}_j + \alpha X_i + \underline{g}_i + \underline{u}_{ij}$	Yes[a]	No
	E3	$\underline{P}_{ijk} = \mu + \underline{y}_j + \underline{C}_k + \alpha X_{i(k)} + \underline{g}_{i(k)} + \underline{u}_{i(k)j}$	No	Yes
	E4	$\underline{P}_{ijk} = \mu + \underline{y}_j + \underline{C}_k + \alpha X_{i(k)} + \underline{g}_{i(k)} + \underline{u}_{i(k)j}$	Yes[a]	Yes
	E5	$\underline{P}_{ijk} = \mu + y_j + C_k + \alpha X_{i(k)} + \underline{u}_{i(k)j}$	No	Yes
Confirmatory	C1	$\underline{P}_i = \mu + \beta R_i + \alpha X_i + \underline{u}_i$	No	—
	C2	$\underline{P}_i = \mu + \beta R_i + \alpha X_i + \underline{u}_i$	Yes[b]	—

Models used for analysis of the explanatory and confirmatory data sets: μ, intercept; y, year; X, marker state (1/0); α, QTL effect; g, genotypic effect; C, substructure effect; and u, residual effect. Underlined terms are random with a proper variance except for the genotypic effect in the case of structuring by pedigree information. The last two columns indicate whether pedigree information and/or population substructure information were included in the variance–covariance (VCOV) model for the genotypic effects.

[a] $\underline{g}_i \sim N(0, 2A\sigma_g^2)$.
[b] $\underline{u}_i \sim N(0, 2A\sigma_u^2)$.

As standard fixed regression and analysis of variance models, including t-tests and simple correlations, are still more popular than mixed models for association analyses (Ford-Lloyd *et al.* 2001; Remington *et al.* 2001; Ivandic *et al.* 2002; Kraakman *et al.* 2004; Skøt *et al.* 2005), we performed, for comparison with the above mixed models, association analyses under the following fixed regression model (E5):

$$\underline{P}_{i(k)j} = \mu + C_k + \alpha X_{i(k)} + g_{i(k)} + y_j + \underline{u}_{i(k)j}. \qquad (3)$$

To corroborate associations found in our exploratory data set, we performed additional association analyses in a second, independent, confirmatory data set. The confirmatory data did not have replications over years, so that the year term disappears from the model, together with the index j, leading to a confounding of residual genotypic effects and residual effects due to experimental errors (and possible genotype-by-year interactions) in one random term, $\underline{u}_i$. For the confirmatory data set, scores were available for the resistance-related binary (1/0) R1-1400 marker, R_i, and we thought it sensible to include this marker as a covariable/cofactor. The model for the confirmatory data then becomes

$$\underline{P}_i = \mu + \beta R_i + \alpha X_i + \underline{u}_i \qquad (4)$$

(model C1), where β represents the effect of the R1-1400 marker. We fitted model C1 (Equation 4) with $\underline{u}_i \sim N(0, \sigma_u^2)$, thus ignoring the pedigree relationships (Table 1). Finally, we included the pedigree, $\underline{u}_i \sim N(0, 2A\sigma_u^2)$, in our model C2 for the confirmatory data (Table 1).

RESULTS

In general, for the genotypes included in the exploratory and confirmatory data set, marker bands were present at moderate to low frequencies (median frequency = 0.22 and 0.27 in the exploratory and confirmatory data sets, respectively), although band presence ranged from rare (0.05 and 0.02, respectively) to almost fixed (0.92 and 0.86, respectively) (Figure 1). Marker band frequencies are not very useful as estimates of allele frequencies, because the number of alleles in a genotype (from one in a simplex to four in a quadruplex) cannot be determined with a dominant marker system like NBS. The fact that most of the bands were at relatively low frequencies points to one of the advantages of using an association mapping approach, as many of the polymorphisms would have passed unnoted in a conventional approach based on the segregation of a specific cross. A negative effect of the low frequencies of the markers is that the power of the test is reduced.

LD was generally low between markers, with most r^2-values <0.10 (Figure 1). While the low LD between markers indicates that most markers are independent of each other, it is difficult to conclude about distances between markers. Values of r^2 of <0.10 have been observed between markers as close as 1500 bp apart in maize (Remington *et al.* 2001) or at distances >10 cM in barley (Kraakman *et al.* 2004). The extent of LD in potato is not known, but it is likely to be large as genotypes are autotetraploid cultivars with a narrow genetic base and with relatively few generations since their introduction in Europe. There are, however, some examples in potato where a strong decay in LD was observed at distances of <1 cM (Gebhardt *et al.* 2004).

