

Methodology article

Open Access

MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments

Jean-Baptiste Veyrieras*¹, Bruno Goffinet² and Alain Charcosset¹

Address: ¹UMR, INRA UPS-XI INAPG CNRS Génétique Végétale, Ferme du Moulon, 91190 Gif-sur-Yvette, France and ²BIA, Chemin de Borde Rouge BP27 31326, Castanet Tolosan Cedex, France

Email: Jean-Baptiste Veyrieras* - veyrieras@moulon.inra.fr; Bruno Goffinet - goffinet@toulouse.inra.fr; Alain Charcosset - charcos@moulon.inra.fr

* Corresponding author

Published: 8 February 2007

Received: 14 November 2006

BMC Bioinformatics 2007, 8:49 doi:10.1186/1471-2105-8-49

Accepted: 8 February 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/49>

© 2007 Veyrieras et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Integration of multiple results from Quantitative Trait Loci (QTL) studies is a key point to understand the genetic determinism of complex traits. Up to now many efforts have been made by public database developers to facilitate the storage, compilation and visualization of multiple QTL mapping experiment results. However, studying the congruency between these results still remains a complex task. Presently, the few computational and statistical frameworks to do so are mainly based on empirical methods (e.g. consensus genetic maps are generally built by iterative projection).

Results: In this article, we present a new computational and statistical package, called MetaQTL, for carrying out whole-genome meta-analysis of QTL mapping experiments. Contrary to existing methods, MetaQTL offers a complete statistical process to establish a consensus model for both the marker and the QTL positions on the whole genome. First, MetaQTL implements a new statistical approach to merge multiple distinct genetic maps into a single consensus map which is optimal in terms of weighted least squares and can be used to investigate recombination rate heterogeneity between studies. Secondly, assuming that QTL can be projected on the consensus map, MetaQTL offers a new clustering approach based on a Gaussian mixture model to decide how many QTL underly the distribution of the observed QTL.

Conclusion: We demonstrate using simulations that the usual model choice criteria from mixture model literature perform relatively well in this context. As expected, simulations also show that this new clustering algorithm leads to a reduction in the length of the confidence interval of QTL location provided that across studies there are enough observed QTL for each underlying true QTL location. The usefulness of our approach is illustrated on published QTL detection results of flowering time in maize. Finally, MetaQTL is freely available at <http://bioinformatics.org/mqtl>.

Background

In the last two decades, the advent of molecular markers and their use in linkage mapping experiments has tremendously increased the potential of quantitative genetics.

Linkage mapping experiments now provide an efficient tool to identify regions of the genome where polymorphism affects the variation of quantitative traits, called Quantitative Trait Loci (QTL). Although a large number of

advanced statistical methods have been developed to improve the localization of QTL, the limited number of recombination events available in routinely used pedigree designs for QTL mapping lead essentially to an approximate mapping of the QTL (see for instance [1]). This is mainly due to both a few mating generations and a restricted number of sampled individuals (generally a few hundreds). From results of QTL experiments gathered over a wide range of plant species, [2] have shown that confidence intervals around most likely QTL positions are, on average, approximately 10 cM, which usually includes several hundreds of genes. More recent advances in the area of molecular biology have allowed researchers to carry out positional cloning of QTL (see for instance the review of [3]) but this approach still remains extremely expensive both in terms of time and resources. Also several authors [2,4] have pointed out that QTL detection is statistically biased both in the true number of QTL, which is underestimated since only QTL with large effects are detected, and in the QTL effects which are over estimated as only significant effects are reported (a phenomenon has commonly referred to as the Beavis effect [5]). Even though QTL mapping experiments must be considered with an awareness of these limitations, they have become commonplace and have greatly improved our knowledge about the genetic component of complex traits.

Since the first publication of a QTL localization using molecular data [6], more and more species and traits have been studied and many of these results have been made available via public databases. One of the main purposes of these databases was to help researchers to compare results from different QTL studies, to study the congruency of QTL locations in order to address the following question: "do QTL identified for a given trait in a population correspond to those detected in other populations?". In theory one would expect that the variation of a quantitative trait within a species is explained by a finite number of genes. Thus QTL congruency investigation will be a relevant approach to improve knowledge on trait genetics and several publications have pointed out its usefulness [7-12]. Nevertheless, the combination of results from linkage studies can be tedious since, even if several studies focus on the same trait within the same species, family structures, sample sizes, marker maps, or QTL detection methods may differ between studies. Some methods have been recently developed to tackle the issues raised by between QTL studies heterogeneity. Integration of genetic maps and QTL locations by iterative projections on a reference map is now widely used to position both markers and QTL on a single and homogeneous consensus map (see for instance [13]). However this process yields a consensus marker map for which both the statistical properties and biological "reality" can't be clearly assessed, even if a robust ordered marker map was used as reference. [14]

proposed an original approach using graph theory to integrate various types of maps (genetic, physical or sequence-based) but it mainly dealt with dissection of marker order inconsistencies between maps. From up to now it seems that there is no efficient methodological framework to build reliable consensus marker maps on which markers and candidate genes from different mapping experiments can be both ordered and positioned (except by merging raw mapping data from multiple populations as proposed by [15] and [16]).

In order to study QTL congruency, [17] proposed an original approach based on a meta-analysis strategy. Meta-analysis, which is mainly used in medical, social, and behavioral sciences, aims to pool results across independent studies in order to combine them in a single result or estimate. The relevance of meta-analysis investigations in genetics and evolution has been discussed and pointed out by several authors in the last decade (see for instance [18-21]). More recently [22] developed another meta-analysis based approach to overcome the between-study heterogeneity and to refine both QTL location and the magnitude of the genetic effects. Yet both the method of [17] and [22] are limited to a small number of underlying QTL positions (from one to four for the former and only one for the later) which is a serious limitation for a whole genome study of QTL congruency. Even if the average number of QTL per experiment is around four in plants [2,12], one would expect that more than four genes can be involved in the trait variation on a single chromosome.

To remove these impediments we have developed a new 2-stage meta-analysis procedure in order to integrate multiple independent QTL mapping experiments. Our aim was to create a global framework to evaluate the homogeneity of both genetic marker and QTL mapping results from literature and public databases. The first part of our meta-analysis procedure consists in building a consensus genetic marker map that takes into account the statistical properties of genetic distance estimates using a Weighted Least Squares (WLS) strategy. Secondly, once the consensus marker map has been built, the QTL locations can be projected on to the map. We also propose a new clustering algorithm based on a Gaussian mixture model in order to identify the number of underlying QTL which best explain the observed distribution of QTL positions in the mapping experiments. As it has been emphasized by [17], the crucial point at this step is to find an unbiased criterion to select the correct number of QTL. In the context of Gaussian mixture, a large variety of model choice criteria have been reported in the literature. We explore, by means of simulations, the properties of some of these criteria for our particular mixture model. These new methods have been implemented into a Java package called **MetaQTL**. Finally, as an example, we applied our new approach to

QTL detection results gathered for flowering time in maize.

Results

Meta-analysis of genetic maps

Input genetic map information

Consider a set of n genetic mapping experiments concerning the same linkage group. These different experiments may involve different kinds of population pedigree. We consider that for each experiment $i = 1, \dots, n$ only the estimated distances between ordered markers along the linkage group are available. We use c_i, N_i, M_i to denote the population cross design, the population size and the number of markers on the i^{th} genetic map, respectively. Let's suppose that two markers m_j and m_k have been positioned on the i^{th} map, $\hat{r}_{i,jk}$ stands for the estimated recombination rate between markers m_j and m_k and $\hat{d}_{i,jk} = f[\hat{r}_{i,jk}]$ the corresponding estimated distance, where f is the mapping function which is assumed to be the same in the n mapping experiments (without loss of generality). Applying the classical asymptotic Gaussian distribution of the maximum-likelihood estimation of the parameter, we assume that the $\hat{r}_{i,jk}$ are normally distributed around the true recombination rate $r_{i,jk}$ between markers m_j and m_k with a variance $\text{var}(\hat{r}_{i,jk}) = \eta_{i,jk}^2$. This variance $\eta_{i,jk}^2$ depends on the cross design c_i , the value of $r_{i,jk}$, the sample size N_i and the amount of information supplied by the marker pair m_j and m_k sampled population (see Additional File 1 for expression of η).

