Structural mapping: how to study the genetic architecture of a phenotypic trait through its formation mechanism

Chunfa Tong*, Lianying Shen*, Yafei Lv, Zhong Wang, Xiaoling Wang, Sisi Feng, Xin Li, Yihan Sui, Xiaoming Pang and Rongling Wu

Submitted: 28th July 2012; Received (in revised form): 3rd September 2012

Abstract

Traditional approaches for genetic mapping are to simply associate the genotypes of a quantitative trait locus (QTL) with the phenotypic variation of a complex trait. A more mechanistic strategy has emerged to dissect the trait phenotype into its structural components and map specific QTLs that control the mechanistic and structural formation of a complex trait. We describe and assess such a strategy, called structural mapping, by integrating the internal structural basis of trait formation into a QTL mapping framework. Electrical impedance spectroscopy (EIS) has been instrumental for describing the structural components of a phenotypic trait and their interactions. By building robust mathematical models on circuit EIS data and embedding these models within a mixture model-based likelihood for QTL mapping, structural mapping implements the EM algorithm to obtain maximum likelihood estimates of QTL genotype-specific EIS parameters. The uniqueness of structural mapping is to make it possible to test a number of hypotheses about the pattern of the genetic control of structural components. We validated structural mapping by analyzing an EIS data collected for QTL mapping of frost hardiness in a controlled cross of jujube trees. The statistical properties of parameter estimates were examined by simulation studies. Structural mapping can be a powerful alternative for genetic mapping of complex traits by taking account into the biological and physical mechanisms underlying their formation.

Keywords: electrical impedance spectroscopy; QTL mapping; structural components; complex phenotype

*These authors contributed equally to this work.

Corresponding authors. Rongling Wu, Center for Statistical Genetics, Pennsylvania State University, Horshey, PA 17033, USA. Tel: +1-717-531-2037; Fax: +1-717-531-0480; E-mail: rwu@phs.psu.edu; Xiaoming Pang, Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China. Tel: +86-10-6233-6283; Fax: +86-10-6233-6164; E-mail: xmpang@bjfu.edu.cn

Chunfa Tong is a visiting scholar working at Beijing Forestry University and The Pennsylvania State University. His research interest is in genetic mapping and modeling of complex traits.

Lianying Shen is a Professor of Horticultural Sciences at Hebei Agricultural University. She develops jujube cultivars that adapt to the environment in North China.

Yafei Lv is a Master’s student in Forest Genetics and Tree Breeding in the Center for Computational Biology at Beijing Forestry University. His research focuses on the statistical genetics of quantitative traits in forest trees.

Zhong Wang is a computer scientist at The Pennsylvania State University. He develops computer algorithms for statistical genetic models.

Xiaoling Wang is a PhD candidate in horticultural sciences at Hebei Agricultural University. Her thesis topic is on genetic mapping of jujube trees.

Sisi Feng is a Master student in forest genetics at Beijing Forestry University. Her thesis topic is on the development of computational tools to study population genetics.

Xin Li is a Master student in forest genetics at Beijing Forestry University. Her thesis topic is on the development of computational tools to study quantitative genetics.

Yihan Sui is a fourth-year undergraduate student in biological sciences at Beijing Forestry University. Her research interest is in the development of statistical and bioinformatics approaches to study the genetic control of complex traits and diseases.

Xiaoming Pang is an Associate Professor of Tree Breeding at Beijing Forestry University. His research interest focuses on the utilization of biotechnologies to study population genetic diversity in fruit trees.

Rongling Wu is a Professor of Biostatistics and Bioinformatics and the Director of the Center for Statistical Genetics at The Pennsylvania State University. He founded the Center for Computational Biology at Beijing Forestry University. He is interested in quantitative and statistical genetics.
INTRODUCTION

The final phenotype of a trait is the consequence of interactions of its internal and external structures and processes [1, 2]. It is such a phenotype that forms the features of an organism, allowing it to survive and interact with its environment. A central theme of biology is to study the intimate association of the phenotype’s structure or form and its function or activity [3]. Structure is the basic component of the phenotype that can be understood at the molecule, cell or tissue level, whereas function refers to the capacity of how different structures are used. In order to identify the function of a phenotype, therefore, we first need to understand how the shape and properties of a structure determine its functions [4, 5].