For the exploratory data, the clustering of genotypes using STRUCTURE did not produce a clear discrimination of the genotypes into hypothetical groups ($K = 2$, 3, and 4). Figure 2 shows the results for a three-groups solution ($K = 3$). Although the bars indicate that some genotypes have a genetic background with a large fraction from one of the three hypothetical subpopulations, substantial intermixing between the groups was evident. Nevertheless, we concluded on the existence of three groups and as Remington *et al.* (2001) did, we assigned each genotype to the group with which the
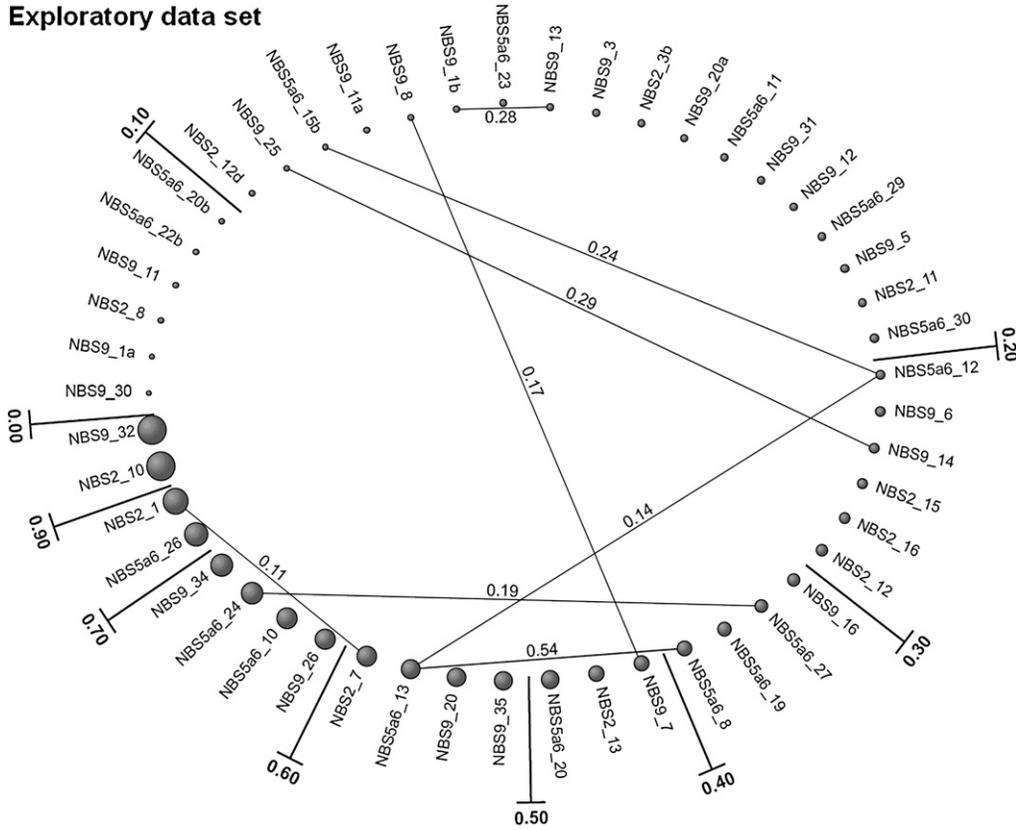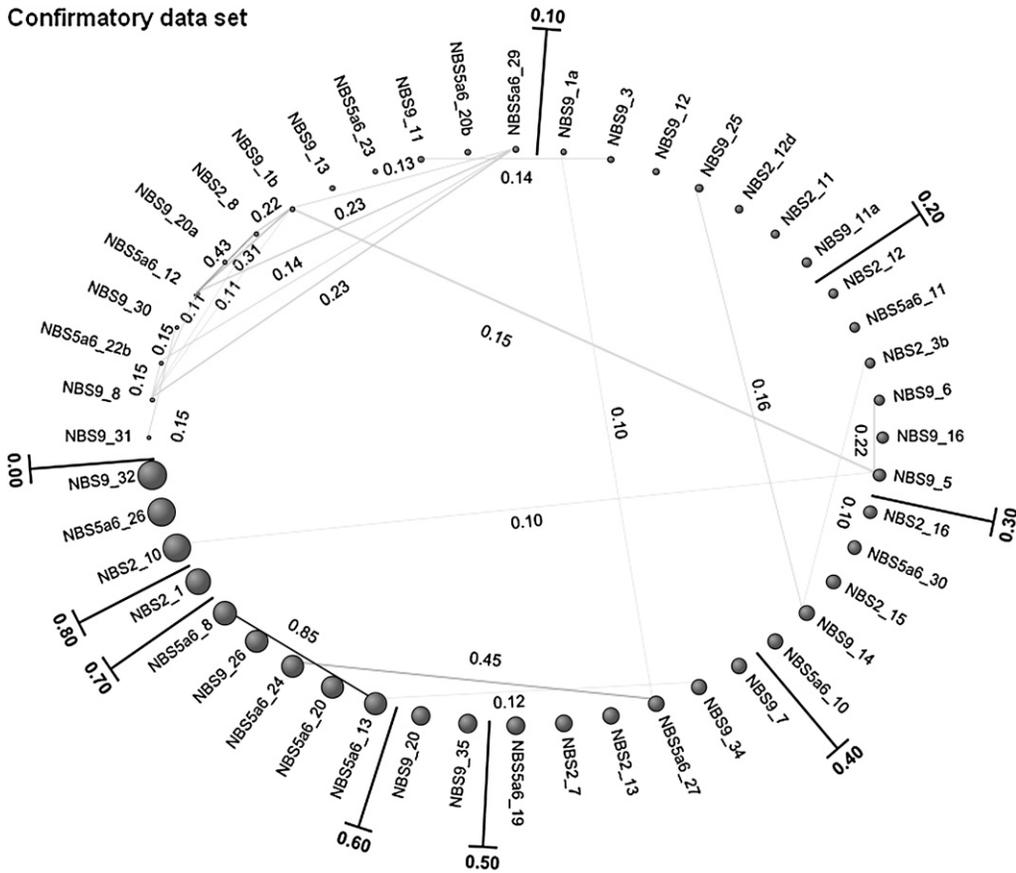
FIGURE 1.—LD between NBS markers in an exploratory and a confirmatory data set. For the exploratory and confirmatory data, name, band frequency, and LD (r^2) between the 49 NBS markers are shown. The size of the symbol is