Since mapping functions are generally one to one functions, the functional invariance property of the maximum-likelihood estimate can be applied. Thus $\hat{d}_{i,jk}$ is also normally distributed around the true distance denoted $d_{i,jk} = f[r_{i,jk}]$. To obtain the variance of $\hat{d}_{i,jk}$, denoted $\gamma_{i,jk}^2$, we use the first term of the Taylor expansion of the inverse of the mapping function leading to the approximation:

$$\gamma_{i,jk}^2 = \text{var}(d_{i,jk}) \approx \eta_{i,jk}^2 \times \left(\frac{\partial f[r_{i,jk}]}{\partial r} \right)^2$$

Now suppose the n experiments are consistent with the following assumptions:

- Assumption 1 : they come from independent population samples. This implies that $\text{cov}(\hat{r}_{i,jk}, \hat{r}_{i',jk}) = 0$ and $\text{cov}(\hat{d}_{i,jk}, \hat{d}_{i',jk}) = 0$ for any pair of markers m_j and m_k which have been mapped in population i and i' , $i \neq i'$ and $(i, i') \in [1..n]^2$.

- Assumption 2 : there is no interference, i.e in each mapping experiment the recombination events occur independently in each marker interval. This is surely an idealization, but presently, most of the statistical models used to build genetic marker maps are based on this assumption. Thus for a given mapping experiment i , both the ordered marker interval recombination rate and distance estimates are independent, i.e $\text{cov}(\hat{r}_{ij(j+1)}, \hat{r}_{i,(j+1)(j+2)}) = 0$ and $\text{cov}(\hat{d}_{ij(j+1)}, \hat{d}_{i,(j+1)(j+2)}) = 0$ for $i \in [1, \dots, n]$ and $j \in [1, \dots, M_i - 2]$.

- Assumption 3 : the "true" marker order and recombination rate are supposed to be the same in the different populations, i.e $r_{i,jk} = r_{i',jk}$ if markers m_j and m_k have been mapped in population i and i' , $i \neq i'$ and $(i, i') \in [1..n]^2$.

- Assumption 4 : all the genetic maps are connected. Mathematically, this means that if we consider maps as vertices and common markers as edges, then the corresponding graph is supposed to be connected.

Meta-analysis model

We define $\hat{D} = (\hat{d}_{i,jk})$ and $\Gamma = \text{diag}(\gamma_{i,jk}^2)$ the vector of ordered marker interval distance estimates and the diagonal terms of the variance covariance matrix of \hat{D} . We assume that a total of M distinct markers have been mapped in the n populations. The aim of the meta-analysis is to combine all the available information on marker order and positions in order to build a consensus linkage group on which the M markers are positioned. To do so we introduce $Y = (\gamma_1, \dots, \gamma_M)$ the vector of the "true" positions of these M markers on the consensus linkage group, where the γ_i 's can be either positive or negative depending on an arbitrary zero-reference on the chromosome (hereafter we suppose $\gamma_1 = 0$). If the n mapping experiments are consistent with the previous assumptions and assuming that the distances on the linkage group are additive we propose to estimate Y by solving the following linear system:

$$\hat{d}_{i,jk} = \gamma_k - \gamma_j + \epsilon_{i,jk}$$

where $(j, k) \in [1, \dots, M]^2$, $i \in [1, \dots, n]$, $\hat{d}_{i,jk}$ is the distance estimate of the interval between marker m_j and m_k consecutive on the i^{th} experiment, $\gamma_k - \gamma_j$ is the true distance between these markers, $\varepsilon_{i,jk} \sim \mathcal{N}(0, \gamma_{i,jk}^2)$ is the error term. If assumption 4 holds we are ensured that this system has at least one solution. Applying a classical weighted least squares (WLS) strategy, the optimal solution is the one which minimizes the target function,

$$\chi = \sum_{i=1}^n \sum_{jk} \frac{[d_{i,jk} - (\gamma_k - \gamma_j)]^2}{\gamma_{i,jk}^2}$$

Let's introduce the design matrix A such that $\chi = \text{T}(\hat{D} - AY)\Gamma^{-1}(\hat{D} - AY)$. Then the value of Y which minimizes χ is given by:

$$\hat{Y} = (\text{T}A\Gamma^{-1}A)^{-1}(\text{T}A\Gamma^{-1}\hat{D})$$

which is also a maximum-likelihood estimation of Y with variance-covariance matrix given by $(\text{T}A\Gamma^{-1}A)^{-1}$. Thus \hat{Y} gives both the marker positions and the marker order along the consensus linkage group. The goodness-of-fit of the model can be evaluated by the means of a chi-square test as $\chi \sim \chi_{q-M+1}^2$ where q is the length of the vector \hat{D} , i.e the number of marker intervals over the n experiments. As an illustration, let's consider the following idealized scenario : suppose that the n gathered genetic maps share the same markers, i.e $M_i = M$ for $i = 1, \dots, n$. In this simple case the computation of \hat{Y} is straightforward:

$$\begin{cases} \hat{y}_1 & = 0 \\ \hat{y}_{j+1} - \hat{y}_j & = \frac{\sum_{i=1}^n \gamma_{i,j(j+1)}^{-2} \hat{d}_{i,j(j+1)}}{\sum_{i=1}^n \gamma_{i,j(j+1)}^{-2}} \quad j \in [1, \dots, M-1] \end{cases}$$

and χ is the sum of $M - 1$ terms each distributed as a chi-square with $n - 1$ degree of freedoms. This is equivalent to for each marker intervals testing if the distances are homogeneous between populations using a classical test of equal means. This trivial example illustrates how our WLS approach can be used to test for the homogeneity of the recombination rate between several mapping experiments. This can be viewed as an alternative to the M-test devised by [23] when raw data are not available.

Meta-analysis of QTL

Input QTL map information

Suppose that for a given trait a QTL detection has been carried out in the n mapping experiments. The minimal information supplied by the i^{th} QTL experiment consists of a set of estimated positions of the QTL, denoted \hat{x}_{ij} , and the corresponding proportion of variance explained by each QTL, the r-squares values λ_{ij} . Here $j \in [1, \dots, q_i]$ and q_i is the number of QTL detected in the i^{th} mapping experiment on the current linkage group (generally $q_i = 1$ or possibly 2). The confidence intervals (CI) of the \hat{x}_{ij} 's, denoted v_{ij} , can also be reported. The construction of the CI may have been performed by different approaches:

- support interval [24,25].
- likelihood method [26].
- bootstrapping [27,28].

When the CI is not available it is possible to obtain an approximation of the CI by applying the empirical formula proposed by [29]. By means of intensive simulations they showed that for either a backcross or a F_2 population the expected CI, at 95% level, can be expressed as $CI(95) \approx 530/(N\lambda)$ where N is the population size and λ the proportion of variance explained by the QTL. More recently [30] have derived simple analytical equations which are in good agreement with the formula of [29].

Whatever the method used to estimate the uncertainty on the QTL locations, we assume that the \hat{x}_{ij} 's are normally distributed around the true position x_{ij} of the j^{th} QTL: $\hat{x}_{ij} \sim \mathcal{N}(x_{ij}, \sigma_{ij}^2)$ where σ_{ij}^2 is the variance of the estimated position which can be deduced from the confidence interval v_{ij} . For a CI of $\beta\%$ (β depends on the method used to compute the CI), the standard deviation σ_{ij} can be estimated as $\sigma_{ij} = v_{ij}/(2u_\beta)$ where u_β is the double-sided β -percentile of a centered normalized gaussian. This Gaussian approximation based on the classical asymptotic theory has been suggested by [17], even though this is not perfectly correct for QTL with small effects [31].

Furthermore the n QTL mapping experiments are assumed to be consistent with the following assumptions:

- Assumption 1 : they are independent. This can be considered as correct when the individuals measured in the different populations have been generated independently.

Independence between experiments i and i' means independence between \hat{x}_{ij} and $\hat{x}_{i'j}$.

- Assumption 2 : for a given trait there is a finite number of underlying QTL which cosegregate in the mapping experiments: this means that the populations share the same trait determinism with potentially different allelic configurations at the QTL. In other word there is a finite number of true QTL positions on the linkage groups, i.e $\{x_{ij} | (i, j) \in [1, \dots, n] \times [1, \dots, q_i]\}$ can potentially contain redundancy.

In addition to the two previous assumptions we also assume that the detected QTL locations are independent within experiments. This is not really true when the QTL detection does not properly take into account linked QTL. But with the advent of composite interval mapping strategy [32,33] multiple-QTL model can now be fitted by adding properly chosen cofactors which limit the impact of linkage between QTL on the position estimates. Therefore we assume that \hat{x}_{ij} and $\hat{x}_{i'j}$ are independent for all $j \neq j'$.