Traditional approaches for mapping quantitative trait loci (QTLs) that control a phenotype do not consider the structural basis of the phenotype, thus failing to relate the structure of the phenotype with its function. The genetic mapping of complex traits can be improved through the dissection of the phenotypes into the underlying structural components. By measuring the structure of organic and inorganic materials, electrical impedance spectroscopy (EIS) has been increasingly used to study the function and properties of biological issues [3, 6–11]. The EIS method is based on the physical principle that the signal of alternating current (AC) will change in its amplitude and phase when it passes through a tissue, arising from the polarization and relaxation of the tissue. Based on those changes, the impedance of the tissue that is composed of a real (resistance) and an imaginary part (reactance) in a complex plane can be determined. When the real and imaginary part is measured at different frequencies, an impedance spectrum is observed [9]. Because the proportion of the AC passing through the tissue depends on its frequency and the tissue properties, such as plasma membranes, cell volumes and intracellular conductivities, the electrical impedance of the tissue at a series of frequencies provides information about the cell population.

It has been observed that intracellular resistance is highly correlated with physiological and pathological variables of tissues, such as frost hardness [8, 9] and cancer risk [3, 10], through changes in cellular features and processes. With a proper equivalent electrical model, it is possible to study the tissue properties according to the changes in the parameters of the model [12]. In this article, we describe and assess a computational model—structural mapping—for mapping specific QTLs that determine the function of complex traits using EIS parameters that quantify tissue properties. In the EIS analysis, a number of robust mathematical models have been established to describe the dynamic properties of impedance through the tissue [7, 13, 14]. These models are integrated with genetic mapping derived to map dynamic QTLs [15–17], characterizing the genetic effects of QTLs on the trait by testing the mathematical parameters. Structural mapping is shown to be powerful for mapping internal structural properties of phenotypic formation and can be used to study the genetic control mechanisms of structural–functional relationships that pervade the kingdom of biology.

MODEL

EIS modeling

The biological features of a tissue can be quantified by an equivalent electrical circuit analysis [9, 13]. The impedance spectra of tissues can be viewed as an equivalent circuit with two distributed circuit elements (DCE1 and DCE2) in series with a resistor R (Figure 1) [9]. Below, we describe this double-DCE model, which is mostly derived from Repo et al. [9]. Both DCE-elements consist of a parallel arrangement of a resistor R1 and R2 and a constant phase element \(Z_{\text{CPE1}}\) and \(Z_{\text{CPE2}}\), respectively. According to Macdonald [13], the impedance of the constant phase elements (CPEs) is expressed as:

\[
Z_{\text{CPE1}} = \frac{1}{(\phi \omega C_1)^{\psi_1}}, \quad \text{and} \quad Z_{\text{CPE2}} = \frac{1}{(\phi \omega C_2)^{\psi_2}},
\]

where \(\omega = 2\pi f\) is the angular velocity (\(f\) is the frequency), \(\phi\) is the imaginary unit, \(C_1\) and \(C_2\) are the specific capacitance for DCE1 and DCE2, respectively, and \(\psi_1\) and \(\psi_2\) are two distribution coefficients of the relaxation times for the two DCEs.

The impedance of the DCEs can be derived as:

\[
Z_{\text{DCE1}} = \frac{R_1}{1 + R_1 (\phi \omega C_1)^{\psi_1}} = \frac{R_1}{1 + (\phi \omega C_1 R_1^{1/\psi_1})^{\psi_1}},
\]

\[
Z_{\text{DCE2}} = \frac{R_2}{1 + R_2 (\phi \omega C_2)^{\psi_2}} = \frac{R_2}{1 + (\phi \omega C_2 R_2^{1/\psi_2})^{\psi_2}},
\]

where \(R_1\) and \(R_2\) are two resistances in the double-DEC model, by using the relationships:

\[
\frac{1}{Z_{\text{DCE1}}} = \frac{1}{R_1} + \frac{1}{Z_{\text{CPE1}}} \quad \text{and} \quad \frac{1}{Z_{\text{DCE2}}} = \frac{1}{R_2} + \frac{1}{Z_{\text{CPE2}}},
\]
Letting \( \tau_1 = C_1 R_1^{1/\Psi_1} \) and \( \tau_2 = C_2 R_2^{1/\Psi_2} \), which are two relaxation times of the DCEs, we have:

\[
Z_{DCE_i} = \frac{R_1}{1 + (\phi_0 \tau_1)\Psi_1} \quad \text{and} \quad Z_{DCE_i} = \frac{R_2}{1 + (\phi_0 \tau_2)\Psi_2}.
\]