proportional to the band frequency of the marker in the population (the larger the symbol is, the higher the frequency). Markers are clockwise ordered according to increasing band frequency with scaling marks appearing at steps of 0.1. Pairs of markers with $r^2 > 0.10$ are connected by lines.
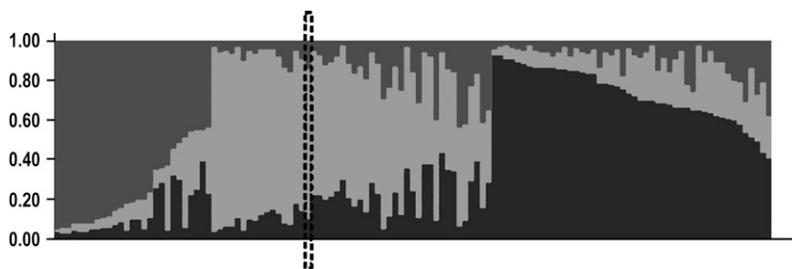
FIGURE 2.—STRUCTURE group probabilities for exploratory data. The bar plot represents the genetic background of 123 potato cultivars. Genotypes are represented on the horizontal axis, each one by a bar composed of three regions according to the proportion of the genetic background that was inferred to come from each of three different groups in the population (indicated by different degrees of shading). The dotted box highlights one genotype as an example of a genotype that has a major proportion of its background coming from the group with light shading.