Pre-processing

The first step is to apply our WLS strategy to the n mapping experiments in order to build a consensus linkage group. Then the QTL locations are projected on the consensus linkage group using a simple scaling rule between the original QTL flanking marker interval and the corresponding one on the consensus chromosome. For a given QTL location the new confidence interval (if available) on the consensus linkage group is computed by taking into account the average scaling between the original and the consensus chromosome. This is done by computing the sum over the common marker intervals of the ratio of the interval lengths weighted by the probability that the QTL position lies in this interval. There are two possible strategies to approximate this probability. The first one relies on a rough approximation using a Gaussian distribution around the most likely position \hat{x}_{ij} of the j^{th} QTL, namely

$$\Pr(\text{QTL } j \text{ in } m) = \frac{\int_{u_m}^{u_{m+1}} \phi[(u - x_{ij})/\sigma_{ij}] du}{\int_0^L \phi[(u - x_{ij})/\sigma_{ij}] du} \text{ where } \phi\{u\} \text{ is the}$$

density function of a centered normalized Gaussian distribution, m is the index of the marker interval, u_m and u_{m+1} are the absolute positions of the flanking markers on the original map of total length L . If the LOD score profile is available, a more accurate strategy can be applied by sub-

stituting ϕ for the density function which best fits the profile.

The meta-analysis model

The purpose of the QTL meta-analysis is to evaluate, for a given trait, the degree of congruency of the QTL detected in the n mapping experiments. By assuming that there is a finite number of true QTL locations, [17] proposed a clustering based approach to both classify the observed QTL and estimate the positions of the underlying QTL. Their method proceeds by testing all the possible QTL combinations and then choosing the one which maximizes a penalized log-likelihood. Although interesting, this method suffers from a categorical repartition of the QTL in the clusters, which is a limit case of Gaussian mixture models. We propose to adopt a similar clustering strategy but with a more standard Gaussian mixture model which allows QTL to be probabilistically distributed into clusters.

In order to lighten the notation we denote by q the total number of observed QTL locations and we ignore the mapping experiment subscripts so that $\hat{X} = (\hat{x}_1, \dots, \hat{x}_q)$ and $\Sigma = (\sigma_1, \dots, \sigma_q)$. Then, let's suppose there are $K \geq 1$ true QTL located at $X^{[K]} = (x_1^{[K]}, \dots, x_K^{[K]})$ which segregate in at least one of the n QTL mapping experiments. Since the QTL position estimates \hat{X} are normally distributed around their true positions, the problem of finding the K underlying true positions can be viewed as a Gaussian mixture problem where the variances of each observation are known. Thus the log-likelihood of the observations can be written as follows:

$$L(X, \Sigma; \Theta^{[K]}) \propto \sum_{i=1}^q \log \left[\sum_{j=1}^K \pi_j^{[K]} \phi \left[(x_i - x_j^{[K]}) / \sigma_i \right] \right] \tag{1}$$

where $\Theta^{[K]} = (X^{[K]}, \Pi^{[K]})$ denotes the parameters of the model, $\Pi^{[K]} = (\pi_1^{[K]}, \dots, \pi_K^{[K]})$ are the mixing proportions, which sum to one, and $\phi\{x\}$ is the density function of a centered normalized Gaussian distribution. We assume without loss of generality that $x_1^{[K]} < x_2^{[K]} < \dots < x_K^{[K]}$ and that $\pi_j^{[K]} \neq 0$, for $j \in [1, \dots, K]$. In other word the distribution of the observed QTL locations is shaped by a mixture density where the components $x_j^{[K]}$ are the positions of the true QTL on the linkage group and the mixing proportions π_j represent the proportion of QTL related to the j^{th}

true QTL which have been detected in the n mapping experiments.

Maximizing 1 can be achieved via a standard EM algorithm [34] by using the following parameter updates (M-step):

$$x_j^{[K]} = \frac{\sum_{i=1}^q t_{ij}^{[K]} x_i \sigma_i^{-2}}{\sum_{i=1}^q t_{ij}^{[K]} \sigma_i^{-2}} \text{ and } \pi_j^{[K]} = \frac{1}{q} \sum_{i=1}^q t_{ij}^{[K]}$$

where $t_{ij}^{[K]}$ is the conditional probability that \hat{x}_i belongs to the j^{th} meta-QTL. This conditional probability $t_{ij}^{[K]}$ is obtained by applying a simple Bayes' rule evaluated at the current parameter estimates (E-step):

$$t_{ij}^{[K]} = \frac{\pi_j^{[K]} \phi \left[(x_i - x_j^{[K]}) / \sigma_i \right]}{\sum_{j'=1}^K \pi_{j'}^{[K]} \phi \left[(x_i - x_{j'}^{[K]}) / \sigma_i \right]}$$

The EM-algorithm is run until reaching convergence: this yields the maximum-likelihood estimate denoted $\tilde{\Theta}^{[K]} = (\tilde{X}^{[K]}, \tilde{\Pi}^{[K]})$. Finally, once $\tilde{\Theta}^{[K]}$ has been obtained the variance-covariance matrix of the parameter estimates, conditionally to the current model, can be computed by applying the Supplemental EM (SEM) strategy proposed by [35].

The problem is that we do not know K , i.e the number of true QTL positions. Since the mixture model of K components is nested into the model with $K + 1$ components, the likelihood ratio test (LRT) should be suitable. However, as discussed by many authors (see for instance [36,37]) the LRT statistic does not follow the usual χ^2 distribution due to testing a null hypothesis on the boundary of the parameter space (i.e the regularity conditions on the loglikelihood do not hold). Another strategy is to use the Kullback-Leibler information in order to derive the information criterion which is widely used to select a statistical model. In particular, the Kullback-Leibler information can be viewed as a measure of goodness-of-fit of a statistical model. Here for a given value of K , minimizing the Kullback-Leibler information is equivalent to maximizing the negentropy KL,

$$KL = - \int_{\mathcal{X}} g(X, \Sigma) L(X, \Sigma; \Theta^{[K]}) dX$$

where $g(\hat{X}, \Sigma)$ is the the true underlying density function. Thus, from the point of view of the negentropy maximization principle, the goodness of the model can be evaluated by the expected log-likelihood. Note that the negentropy maximization principle naturally leads to the maximization of the log-likelihood. However, the maximized log-likelihood is a naive estimate of the expected loglikelihood: since the same data set \hat{X} is used for both the estimation of the parameter and the estimation of the expected log-likelihood, $L(\hat{X}, \Sigma; \Theta^{[K]})$ is a biased estimator of the expected loglikelihood. Its bias is defined by,

$$B = E_{\hat{X}} \left[L(\hat{X}, \Sigma; \tilde{\Theta}^{[K]}) \right] - E_Y \left[L(Y, \Sigma; \tilde{\Theta}^{[K]}) \right]$$

and the use of $L(\hat{X}, \Sigma; \tilde{\Theta}^{[K]}) - B$ is justified as an estimate of KL. There are different strategies to estimate this bias and several information based criteria have been reported in the mixture model literature in order to tackle the issue raised by choosing the number of components (see Additional File 2). In the next *Simulation section*, we propose to evaluate the ability of some of these information based criteria to determine the optimal number of QTL.

Simulation study

For the sake of concision, in this section we only present simulations for the QTL meta-analysis (simulations for the meta-analysis of genetic maps are described in Additional File 3). We assume that the complexity which shapes the distribution of the observed QTL along the chromosome can be represented by our mixture model. In order to explore mixture configurations which are realistic we have assumed that the QTL effects have a L-shaped distribution (i.e most of the detected QTL in mapping experiments have a small effect and only a few show a strong effect, or in other words, most of the detected QTL have large confidence intervals). Consequently this implies that Σ^{-1} , the inverse of the QTL standard deviations, has also a L-shaped distribution (i.e the smaller the effect of the QTL the larger the confidence interval of the estimated QTL position). Then for a given value of the number of true QTL, K , we randomly generated configurations as follows:

1. Draw Σ from a inverse gamma distribution (this simply mimics a L-shaped distribution).

2. Generate the mixing proportions by choosing them over the discrete uniform [0.1, 0.9] distribution subject to

$$\text{constraint } \sum_{k=1}^K \pi_k = 1 .$$

3. Draw from a multinomial, with frequencies equal to the mixing proportions, the origins of the q observed QTL $Z = (z_1, \dots, z_q)$ where $z_{ij} = 1$ if the i^{th} observed QTL belongs to the j^{th} true QTL, 0 otherwise.

4. Generate the true QTL positions, $X = (x_1, \dots, x_k)$, subject to constraint $x_k + \tau_{\min} < x_{k+1} < x_k + \tau_{\max}$ where τ_{\min} and τ_{\max} are defined so that the distance between x_k and x_{k+1} lies between δ_{\min} and δ_{\max} . The distance δ is defined as the mahalanobis distance between x_k and x_{k+1} :

$$\delta = \sqrt{\frac{(x_k - x_{k+1})^2}{a_k^2 + a_{k+1}^2}} \text{ where } a_k = \left(\sum_{i=1}^q z_{ik} \sigma_i \right) / \left(\sum_{i=1}^q z_{ik} \right) \text{ is the}$$

average standard deviation for the k^{th} true QTL. This measures the separation between consecutive true QTL relatively to the precision of the experiments: $\delta \leq 2$ corresponds to tightly or moderately separated QTL, while $\delta \geq 3$ corresponds to well separated QTL.