Thus, the total complex impedance of the double-DCE is expressed as:

\[
Z = R + Z_{DCE_i} + Z_{DCE_2} = R + \frac{R_1}{1 + (\phi_0 \tau_1)\Psi_1} + \frac{R_2}{1 + (\phi_0 \tau_2)\Psi_2}, \tag{1}
\]

where the real part is

\[
R + \frac{R_1(1 + (\tau_1 \Psi_1)\cos(\Psi_1 \pi/2))}{1 + 2(\tau_1 \Psi_1)\cos(\Psi_1 \pi/2) + (\tau_1 \Psi_1)^2} + \frac{R_2(1 + (\tau_2 \Psi_2)\cos(\Psi_2 \pi/2))}{1 + 2(\tau_2 \Psi_2)\cos(\Psi_2 \pi/2) + (\tau_2 \Psi_2)^2}.
\]

and the imaginary part is

\[
-\frac{R_1(\tau_1 \Psi_1)\sin(\Psi_1 \pi/2)}{1 + 2(\tau_1 \Psi_1)\cos(\Psi_1 \pi/2) + (\tau_1 \Psi_1)^2} - \frac{R_2(\tau_2 \Psi_2)\sin(\Psi_2 \pi/2)}{1 + 2(\tau_2 \Psi_2)\cos(\Psi_2 \pi/2) + (\tau_2 \Psi_2)^2}.
\]

In sum, the double-DCE model contains three resistances \((R, R_1 \text{ and } R_2)\), two relaxation times \((\tau_1 \text{ and } \tau_2)\) and two distribution coefficients \((\Psi_1 \text{ and } \Psi_2)\) of the relaxation times. Mathematical interpretations of all these parameters are diagrammed in Figure 1.

**Regression model**

Consider a backcross population of \( n \) progeny in which a panel of markers are genotyped and EIS values are measured at \( T \) frequencies. Assume that a set of QTLs (forming a total of \( J \) genotypes) affect EIS curves in this population. We denote the total complex impedance of progeny \( i \) at frequency \( t \), composed of imaginary (I) and real (R) part, as \( y_i(t) = (y_{i1}(t), y_{i2}(t)) \). Structural mapping integrates the double-DCE (1) into a statistical model for QTL mapping, expressed as:

\[
y_i(t) = \sum_{j=1}^{J} z_j \left[ R_j + \frac{R_{1j}}{1 + (\phi_0 \tau_{1j})\Psi_{1j}} + \frac{R_{2j}}{1 + (\phi_0 \tau_{2j})\Psi_{2j}} \right] + e_i(t), \tag{2}
\]

where \( y_i(t) \) is the complex impedance value of progeny \( i \) at frequency \( t \); \( z_j = 1 \) if progeny \( i \) has the \( j \)th QTL genotype and \( z_j = 0 \) otherwise; \((R_j, R_{1j}, R_{2j}, \tau_{1j}, \tau_{2j}, \Psi_{1j}, \Psi_{2j})\) are a set of EIS parameters for QTL genotype \( j \); and \( e(t) \) is a residual error distributed as a complex normal distribution with mean 0 and variance \( \sigma^2 \). The complex vector \( y_i(t) \) can be divided into two \( T \)-dimensional subvectors, one for the real part and the other for the imaginary part. For simplicity, we assume that the two parts are not correlated, but each part follows a multivariate normal distribution. Therefore, the phenotype vector \( y_i(t) \) has the structure of covariance matrix, expressed as:

\[
\Sigma = \begin{bmatrix}
\sigma_1^2 & 0 \\
0 & \sigma_2^2
\end{bmatrix}
\]

\[
\Sigma_1 = \begin{bmatrix}
1 & \rho_1 & \cdots & \rho_1^{T-1} \\
\rho_1 & 1 & \cdots & \rho_1^{T-2} \\
& & \ddots & \vdots \\
\rho_1^{T-1} & & \ddots & 1 \\
\rho_2 & \cdots & \cdots & 1 \\
& & & & \\
\rho_2^{T-1} & \cdots & \cdots & 1
\end{bmatrix}
\]

and

\[
\Sigma_2 = \begin{bmatrix}
1 & \cdots & \cdots & \cdots \\
\rho_2 & \cdots & \cdots & \cdots \\
& & \ddots & \vdots \\
\rho_2^{T-1} & & \ddots & 1
\end{bmatrix}
\]

where \( \rho_1 \) and \( \rho_2 \) \((-1 < \rho_1, \rho_2 < 1)\) are the proportion parameters with which the correlations decay with frequency lag.