genotype shared the largest genetic background fraction. For the confirmatory data, from the analysis by STRUCTURE we could not derive any useful grouping for later use in association analyses.

Before fitting the association models E1–E5 to the exploratory data and models C1 and C2 to the confirmatory data (see Table 1 for all models), we investigated whether the pedigree relationships contributed anything to a better description of the genetic covariances between genotypes. For the exploratory data we compared the fit of models E2 and E1, leaving out marker information, where the first model did contain a pedigree-structured genotypic variance–covariance matrix and the second model did not. Comparison of the Bayesian information criterion indicated a clear improvement of model E2 over E1, 654.1 vs. 684.3 (a lower number is better). An analogous comparison can be made between models E4 and E3, with a Bayesian information criterion (BIC) of 659.9 vs. 687.1, leading to the same conclusion. For the confirmatory data, we compared model C2 with C1, with a BIC of 131.1 vs. 133.6, showing that for this data set inclusion of pedigree information led to only marginal improvement of the model fit.

The P-values of the Wald tests for NBS markers with at least one significant result across models E1–C2 are presented in Figure 3. On the basis of the BIC (lowest values) the models of choice should be E2, with pedigree-structured genotypic variances for the exploratory data, and C2 for the confirmatory data, again with pedigree-structured genotypic variances and covariances. The other models are presented to show the effects of omitting pedigree information and/or structured association. Model E5 (Equation 3) serves as an anchoring point as it represents a default model of choice for many association analyses in the literature. The rows in Figure 3 are sorted in ascending order by the P-values found in model E5. At one glance it is obvious from Figure 3 that model E5 produced far more significant marker–trait associations than any other model. However, as deviance tests pointed to the desirability of inclusion of pedigree information, plus as it appeared natural to choose random instead of fixed genotype-by-year interaction, we know that model E2 provides a better fit than model E5. Therefore, the significant

associations for model E5 that are not reproduced by model E2 are false positives, stemming from the use of an inadequate marker–trait association model. With an analogous argument, we find that because of the omission of pedigree information, model E1 is too liberal in comparison to model E2, as is model E3 to E4 and model C1 to C2. In all three model comparisons, P-values tend to increase, i.e., become less significant, with the inclusion of pedigree information, emphasizing too liberal testing in the absence of appropriate structuring of the genotypic variance–covariance matrix. From the significant markers detected by the application of model E2 only the markers NBS2_8 and NBS5a6_10 carry over from the exploratory data to the confirmatory data as analyzed by model C2.

Marker NBS5a6_10 showed the highest association with the trait in the exploratory data set ($P < 0.001$) and had a positive effect on resistance (Table 2). Band presence of this marker was associated with an increase in resistance score of 0.95 ± 0.244 as estimated in model E2, an effect that was consistent with the one of 0.74 ± 0.326 in model C2 found for the confirmatory data set. The other marker consistently associated with resistance was NBS2_8 with an increase of 0.98 ± 0.370 in resistance score for the exploratory data set as estimated in model E2, while for the confirmatory data fitting model C2 the effect was 1.91 ± 0.697. In the exploratory data, NBS5a6_10 and NBS2_8 had effects of similar magnitude, but the test statistic was much lower for NBS2_8, possibly because of the marker band being present at low frequency. In the confirmatory data set, this marker appeared at an even lower band frequency, but still showed a significant effect, supporting its association with the trait.

DISCUSSION

Although widely used in human genetics, association mapping has recently emerged in the plant breeding arena. A major reason for the interest in association mapping approaches in plant breeding is that the extensive phenotypic evaluations routinely produced by breeding programs and cultivar evaluation systems contain a large potential for studying the genetic

Model

Marker	E1	E2	E3	E4	E5	C1	C2
NBS2_10							
NBS9_35							
NBS2_16							
NBS5a6_15b							
NBS9_5							
NBS9_30							
NBS9_20							
NBS9_11							
NBS5a6_23							
NBS2_3b							
NBS9_8							
NBS9_34							
NBS2_8							
NBS9_1a							
NBS9_26							
NBS5a6_10							
NBS5a6_20							
NBS2_15							
NBS2_12							
NBS9_14							
NBS9_12							
NBS2_11							
NBS9_16							
NBS5a6_27							
NBS9_25							
NBS9_3							
NBS2_1							
NBS9_7							
NBS9_20a							
NBS5a6_22b							
NBS5a6_12							
NBS9_11a							
NBS2_7							

■ Strong association, $P < 0.001$
▨ Moderate association, $0.001 < P < 0.01$
▧ Weak association, $0.01 < P < 0.05$
□ No association, $P > 0.05$

FIGURE 3.—Selected NBS markers by different models. The numbers of strong ($P < 0.001$), moderate ($0.001 < P < 0.01$), weak ($0.01 < P < 0.05$), and not significant ($P > 0.05$) marker–trait associations between 49 NBS markers and late blight resistance resulting from the fit of the models described in Table 1 are shown.

mechanisms underlying trait variation in many crops. Sufficient conditions for unlocking that potential are targeted marker systems in combination with sound statistical procedures. In this article, we showed how 25 years of historical phenotypic data as collected within the Dutch potato cultivar evaluation system could be exploited to investigate the genetic basis for late blight resistance in potato, a challenging disease with large economic consequences. We tested for associations between 49 NBS markers, a new type of marker targeting resistance genes and their analogs, and late blight resistance, using mixed-model methodology. To assess the consistency of our procedure we compared the results of the association analyses as performed on the Dutch variety trials with analyses on a second independent set made up of passport data that were published as additional material to GEBHARDT *et al.* (2004).