We stress that this process is not an attempt to describe reality, nevertheless it makes it possible to cover a large range of possible repartitions of the QTL. Finally, for each of the 4 distance constraints considered ($\delta_{\min} = 1, 2, 3, 4$ and $\delta_{\max} = \delta_{\min} + 1$), 50 configurations were generated. For a given value of K , the following scenario was repeated 100 times:

1. Draw a sample \hat{X} of size q .
2. Run the EM-algorithm to obtain $\tilde{\Theta}^{[K]}$ for $K = 1, \dots, q$.
3. Choose the best model according to each criterion.

Since the goal of QTL meta-analysis is to obtain a better predictive inference of the true QTL locations we have compared the two alternative strategies:

- Strategy 1 : choose the model with as many true QTL as the number of observed QTL. It is the naive model, $\bar{x}_i(1) = \hat{x}_i$
- Strategy 2 : choose the best model K according to the model choice criterion, $\bar{x}_i(2) = \sum_{j=1}^K t_{ij}^{[K]} \hat{x}_j^{[K]}$

For each strategy $s = 1, 2$, the measure of performance used was the mean squared error of prediction defined as follows:

$$\text{MSEP}(s) = \frac{1}{q} \sum_{i=1}^q E[(x_i - \bar{x}_i(s))^2]$$

Absolute values of these MSEP are not of interest here because our goal is comparison of strategies; hence, we consider the ratios $\text{MSEP}(2)/\text{MSEP}(1)$ for 5 different information based criteria:

- $\text{AIC} = -2L(\hat{X}, \Sigma; \tilde{\Theta}^{[K]}) + 2\nu$
- $\text{AIC}_c = -2L(\hat{X}, \Sigma; \tilde{\Theta}^{[K]}) + 2\nu + \frac{2\nu(\nu + 1)}{q - \nu - 1}$
- $\text{AIC3} = -2L(\hat{X}, \Sigma; \tilde{\Theta}^{[K]}) + 3\nu$
- $\text{BIC} = -2L(\hat{X}, \Sigma; \Theta) + \nu \log(q)$
- $\text{EIC} \approx \text{AIC} - K + 1$, which was obtained by means of simulations (data not shown).

where $\nu = 2K - 1$ is the number of free parameters of the model and q the number of observed QTL along the chromosome (see Additional file 2 for theoretical details on each above criterion).

In Figure 1 we summarized the result of simulations for several values of K and q by averaging over the distance constraint configurations ($\delta_{\min} = 1, 2, 3$ and 4). At first sight the 5 criteria seem to have the same behavior whatever the configuration, except for AIC3 which crucially underperforms for small values of q (this can be explained by the higher penalty of AIC3 comparing to the other criteria for small values of q). For reasonable sample size relatively to the true number of components the meta-analysis appears to be more efficient than strategy 1. Since the AIC criterion has relatively good performance in these simulations we assume that there is no need for a specific theory to deal with this kind of mixture models and that this criterion can be used to carry out model selection in this context. So, in Figure 2 we focus on the AIC criterion for the different distance configurations $\delta_{\min} = 1, 2, 3$ and 4 . This clearly shows that, for configurations with reasonable separation between the true positions of the QTL, the meta-analysis performs relatively well. It is worth noting that the better the probability to choose the true model, the better the quality of the QTL position estimates. In order to evaluate the ability of the meta-analysis to improve the precision on the "true" QTL locations we computed the quantities $|x_i - \hat{x}_i(s)|$ and calculated the

quantiles at 95 and 90% of its empirical distribution over all the QTL for the two strategies. The smaller this confidence interval, the better the estimated position $\hat{x}_i(s)$. We reported in Supplementary Table 1 (see Additional File 4) the average ratios of these quantities between the two strategies. Hence, if there are actually one, two, three or four different QTL locations with a reasonable separation ($\delta_{min} \geq 2$), we can see that the meta-analysis gives better estimates of the QTL locations and makes it possible to reduce the length of the 95% CI (in most of the situations this length is halved). According to [29] to halve a CI in a QTL experiment, one needs to use at least two times the initial number of individuals.

Implementation

The previous methods have been implemented into a Java package called **MetaQTL**. This package includes also additional programs to format, organize or visualize the data. All the programs in **MetaQTL** are command line programs (see Supplementary Table 2 in Additional File 4 for a complete list of the programs available in the package). Each program performs a specific task and the programs can be combined by the user as a group to run a complete analysis. Thanks to its flexible and modular implementation, **MetaQTL** could also be integrated in more elaborated softwares if needed. First, before running meta-analysis one needs to store the different QTL studies into a database. To do this **MetaQTL** uses a simple multiple plain text files database. Each file corresponds to a table and the database is organized as follows:

- Experiment file: stores descriptions on mapping experiments (name, population type and size, reference, ...).
- Genetic map directory : contains one file per input marker map. Each file contains the corresponding genetic marker map.
- QTL map directory : contains one file per QTL mapping experiments. For each QTL mapping experiment the file provides the properties of each detected QTL (trait, position, confidence interval, r-square, ...).
- Trait ontology file: describes how the traits are related together using a simple hierarchical relationship scheme. This information can then be used to group the QTL according to the ontology in subsequent analyses.

Once the database created, **MetaQTL** first checks the input data files and then summarizes their content into a set of XML files. All the programs of **MetaQTL** use these XML files as inputs. Utilities are provided to convert them in various plain text file formats if required (for more details

on using **MetaQTL** see the user manual in Additional File 5).

Application

Recently, [12] made a bibliographical review of QTL studies relative to 4 traits related to flowering time in maize: days to pollen shed (DPS), silking date (SD), plant height (HT) and leaf number (LN). From the 22 QTL studies they reported, we excluded 6 experiments for which QTL detection was based on ANOVA with a low density of markers and 2 other for which it was not possible to get exact information on either the genetic linkage map or the QTL locations. In addition to these 15 mapping experiments we considered 3 other recent experiments (details of these 18 QTL studies are given in Additional File 6). We focus here on chromosome 8 and we present results by using for each step of the meta-analysis the corresponding program name of **MetaQTL**.

Result of InfoMap

Among the 153 distinct markers which have been positioned over the 18 mapping experiments on the chromosome 8, only 53 markers are observed in at least two different mapping experiments. We restricted the meta-analysis to these 53 markers. Only one order inconsistency was detected between [38] and [39] concerning markers *umc89a* and *umc12a*. As in [38] *umc12a* is very close to *umc89a* (less than 2 cM) we have decided to ignore this marker in this mapping experiment. Over the 18 mapping experiments the mean interval distance was about 18.9 cM with an average of 8.7 markers per mapping experiment and it existed at least one common marker path which connected all the mapping experiments together (insuring that the WLS can be applied).

Result of ConsMap

The consensus linkage group of chromosome 8 is depicted in Figure 3. The goodness-of-fit of the consensus chromosome is relatively bad: $\lambda = 365.31$ with $\lambda \sim \chi_{87}^2$. It could be due to some heterogeneities in recombination rate among mapping experiments, located in the filled marker intervals of Figure 3. Note that variability of recombination rate in maize was first reported by [40] and, more recently, [41] demonstrated that exotic inbred lines exhibit higher recombination rate than U.S. inbreds origin along chromosome 1 (see also [42]). On the other hand, since no information about the marker configurations in each individual mapping experiment was available, the variances of the distance estimates have been computed by assuming no missing data and no ambiguities (dominance) in the original data sets. This is surely too optimistic and some data sets may have included missing data and/or dominant markers. Therefore the pre-