**Mixture model**

QTL genotypes are unknown, but can be inferred from marker genotypes \((M)\) and phenotypes \((y)\). To do that, structural mapping is founded on a mixture model, expressed as:

\[
L(\Omega|y, M) = \prod_{i=1}^{n} \left[ p_{1i} f_1(y_i, \mu_i) + \cdots + p_{Ji} f_J(y_i, \mu_J) \right]
\]

where \( \Omega \) contains the unknown parameters being estimated for the impedance values of different QTL genotypes, QTL positions, and residual
(co)variances; \( p_{ji} \) is the conditional probability of QTL genotype \( j \), conditional on the genotypes of the two flanking markers of progeny \( i \) [18]; and \( f_i(y_i, \mu_i) \) is a probability density function of multivariate complex normal distribution expressed as:

\[
f_i(y_i, \mu_i) = (2\pi)^{-1} \left( \sigma_1^2 \sigma_2^2 \right)^{-\frac{1}{2}} |\Sigma_1|^{-\frac{1}{2}} |\Sigma_2|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2\sigma_1^2} R(y_i - \mu_i)^\top \Sigma_1^{-1} R(y_i - \mu_i) \right. \\
\left. -\frac{1}{2\sigma_2^2} I(y_i - \mu_i)^\top \Sigma_2^{-1} I(y_i - \mu_i) \right] \tag{6}
\]

where \( \mu_i \) is the mean vector of \( y_i \) if it has the \( j \)th QTL genotype; and \( R(\cdot) \) and \( I(\cdot) \) stand for the real and imaginary parts of a complex vector, respectively.

**Parameter estimation**

In QTL mapping, we usually estimate QTL-effect parameters by assuming the QTL at a particular position in the marker interval. This procedure is repeated throughout the entire linkage group, allowing the likelihood profile to be drawn. The peak of the profile over the map position is the estimated
location of the putative QTL. This approach is used for structural mapping by fixing both the QTL position and residual correlations, $\rho_1$ and $\rho_2$ (since the correlations have a defined interval $[-1, 1]$).

We then implement the EM algorithm to estimate the other parameters related to the genotype-specific impedance values and the variances. Denote $\theta_j = (R_j, R_{1j}, R_{2j}, \tau_{1j}, \tau_{2j}, \psi_{1j}, \psi_{2j})'$ for QTL genotype $j$ and use the first-order Taylor expansion to approximate $\mu_j$ by:

$$\mu_j(\theta_j + \Delta \theta_j) \approx \mu_j(\theta_j) + \left(\frac{\partial \mu_j}{\partial \theta_j}\right)' \Delta \theta_j.$$  

By differentiating the log likelihood with respect to the unknown parameters and setting the derivatives to equal 0, we obtain:

$$\Delta \theta_j = \left\{ \sum_{i=1}^n P_{ji} \left[ \frac{1}{\sigma_1^2} \left( \frac{\partial R(\mu_i)}{\partial \theta_j} \right)' \Sigma_1^{-1} \left( \frac{\partial R(\mu_i)}{\partial \theta_j} \right) \right] \right.$$  

$$+ \frac{1}{\sigma_2^2} \left( \frac{\partial I(\mu_i)}{\partial \theta_j} \right)' \Sigma_2^{-1} \left( \frac{\partial I(\mu_i)}{\partial \theta_j} \right) \right\}^{-1} \times$$  

$$\sum_{i=1}^n P_{ji} \left[ \frac{1}{\sigma_1^2} \left( \frac{\partial R(\mu_i)}{\partial \theta_j} \right)' \Sigma_1^{-1} R(y_i - \mu_i) \right] +$$  