When LD decays over short distances, the number of random molecular markers that are necessary to find significant associations can be substantial. NBS markers allowed us to focus directly on interesting genomic regions. With a relatively reduced number of NBS markers we singled out two markers that showed consistent effects across populations for resistance to late blight. The absence of map information hindered comparison of our detected associations with previously reported QTL. Nevertheless, since we had information on R1-1400, a marker linked to a major resistance gene that maps on chromosome V, we could check for its association with either one of the two NBS markers. The observed LD (r^2) between R1-1400 and NBS5a6_10 was 0.023 and that between R1-1400 and NBS2_8 was 0.011, suggesting that the markers map either to a distant region on chromosome V or to other chromosomes. Sequence information for marker NBS2_8 did not produce a clear indication for its location, but for NBS5a6_10 a strong homology with genes of the R3 complex (results not shown) located this marker on chromosome XI (HUANG *et al.* 2004).

The mixed-model framework is for various reasons highly appropriate to study marker–trait associations. Many historical data sets in plant breeding are severely unbalanced in that the set of genotypes changes between years and often also between trials within years. Within plant breeding, mixed models constitute the method of choice to deal with unbalanced data across multiple trials. We extended the standard phenotypic analysis of multiple trials by mixed models to arrive at models suitable for association mapping by introducing marker genotype information as covariables at the levels of the genotypic factor, thereby creating so-called mixed factorial regression models (VAN EEUWIJK *et al.* 1996; DENIS *et al.* 1997; MALOSETTI *et al.* 2004). The introduction of molecular marker information at the genotypic level, corresponding to the simultaneous modeling of multiple trials while acknowledging the correlation structure between those trials, distinguishes our approach from that of YU *et al.* (2006), who present a model for identifying marker–trait associations for a single set of observations on a collection of genotypes.

Mixed models guarantee reliable inference through the explicit modeling of correlations induced by genetic and environmental causes. In genetic contexts, from the genotypic dimension, correlations between observations can arise in many ways. First, observations taken on the same genotype across multiple trials and replicates within trials can be expected to be correlated. Second, genetic correlations between observations made on different genotypes follow from these genotypes sharing ancestral history, *i.e.*, pedigree relations,

TABLE 2

Test statistics and effect sizes for consistent markers

Marker	Exploratory data set (123 cultivars)					Confirmatory data set (81 cultivars)				
	Freq	Wald	P	Effect	SE	Freq	Wald	P	Effect	SE
NBS5a6_10	0.63	15.1	<0.001	0.95	0.244	0.40	5.2	0.023	0.74	0.326
NBS2_8	0.09	7.1	0.008	0.98	0.370	0.05	7.5	0.006	1.91	0.697

A description of consistent marker–trait associations across two independent data sets is shown. The exploratory data set spans 25 years of Dutch variety trials and the confirmatory data set is based on passport data used earlier by GEBHARDT *et al.* (2004). Columns show marker band frequency (Freq), the Wald test statistic for marker–trait association with late blight resistance, the *P*-value for the Wald statistic, the estimated QTL effect, and the standard error (SE).

or from these genotypes having been subjected to similar selection forces and/or drift processes. To model correlations between multiple observations on the same genotype, we can choose genotypic effects to be random in an appropriate mixed model. To account for correlations due to pedigree, selection, and drift, we can further structure the variance–covariance matrix of the random genotypic effects. For rough-grained structure following from especially drift and to a lesser extent selection, we can introduce a random factor whose levels coincide with groups of genotypes in such a way that most or all of LD is explained by differences between those groups, where the groups themselves are relatively homogeneous and in, or close to, Hardy–Weinberg and linkage equilibrium. For the imposition of more subtle structure on the genotypic variance–covariance matrix, coancestry relations could be calculated from known pedigree relations, while as an alternative or complementary means for capturing fine-grained correlation structure molecular marker-based estimates for pairwise genotypic relatedness can be used (LYNCH and RITLAND 1999; WANG 2002; MILLIGAN 2003).

Fine-grained pedigree information can be hard to acquire and therefore many articles on association mapping address a coarse level of genotypic correlation by classifying genotypes into more or less homogenous groups. Correcting for these group effects is an indirect way of correcting for relatedness between genotypes and therefore less spurious associations are expected. However, defining homogeneous groups in a collection of genotypes is not always a simple task. For our potato data, we could not find a clear group structure in the exploratory data and no structure at all in confirmatory data using the methodology proposed by PRITCHARD *et al.* (2000b). The low number of dominant markers that we used will certainly have contributed to our problems in identifying groups. Still, the absence of a clear group structure in collections of elite breeding germplasm has been observed before in similar works (KRAAKMAN *et al.* 2004; SIMKO *et al.* 2004b). Elite breeding genotypes have a relatively narrow genetic base resulting from closely related breeding history, so it can be expected that such populations consist mainly of