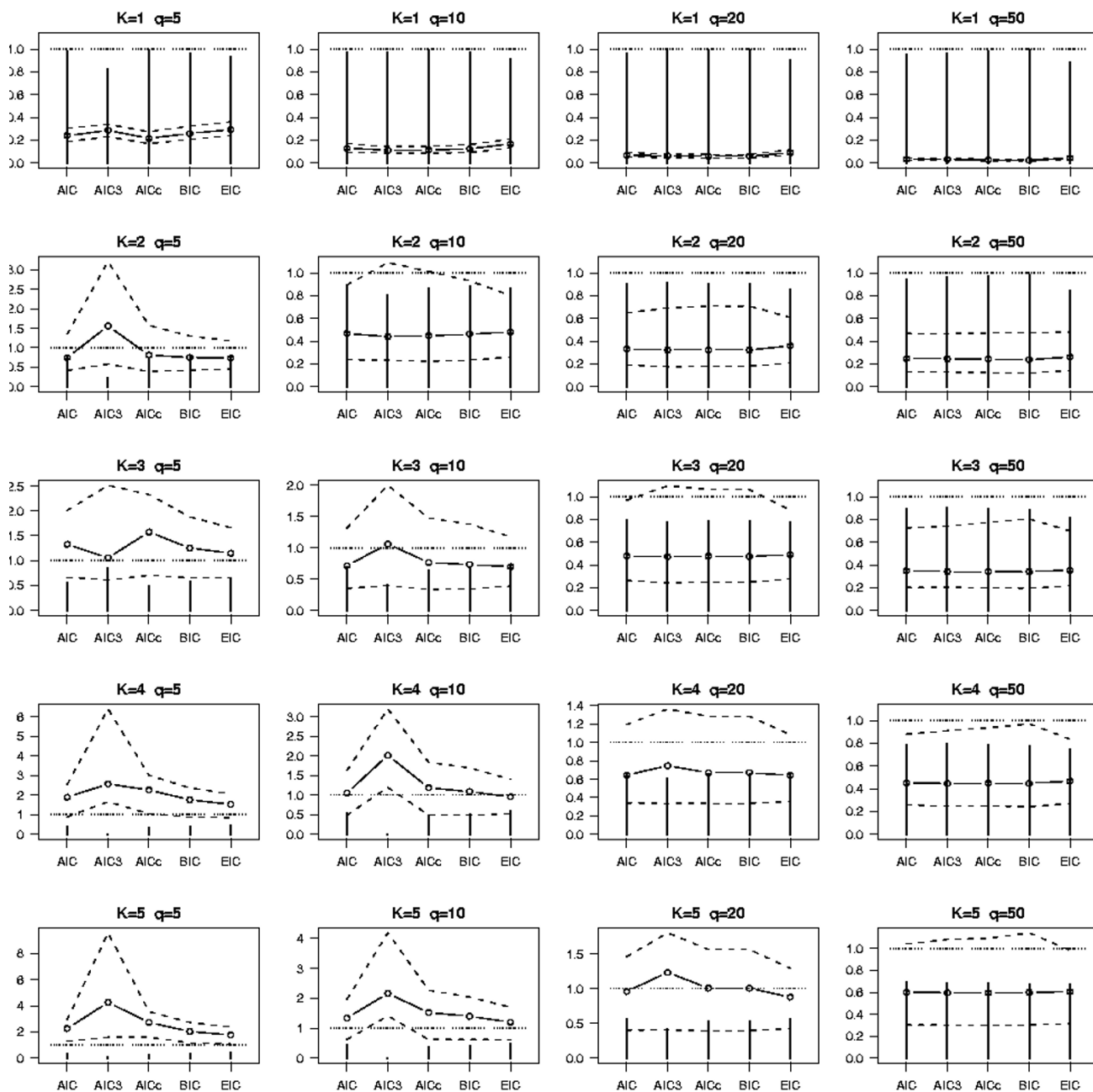


Figure 1
Comparison of model choice criteria: a simulation study. Simulation results for different values of the true number of QTL, K , and the number of observed QTL, q . The vertical bars indicate the probability that the best model selected by the criterion is the true model. The open circles, respectively the dotted lines, represent the mean, respectively the 0.1% and 0.9% quantiles, of the ratios $MSEP(2)/MSEP(1)$ for each criterion.

cision on the distance estimate may have been overestimated for some marker intervals.

Result of QTLProj

From the 18 QTL studies we projected 34 QTL on the consensus chromosome 8. Among these 34 QTL, 16 (47%)

are related to SD, 10 (29%) to DPS and 8 (24%) to HT. The distribution of the r-square values clearly shows a L-shape: 75% of the QTL have r-square values lower than 12%. For 17 QTL a CI was reported (build from a 1-LOD support) from which we computed the standard devia-

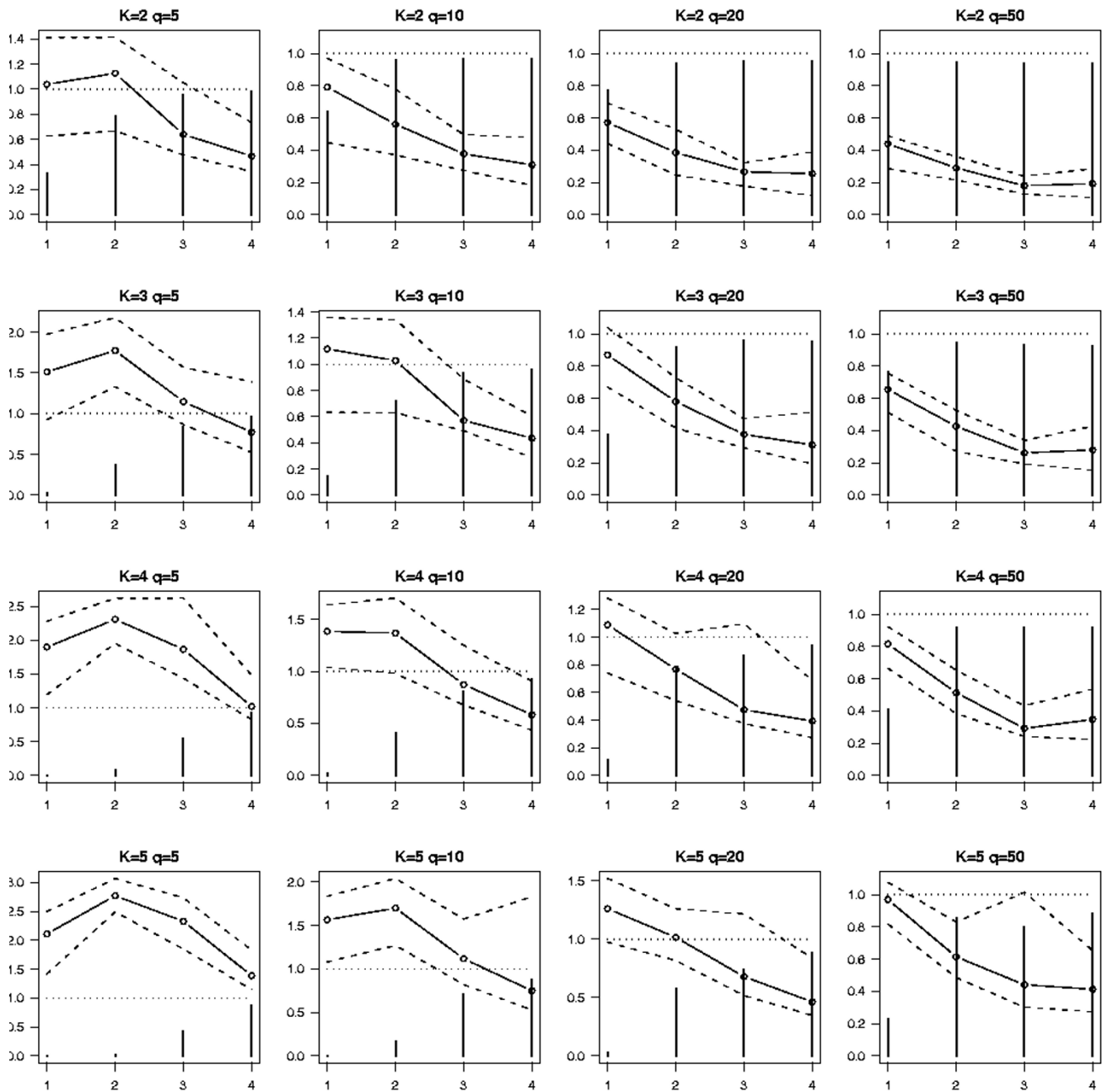


Figure 2
Performance of AIC: a simulation study. Behavior of the AIC criterion for the 4 distance constraints, $\delta_{\min} = 1, 2, 3, 4$, and for different values of the true number of QTL, K , and the number of observed QTL, q . The vertical bars indicate the probability than the AIC criterion has selected the true model. The open circles, respectively the dotted lines, represent the mean, respectively the 0.1% and 0.9% quantiles, of the ratios $MSEP(2)/MSEP(1)$.

tions assuming that a 1-LOD support corresponds in fact to a 90% CI. For the other QTL we derived the standard deviations from the formula proposed by [29]. Then models from $K = 1$ to $K = 10$ QTL were considered and their parameters estimated by applying our EM-algorithm.