one major gene pool rather than of the intermixing of very distinct subpopulations. Studies where a clear population structure was observed generally involved broad-based germplasm and the groupings discriminated between rather distinct genetic backgrounds. For example, in ryegrass, groups corresponded to different species (SKØT *et al.* 2005) and in maize to different heterotic groups (REMINGTON *et al.* 2001). In addition to the difficulty of finding structure in the population, even when groups are found, the correction for group effects may remove part but not all of the genetic correlation. Therefore, substantial polygenic effects may still persist and affect the results. For our data, the influence of the groups (model E3 *vs.* E1 and model E4 *vs.* E2) was very small, especially in comparison to the influence of the pedigree.

Ignoring correlations between genotypes will lead to liberal tests of marker–trait associations in the same way as testing of whole-plot treatments over subplot errors in split-plot designs will in general lead to liberal tests for the whole-plot treatment factor. With regard to QTL detection, KENNEDY *et al.* (1992) used simulation to show that failure to account for the variance–covariance structure generated by relatedness in complex pedigree populations facilitates an increase of false positive QTL detection. Our potato data emphasized the same point, as only the inclusion of pedigree information in the mixed model produced consistent significant associations across two independent studies (associations identified by models E2 and C2). In agreement with our results for potato, PARISSEAUX and BERNARDO (2004) identified marker–trait associations for maize that were reproducible over populations, also by using mixed models including pedigree information.

A fixed approach to association mapping, implying neglect of correlations between multiple observations on the same genotypes and of coarse- and fine-grained correlation structures between genotypes, produces a large amount of associations of which the majority will be false positives that will not be reproducible. Our model E5, although including a correction for population structure, closely resembled this approach and indeed produced many not reproducible associations. Still, the fixed approach is not uncommon in the

literature (Ford-Lloyd *et al.* 2001; Remington *et al.* 2001; Ivandic *et al.* 2002; Kraakman *et al.* 2004; Skøt *et al.* 2005), probably due to the larger availability of and familiarity with fixed than mixed models.

Relations between genotypes need to be adequately represented in statistical models for marker–trait associations to guarantee reliable inference. Analogous to Yu *et al.* (2006), we presented a mixed-model framework that offered two ways of representing genotypic relations: by structuring the variance–covariance matrix of the genotypic effects using pedigree or marker information and by introducing a grouping factor to represent structured association. When the pedigree is accurate or when enough marker information is available to estimate reliably individual pairwise genotypic relations, the inclusion of a factor representing structured association is not expected to contribute much to the model fit. The structured association factor merely provides a crude approximation to the real matrix of pairwise genotypic relations, given that the groups can be identified reasonably. When population structure is suspected, but methods to identify groups are hard to apply or fail, the method of genomic control (Devlin and Roeder 1999; Devlin *et al.* 2001) can be an attractive procedure to follow. Genomic control is based on the assumption that population structure will induce overdispersion that will affect the test statistics for association. The amount of overdispersion can be estimated from the empirical values for the test statistic, after which the empirical test statistics can be adjusted on the basis of the estimated overdispersion. This procedure is simple to apply and seems to work well in simple clinical settings (case–control studies). For our purposes, however, when modeling marker–trait associations in plant data coming from multiple environments, using mixed models with complex variance–covariance structures, it is not so obvious how the theory of genomic control can be generalized, as it will be difficult to motivate a unique overdispersion factor across the whole of the data set. Therefore, we feel that a mixed-model approach as discussed in this article is a highly attractive framework for studying marker–trait association for plant data across multiple environments.

LITERATURE CITED

Asins, M. J., 2002 Present and future of quantitative trait locus analysis in plant breeding. Plant Breed. **121:** 281–291.

Ballvora, A., M. R. Ercolano, J. Weiss, K. Meksem, C. A. Bormann *et al.*, 2002 The R1 gene for potato resistance to late blight (Phytophthora infestans) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. Plant J. **30:** 361–371.

Bormann, C. A., A. M. Rickert, R. A. Castillo Ruiz, J. Paal, J. Lübeck *et al.*, 2004 Tagging quantitative trait loci for maturity-corrected late blight resistance in tetraploid potato with PCR-based candidate gene markers. Mol. Plant Microbe Interact. **17:** 1126–1138.

Calenge, F., C. G. Van der Linden, E. Van de Weg, H. J. Schouten, G. Van Arkel *et al.*, 2005 Resistance gene analogues identified through the NBS-profiling method map close to major genes and

QTL for disease resistance in apple. Theor. Appl. Genet. **110:** 660–668.

Collard, B. C. Y., M. Z. Z. Jahufer, J. B. Brouwer and E. C. K. Pang, 2005 An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica **142:** 169–196.

Collins, A., D. Milbourne, L. Ramsay, R. Meyer, C. Chatot-Balandras *et al.*, 1999 QTL for field resistance to late blight in potato are strongly correlated with maturity and vigour. Mol. Breed. **5:** 387–398.

Darvasi, A., A. Weinreb, V. Minke, J. I. Weller and M. Soller, 1993 Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics **134:** 943–951.

Dekkers, J. C. M., and F. Hospital, 2002 The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet. **3:** 22–32.

Denis, J. B., H. P. Piepho and F. A. van Eeuwijk, 1997 Modelling expectation and variance for genotype by environment data. Heredity **79:** 162–171.

Devlin, B., and K. Roeder, 1999 Genomic control for association studies. Biometrics **55:** 997–1004.

Devlin, B., K. Roeder and L. Wasserman, 2001 Genomic control, a new approach to genetic-based association studies. Theor. Popul. Biol. **60:** 155–166.

Ewing, E. E., I. Simko, C. D. Smart, M. W. Bonierbale, E. S. G. Mizubuti *et al.*, 2000 Genetic mapping from field tests of qualitative and quantitative resistance to Phytophthora infestans in a population derived from Solanum tuberosum and Solanum berthaultii. Mol. Breed. **6:** 25–36.

Flint-Garcia, S. A., J. M. Thornsberry and I. V. E. S. Buckler, 2003 Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54:** 357–374.

Ford-Lloyd, B. V., H. J. Newbury, M. T. Jackson and P. S. Virk, 2001 Genetic basis for co-adaptive gene complexes in rice (*Oryza sativa* L.) landraces. Heredity **87:** 530–536.

Gebhardt, C., A. Ballvora, B. Walkemeier, P. Oberhagemann and K. Schüler, 2004 Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol. Breed. **13:** 93–102.

Gilmour, A., B. Cullis, S. J. Welham and R. Thompson, 1998 *ASREML. Program User Manual.* Orange Agricultural Institute, New South Wales, Australia.

Hospital, F., L. Moreau, F. Lacoundre, A. Charcosset and A. Gallais, 1997 More on the efficiency of marker-assisted selection. Theor. Appl. Genet. **95:** 1181–1189.

Huang, S., V. G. A. A. Vleeshouwers, J. S. Werij, R. C. B. Hutten, H. J. van Eck *et al.*, 2004 The *R3* resistance to *Phytophthora infestans* in potato is conferred by two closely linked *R* genes with distinct specificities. Mol. Plant Microbe Interact. **17:** 428–435.

Hutten, R. C. B., and R. van Berloo, 2001 An online potato pedigree data base (http://www.dpw.wau.nl/pv/query.asp).

Ivandic, V., C. A. Hackett, E. Nevo, R. Keith, W. T. B. Thomas *et al.*, 2002 Analysis of simple sequence repeats (SSRs) in wild barley from the Fertile Crescent: associations with ecology, geography and flowering time. Plant Mol. Biol. **48:** 511–527.

Jannink, J. L., and B. Walsh, 2002 Association mapping in plant populations, pp. 59–68 in *Quantitative Genetics, Genomics and Plant Breeding*, edited by M. S. Kang. CAB International, New York.

Jannink, J. L., M. Bink and R. C. Jansen, 2001 Using complex plant pedigrees to map valuable genes. Trends Plant Sci. **6:** 337–342.

Kennedy, B. W., M. Quinton and J. A. van Arendonk, 1992 Estimation of effects of single genes on quantitative traits. J. Anim. Sci. **70:** 2000–2012.

Kraakman, A. T. W., R. E. Niks, P. M. M. M. Van den Berg, P. Stam and F. A. Van Eeuwijk, 2004 Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics **168:** 435–446.

Lynch, M., and K. Ritland, 1999 Estimation of pairwise relatedness with molecular markers. Genetics **152:** 1753–1766.

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

Malosetti, M., J. Voltas, I. Romagosa, S. E. Ullrich and F. A. van Eeuwijk, 2004 Mixed models including environmental

variables for studying QTL by environment interaction. Euphytica **137:** 139–145.

MEYER, R. C., D. MILBOURNE, C. A. HACKETT, J. E. BRADSHAW, J. W. MCNICHOL *et al.*, 1998 Linkage analysis in tetraploid potato and association of markers with quantitative resistance to late blight (*Phytophthora infestans*). Mol. Gen. Genet. **259:** 150–160.