Result of QTLClust

In Table 1 we give the values for the criteria AIC, AIC_c , AIC3 and BIC for the different values of K explored. This clearly shows that the model with 5 QTL is the best one. Then, for the model with 5 QTL, the parameter estimates

Table 1: Model choice criteria for the meta-analysis of flowering time QTL on maize chromosome 8

K	Model Choice Criterion			
	AIC	AIC _c	AIC3	BIC
1	1096.32	1088.94	1088.32	1084.11
2	497.22	490.52	491.22	488.06
3	139.15	133.79	135.15	133.05
4	34.73	31.53	32.73	31.67
5	0.00	0.00	0.00	0.00
6	4.00	8.50	6.00	7.05
7	8.00	18.70	12.00	14.11
8	12.00	31.17	18.00	21.16
9	16.00	46.75	24.00	28.21
10	19.54	66.33	29.54	34.81
34	51.42	43.92	76.42	89.58

All the criteria select the model with 5 QTL (in bold). The values are adjusted with respect to the minimal value of the criterion, i.e. the value for K = 5.

are listed in Table 2 and depicted in Figure 4. First, 3 QTL (1,4 and 5) have been detected in only 22% of the mapping experiments. At least two observed QTL are assigned to each of these 3 QTL without ambiguity.

Secondly, two closely linked QTL (2 and 3) contribute to 75% of the reported QTL. This is strongly consistent with the knowledge of this region where a major QTL, *vgt1*, is tightly linked to another QTL, *vgt2* [43,44]. It is worth noting that the confidence interval of the QTL corresponding to *vgt1* (around 3.8 cM) encompasses a marker interval of approximately 2 cM (at the left of the marker *umc89a*) in which this QTL has been finely mapped by [44] using NIL lines (result not included in our analysis). This congruency lends further credence to the meta-analysis approach.

Discussion and conclusion

Nowadays more and more studies concerning QTL detection are available via public databases and the number of articles dealing with the comparison and/or integration of these results increases [12,45-47]. We believe that our meta-analysis procedure can contribute to facilitate the elaboration of such syntheses by providing a simple statis-

tical framework to establish consensus models for both linkage maps and QTL locations.

First, the WLS strategy we proposed is a step forward to integrate several genetic marker maps. Contrary to iterative projection procedures, this approach provides a well-established statistical machinery (WLS) to assess the goodness-of-fit of the consensus model. It can also be used to test the homogeneity of the distance estimates among different mapping experiments. This can be useful to investigate the possible variation of recombination rate among genotypes (as reported by [41]). As pointed out in the application, this method can suffer from the lack of knowledge about the effective precision on the marker interval distances in each individual mapping experiment due to possible missing data and/or the type of scoring of individual markers (codominance vs dominance). This could be improved by asking researchers to supply the variance estimates of the marker interval distances when they submit their results to a public database. These variance estimates could be used to improve the weight factors in the WLS model. Also, as sometimes robust framework maps are available in the literature or via public databases, the program ConsMap in MetaQTL offers the possibility

Table 2: Parameter estimates of the best meta-analysis model for flowering time on maize chromosome 8

QTL	Position \tilde{X}	Weight $\tilde{\Pi}$	Mahalanobis distance to next QTL	95% CI
1	14.6	0.06	6.21	11.7
2	75.4	0.35	1.26	6.2
3	89.5	0.44	2.64	3.8
4	114.5	0.07	5.20	11.1
5	165.2	0.09	-	13.9

The positions and the lengths of the CI are given in cM. The 95% CI is derived from the conditional variance estimate of the position obtained by applying the SEM strategy. Note that QTL 2 and 3 correspond to *vgt2* and *vgt1*, respectively (following the designation of [43]).

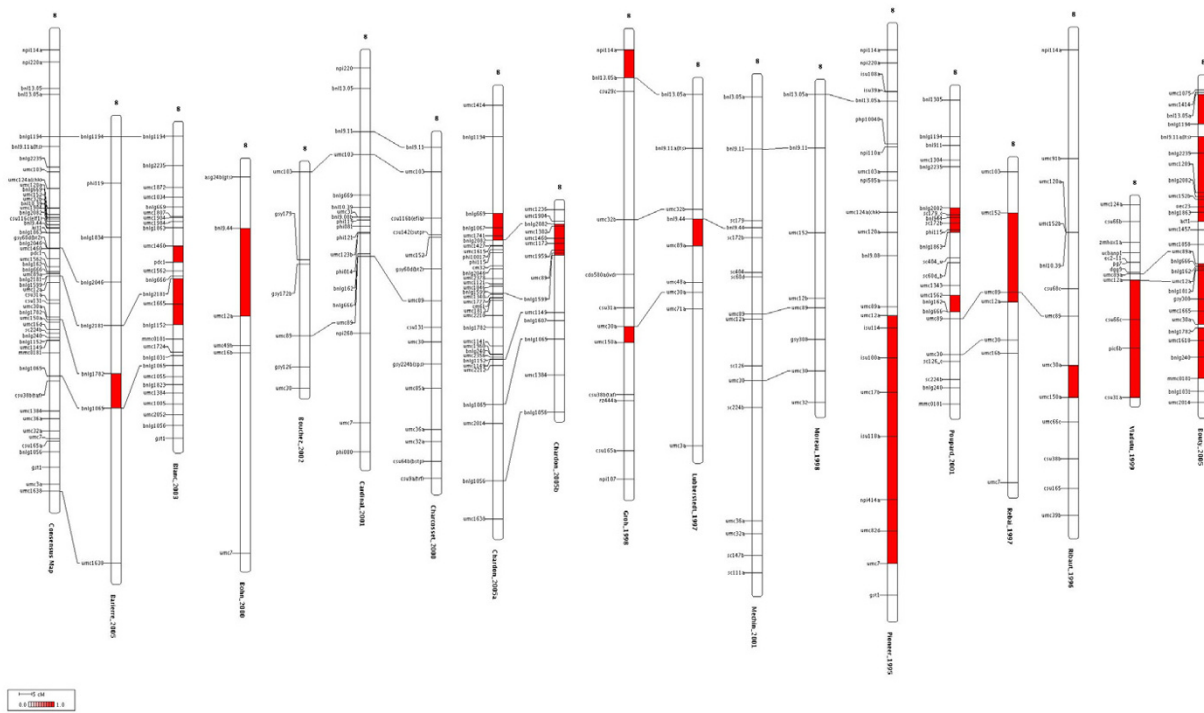


Figure 3
Overview of the maize chromosomes 8 together with the consensus chromosome. Overview of chromosome 8 for the 18 mapping experiments involved in the meta-analysis of flowering time in maize. The first chromosome at the left represents the consensus chromosome obtained by applying the WLS approach as described in the first section of the article (implemented into ConsMap). The filled marker intervals indicate that the standardized residual between the interval distance estimates of the original chromosome and the consensus one exceeded the double-sided 95% percentile of a normalized centered gaussian. This figure has been created by the program MMapView.

to fix a genetic map as a reference (i.e for which the distances between ordered markers are assumed to be the "actual" distances). In this case, only the positions of the markers which are not reported on the reference have to be estimated.

Secondly, for the QTL meta-analysis itself, the Gaussian mixture model used to fit the distribution of the observed QTL locations on the chromosome provides a well-studied statistical inference technique. In this model-based clustering, each "true" QTL is mathematically represented by the Gaussian distribution of its detected positions, which leads to a probabilistic classification of the observed QTL. Contrary to [17] who developed a specific model choice criterion, our simulation results show that AIC gives relatively good performances in our QTL meta-analysis framework. This difference, with regard to the conclusions of [17], may be explained by their discrete formulation of the problem (recall that, instead of using a usual Gaussian mixture likelihood to evaluate the proba-

bility of the data, they assumed that the observations could be categorically assigned to the mixture components). Parameter estimates obtained by this approach were not really the maximum-likelihood estimates of the underlying mixture model. This may have added a bias in the evaluation of the AIC criterion, which could explain the bad performances of AIC in their simulations.