MILLIGAN, B. G., 2003 Maximum-likelihood estimation of relatedness. Genetics **163:** 1153–1167.

OBERHAGEMANN, P., C. CHATOT-BALANDRAS, R. SCHÄFER-PREGL, D. WEGENER, C. PALOMINO *et al.*, 1999 A genetic analysis of quantitative resistance to late blight in potato: towards marker-assisted selection. Mol. Breed. **5:** 399–415.

PARISSEAUX, B., and R. BERNARDO, 2004 In silico mapping of quantitative trait loci in maize. Theor. Appl. Genet. **109:** 508–514.

PAYNE, R. W., and G. M. ARNOLD, 2002 *Genstat Release 6.1. Reference Manual*, Part 3. VSN International, Oxford.

PRITCHARD, J. K., and M. PRZEWORSKI, 2001 Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. **69:** 1–14.

PRITCHARD, J. K., M. STEPHENS and P. DONNELLY, 2000a Inference of population structure using multilocus genotype data. Genetics **155:** 945–959.

PRITCHARD, J. K., M. STEPHENS, N. A. ROSENBERG and P. DONNELLY, 2000b Association mapping in structured populations. Am. J. Hum. Genet. **67:** 170–181.

REMINGTON, D. L., J. M. THORNSBERRY, Y. MATSUOKA, L. M. WILSON, S. R. WHITT *et al.*, 2001 Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA **98:** 11479–11484.

SAS INSTITUTE, 1999 *SAS/Stat User's Guide*, Version 8. SAS Institute, Cary, NC.

SIMKO, I., S. COSTANZO, K. G. HAYNES, B. J. CHRIST and R. W. JONES, 2004a Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. Theor. Appl. Genet. **108:** 217–224.

SIMKO, I., K. G. HAYNES, E. E. EWING, S. COSTANZO, B. J. CHRIST *et al.*, 2004b Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. Mol. Genet. Genomics **271:** 522–531.

SKØT, L., M. O. HUMPHREYS, I. ARMSTEAD, S. HEYWOOD, K. P. SKØT *et al.*, 2005 An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). Mol. Breed. **15:** 233–245.

THORNSBERRY, J. M., M. M. GOODMAN, J. DOEBLEY, S. KRESOVICH, D. NIELSEN *et al.*, 2001 *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. **28:** 286–289.

VAN DER LINDEN, C. G., D. C. A. E. WOUTERS, V. MIHALKA, E. Z. KOCHIEVA, M. J. M. SMULDERS *et al.*, 2004 Efficient targeting of plant disease resistance loci using NBS profiling. Theor. Appl. Genet. **109:** 384–393.

VAN DER VOSSEN, E., A. SIKKEMA, B. T. L. HEKKERT, J. GROS, P. STEVENS *et al.*, 2003 An ancient R gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. Plant J. **36:** 867–882.

VAN EEUWIJK, F. A., J. B. DENIS and M. S. KANG, 1996 Incorporating additional information on genotypes and environments in models for two-way genotype by environment tables, pp. 15–49 in *Genotype-by-Environment Interaction: New Perspectives*, edited by M. S. KANG and H. G. GAUCH, JR. CRC Press, Boca Raton, FL.

VERBEKE, G., and G. MOLENBERGHS, 2000 *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

VISKER, M. H. P. W., L. C. P. KEIZER, H. J. VAN ECK, E. JACOBSEN, L. T. COLON *et al.*, 2003 Can the QTL for late blight resistance on potato chromosome 5 be attributed to foliage maturity type? Theor. Appl. Genet. **106:** 317–325.

WALSH, B., 2002 Quantitative genetics, genomics and the future of plant breeding, pp. 23–32 in *Quantitative Genetics, Genomics and Plant Breeding*, edited by M. S. KANG. CAB International, New York.

WANG, J., 2002 An estimator for pairwise relatedness using molecular markers. Genetics **160:** 1203–1215.

WILSON, L. M., S. R. WHITT, A. M. IBANEZ, T. R. ROCHEFORD, M. M. GOODMAN *et al.*, 2004 Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell **16:** 2719–2733.

WOLFINGER, R. D., 1993 Covariance structure selection in general mixed models. Commun. Stat. A **22:** 1079–1106.

YOUNG, D. N., 1999 A cautiously optimistic vision for marker-assisted breeding. Mol. Breed. **5:** 505–510.

YU, J., G. PRESSOIR, W. H. BRIGGS, I. VROH BI, M. YAMASAKI *et al.*, 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38:** 203–208.

Communicating editor: J. B. WALSH