Thus, our mixture-modelling approach makes it possible to go beyond the limits encountered by [17]: the Akaike like criterion they proposed was limited to models from 1 to 4 QTL. As a consequence, [47] who used the method of [17], was obliged to break chromosomes on distinct segments to carry out the meta-analysis. This subjective division of the chromosome can now be avoided thanks to our method. Simulations have shown that the ratio between the number of observed QTL and the number of "true" QTL is one of the main limiting factor. The number of "true" QTL which can be assessed by the meta-analysis must be reasonable compared to the number of observed

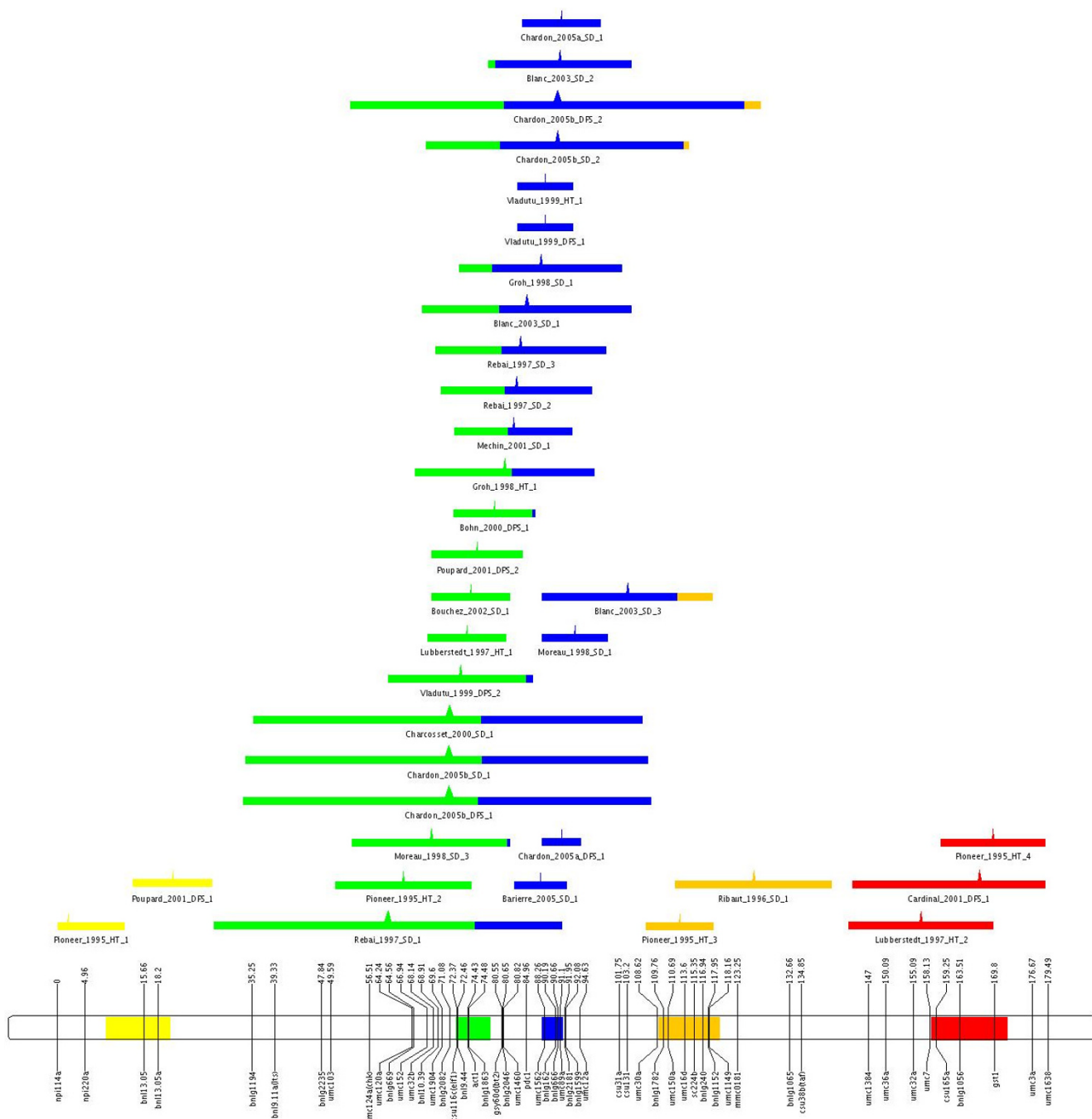


Figure 4
Visualization of the QTL meta-analysis result on maize chromosome 8. Result of the meta-analysis of the 34 QTL projected on the consensus chromosome 8. The CI of the meta-QTL positions are indicated on the chromosome by the filled colored areas. The observed QTL positions are depicted by their most probable position (triangle) and CI (segment). Membership probabilities of each initial QTL with respect to meta-QTL is visualized by the proportions of corresponding colored segments. This figure has been created by using MQTLView.

QTL (at least between 5 or 10 observed QTL per actual location). Note that this also depends on the distance between true QTL. But since there are more and more QTL locations reported for a given trait and since the real number of distinct QTL locations which can be detected with usual experimental designs is limited (only QTL with relatively large effects can be found), we assume that in many cases the ratio between observed and "true" QTL locations will steadily increase and should generally be reasonable. It is worth noting that, provided that the number of observed QTL is appropriate, the meta-analysis is able to separate "true" QTL locations even if they are closely linked (as illustrated in the application with *vgt1* and *vgt2*, and the consistency of the *vgt1* estimated position with fine mapping result of [44], result not included in the meta-analysis).

The ultimate step toward a more accurate identification of QTL relies on finding the underlying genes. Up to now, the majority of QTL isolated in plants have been cloned via positional cloning (see for instance [48]). However positional cloning of QTL is quite expensive both in terms of time and resources due to the necessity to screen recombinant individuals within large population (typically several hundreds) and to characterize these individuals with a very dense set of molecular markers. As an alternative and thanks to the advent of structural and functional genomics, QTL can also be resolved through association mapping of candidate genes. Candidate genes identification is based on a assumption that the polymorphism of the gene is associated with the variation of the trait of interest. Both function and mapping information have to be crossed to establish this assumption. The function of the gene may have been determined in the species of interest, based for instance on mutant analysis. More often, function is hypothesized based on sequence homology with genes the function of which has been established in model species, including possible positional cloning of QTL. Gene mapping information may have been obtained in the species of interest, but may have been also inferred from synteny based projections, as illustrated by [12] for rice to maize. Relevancy of the colocalization between QTL and candidate genes crucially depends on the confidence interval of the QTL positions. For this purpose the reduction of the confidence interval of the QTL is an important goal [2]. The ability of our method to reduce the QTL confidence interval by taking advantage of pooling QTL results could contribute in an increased resolution in selecting candidate genes. It is worth noting that candidate genes are generally mapped on a framework map used as reference for the species of interest (e.g. in maize [49]), while the QTL detections are carried out using specific populations (generally obtained by crossing parents contrasted for the trait(s) of interest).

Therefore, the selection of candidate genes which colocalize with QTL depends also on the process used to merge these different maps. Up to now, no statistical method had been proposed to combine candidate genes and QTL mapped in independent experiments, so we think that our WLS strategy should increase the precision of the integration of candidate gene mapping information.

Finally once candidate genes have been selected and their different haplotypes defined, association studies can be carried out. The identification of a statistically significant association between haplotype variation at a candidate gene and the target trait gives further credence on the role of this gene in the trait variation. Since the last 5 years more and more association studies have been reported in plants [50]. It would be interesting to integrate these new results into a global meta-analysis framework. Further developments are needed to combine onto a synthetic model the different scale of mapping: from linkage mapping (QTL) to fine mapping (association studies).

Availability and requirements

- Project name: MetaQTL
- Project home page: <http://bioinformatics.org/mqtl>
- Operating system: Platform independent
- Programming language: Java
- Other requirements: Java 5.0
- License: GNU GPL

Authors' contributions

B.G and A.C conceived the initial idea. J-B.V. developed and designed the project: he coded the package MetaQTL, developed the simulation tests, applied MetaQTL on the flowering time maize data set. All authors contributed to the interpretation of the results. Then J-B.V. and A.C. drafted the manuscript. All authors read and approved the final version of the manuscript.

Additional material

Additional file 1

Theory and method for the computation of the variance of the recombination rate estimate. This PDF file contains a short review of the theory and the methods to compute the variance of the recombination rate estimator for different kinds of pairwise marker configurations and different types of mapping experiments.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-49-S1.pdf>]

Additional File 2

Model choice by information based criterion. This PDF file deals with a short review of the theory underlying some well-known information based criteria to select the number of components in a mixture model.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-49-S2.pdf>]

Additional File 3

Simulation study for the meta-analysis of genetic maps. This PDF file describes two simulation scenarios which have been used to evaluate our weighted least squares strategy to build a consensus marker map.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-49-S3.pdf>]

Additional File 4

Supplementary Tables. This PDF file contains the two supplementary tables.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-49-S4.pdf>]

Additional File 5

Documentation of MetaQTL. The documentation of the Java package MetaQTL in PDF format.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-49-S5.pdf>]

Additional File 6

Summary of the QTL mapping experiments used in the application. This PDF file gives details on the mapping experiments used in our application.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-49-S6.pdf>]

Additional File 7

MetaQTL Package : jar file and tutorial. This Zip archive contains both the MetaQTL JAR file and the files of the tutorial.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-49-S7.zip>]

Acknowledgements

This work was supported by INRA grant and the authors would like to thank F. Chardon, from INRA, for his help on the gathering process of QTL results relative to flowering time in maize. We would like also to thank C. Lepointevin for her help in testing and debugging **MetaQTL**. We also thank Graham Coop for his help to improve the style of the manuscript.

References

- Kearsey M, Pooni HS: *The genetical analysis of quantitative traits* London: Chapman and Hall; 1996.
- Kearsey M, Farquhar A: **QTL analysis in plants; where are we now?** *Heredity* 1998, **80**:137-142.
- Paran I, Zamir D: **Quantitative traits in plants: beyond the QTL.** *Trends Genet* 2003, **19(6)**:303-306.
- Xu S: **Theoretical basis of the Beavis effect.** *Genetics* 2003, **165(4)**:2259-68.
- Beavis W: **The power and deceit of QTL experiments: lessons from comparative QTL studies.** In *Proceedings of the Forty-Ninth Annual Corn and Sorghum Industry Research Conference* Washington, DC: American Seed Trade Association; 1994:250-266.
- Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD: **Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms.** *Nature* 1988, **335**:521-529.
- Lin YR, Schertz KF, Paterson AH: **Comparative analysis of QTLs affecting plant height and maturity across the poaceae, in reference to an interspecific sorghum population.** *Genetics* 1995, **141**:391-411.
- Paterson AH, Lin YR, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE: **Convergent domestication of cereal crops by independent mutations at corresponding genetic loci.** *Science* 1995, **269**:1714-1718.
- Keightley P, Knott S: **Testing the correspondence between map positions of quantitative trait loci.** *Genet Res Camb* 1999, **74**:323-328.
- Mihaljevic R, Utz HF, Melchinger AE: **Congruency of quantitative trait loci detected for agronomic traits in testcrosses of five populations of european maize.** *Crop Sci* 2004, **44**:114-124.
- Khatkar M, Thomson P, Tammen I, Raadsma H: **Quantitative trait loci mapping in dairy cattle: review and meta-analysis.** *Genet Sel Eval* 2004, **36**:163-190.
- Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A: **Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome.** *Genetics* 2004, **168**:2169-2185.
- Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A, Joets J: **BioMercator: integrating genetic maps and QTL towards discovery of candidate genes.** *Bioinformatics* 2004, **20(14)**:2324-2326.
- Yap I, Schneider D, Kleinberg J, Matthews D, Cartinhour S, McCouch S: **A graph-theoretic approach to comparing and integrating genetic, physical and sequence-based maps.** *Genetics* 2003, **165**:2235-2247.
- Stam P: **Construction of integrated genetic linkage maps by means of a new coputer package: JoinMap.** *Plant J* 1993, **3**:739-744.
- Schiex T: **Carthagene: constructing and joining maximum likelihood maps.** *ISMB* 1997, **5**:258-267.
- Goffinet B, Gerber S: **Quantitative trait loci: a meta-analysis.** *Genetics* 2000, **155**:463-473.
- Allison DB, Heo M: **Meta-analysis of linkage data under worst-case conditions: a demonstration using the human OB region.** *Genetics* 1998, **148**:859-865.
- Van Zandt P, Mopper S: **A meta-analysis of adaptive deme formation in phytophagous insect populations.** *Am Nat* 1998, **152**:595-604.
- Vollestad LA, Hindar K, Moller AP: **A meta-analysis of fluctuating asymmetry in relation to heterozygosity.** *Heredity* 1999, **83**:206-218.
- Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN: **Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease.** *Nat Genet* 2003, **33**:177-182.
- Etzel C, Guerra R: **Meta-analysis of genetic-linkage of quantitative trait loci.** *Am J Hum Genet* 2003, **71**:56-65.
- Morton NE: **The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type.** *Am J Hum Genet* 1956, **8(2)**:80-96.
- Conneally P, Edwards J, Kidd K, Lalouel J, Morton N: **Reports of the committee methods of linkage analysis and reporting.** *Cytogenet Cell Genet* 1985, **40**:356-359.
- Lander ES, Botstein D: **Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps.** *Genetics* 1989, **121**:185-199.
- Dupuis J, Siegmund D: **Statistical methods for mapping quantitative trait loci from a dense set of markers.** *Genetics* 1999, **151**:373-386.
- Darvasi A, Weinreb A, Minke V, Weller J, Soller M: **Detecting marker-QTL linkage and estimating QTL gene effect and**

- map location using a saturated map. *Genetics* 1993, **134**:943-951.
28. Kao C, Zeng Z, Teasdale R: **Multiple Interval Mapping for Quantitative Trait Loci.** *Genetics* 1999, **152**:1203-1216.
 29. Darvasi A, Soller M: **A simple method to calculate resolving power and confidence interval of QTL map location.** *Behav Genet* 1997, **27**(2):125-132.
 30. Visscher PM, Goddard ME: **Prediction of the confidence interval of quantitative trait loci.** *Behavior Genetics* 2004, **34**(4):477-482.
 31. Mangin B, Goffinet B, Rebaï A: **Constructing confidence intervals for QTL location.** *Genetics* 1994, **138**:1301-1308.
 32. Jansen RC: **Interval mapping of multiple quantitative trait loci.** *Genetics* 1993, **135**:205-211.
 33. Zeng ZB: **Precision mapping of quantitative trait loci.** *Genetics* 1994, **136**(4):1457-1468.
 34. Dempster A, Laird N, Rubin D: **Maximum Likelihood from incomplete data via the EM algorithm (with discussion).** *J Roy Stat Soc* 1977, **B 39**:1-38.
 35. Meng X, Rubin DB: **Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm.** *J Am Stat Assoc* 1991, **86**:899-909.
 36. Aitkin M, Rubin D: **Estimation and Hypothesis Testing in Finite Mixture Models.** *Journal of the Royal Statistical Society* 1985, **47**:67-75.
 37. Titterton D, Smith A, Markov U: *Statistical Analysis of Finite Mixture Distributions* New York: John Wiley and Sons; 1985.
 38. Poupard B, Moreau L, Charcosset A: **Analyse de l'épistatsie entre QTL pour 3 caractères agronomiques chez le maïs.** *Tech rep, INRA* 2001.
 39. Mechin V, Argillier O, Hebert Y, Guingo E, Moreau L, Charcosset A, Barriere Y: **Genetic analysis and QTL mapping of cell wall digestibility and lignification in silage maize.** *Crop Sci* 2001, **41**:690-697.
 40. Stadler LJ: **The Variability of Crossing Over in Maize.** *Genetics* 1925, **11**:1-37.
 41. Williams CG, Goodman MM, Stuber CW: **Comparative recombination distances among Zea mays L. inbreds, wide crosses and interspecific hybrids.** *Genetics* 1995, **141**(4):1573-1581.
 42. Ji Y, Stelly DM, De Donato M, Goodman MM, Williams CG: **A candidate recombination modifier gene for Zea mays L.** *Genetics* 1999, **151**(2):821-830.
 43. Vladutu C, McLaughlin J, Phillips RL: **Fine mapping and characterization of linked quantitative trait loci involved in the transition of the maize apical meristem from vegetative to generative structures.** *Genetics* 1999, **153**(2):993-1007.
 44. Salvi S, Tuberosa R, Chiapparino E, Maccaferri M, Veillet S, van Beuningen L, Isaac P, Edwards K, Phillips RL: **Toward positional cloning of Vgt1, a QTL controlling the transition from the vegetative to the reproductive phase in maize.** *Plant Mol Biol* 2002, **48**(5-6):601-13.
 45. Khavkin E, Coe EH: **Mapped genomic locations for developmental functions and QTLs reflect concerned groups in maize (Zea mays L.).** *Theor Appl Genet* 1997, **95**:343-352.
 46. Khavkin E, Coe EH: **The major quantitative trait loci for plant stature, development and yield are general manifestations of developmental gene clusters.** *Maize Newslett* 1998, **72**:60-66.
 47. Chardon F, Hourcade D, Combes V, Charcosset A: **Mapping of a spontaneous mutation for early flowering time in maize highlights contrasting allelic series at two-linked QTL on chromosome 8.** *Theor Appl Genet* 2005, **112**:1-11.
 48. Salvi S, Tuberosa R: **To clone or not to clone plant QTLs: present and future challenges.** *Trends Plant Sci* 2005, **10**(6):297-304.
 49. Falque M, Decousset L, Dervins D, Jacob AM, Joets J, Martinant JP, Raffoux X, Ribiere N, Ridel C, Samson D, Charcosset A, Murigneux A: **Linkage mapping of 1454 new maize candidate gene Loci.** *Genetics* 2005, **170**(4):1957-1966.
 50. Gupta PK, Rustgi S, Kulwal PL: **Linkage disequilibrium and association studies in higher plants: present status and future prospects.** *Plant Mol Biol* 2005, **57**(4):461-485.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