$$\frac{1}{\sigma_2^2} \left( \frac{\partial I(\mu_i)}{\partial \theta_j} \right)' \Sigma_2^{-1} I(y_i - \mu_i) \right\} + \Delta \theta_j^j.$$  

$$\Delta \theta_j = \left\{ \sum_{i=1}^n P_{ji} \left[ \frac{1}{\sigma_1^2} \left( \frac{\partial R(\mu_i)}{\partial \theta_j} \right)' \Sigma_1^{-1} R(y_i - \mu_i) \right] \right.$$  

$$+ \frac{1}{\sigma_2^2} \left( \frac{\partial I(\mu_i)}{\partial \theta_j} \right)' \Sigma_2^{-1} I(y_i - \mu_i) \right\}^{-1} \times$$  

$$\sum_{i=1}^n P_{ji} \left[ \frac{1}{\sigma_1^2} \left( \frac{\partial R(\mu_i)}{\partial \theta_j} \right)' \Sigma_1^{-1} R(y_i - \mu_i) \right] +$$  

$$\frac{1}{\sigma_2^2} \left( \frac{\partial I(\mu_i)}{\partial \theta_j} \right)' \Sigma_2^{-1} I(y_i - \mu_i) \right\} + \Delta \theta_j^j.$$  

where

$$P_{ji} = \frac{p_{ji} \gamma_j(y_i, \mu_i)}{p_{ji} \gamma_j(y_i, \mu_i) + \cdots + p_{ji} \gamma_j(y_i, \mu_j)}, \ j = 1, \cdots, J.$$  

Therefore, the maximum likelihood estimates (MLEs) of $\theta_j$ can be obtained by an iterative procedure; i.e. for the $t$th iteration, we have

$$\theta_j^{t+1} = \theta_j^t + \Delta \theta_j^t.$$  

The procedure for solving equivalent circuit EIS parameters are described as follow. Let

$$C_{1jt} = \frac{1}{1 + 2(\tau_{1j} \omega_0)^{\psi_0} \cos(\psi_{1j}/2) + (\tau_{1j} \omega_0)^{2\psi_0}}$$  

and

$$C_{2jt} = \frac{1}{1 + 2(\tau_{2j} \omega_0)^{\psi_0} \cos(\psi_{2j}/2) + (\tau_{2j} \omega_0)^{2\psi_0}}.$$  

We then have the $t$th element of the real and imaginary part of $\mu_j$ expressed as:

$$R(\mu_j) = R_1 + C_{1jt} R_{1j} \left[ 1 + (\tau_{1j} \omega_0)^{\psi_0} \cos(\psi_{1j}/2) \right] + C_{2jt} R_{2j} \left[ 1 + (\tau_{2j} \omega_0)^{\psi_0} \cos(\psi_{2j}/2) \right]$$  

and

$$I(\mu_j) = -C_{1jt} R_{1j} (\tau_{1j} \omega_0)^{\psi_0} \sin(\psi_{1j}/2) - C_{2jt} R_{2j} (\tau_{2j} \omega_0)^{\psi_0} \sin(\psi_{2j}/2).$$  

The derivatives of the $t$th element of $R(\mu_j)$ and $I(\mu_j)$ with respect to parameters $R_n$, $R_{1n}$, $R_{2n}$, $\tau_{1n}$, $\tau_{2n}$, $\psi_{1n}$, $\psi_{2n}$, and $\psi_{3n}$ can be obtained, which are shown in Box 1. By solving the equations, we obtain the estimates of the EIS parameters for different QTL genotypes.

Similarly, we can get the MLEs of $\sigma_1^2$ and $\sigma_2^2$, given the other parameters, as:

$$\sigma_1^2 = \frac{1}{T} \sum_{i=1}^n \left[ P_{ji} R(y_i - \mu_j) \Sigma_1^{-1} R(y_i - \mu_j) \right]$$  

and

$$\sigma_2^2 = \frac{1}{T} \sum_{i=1}^n \left[ P_{ji} I(y_i - \mu_j) \Sigma_2^{-1} I(y_i - \mu_j) \right].$$  

In the E-step, a posterior probability of progeny $i$ that has a QTL genotype $j$ is calculated as Equation (9). In the M-step, the parameters are estimated by Equations (9) and (10). This procedure is repeated until the iterative value of each parameter converges. The SEs of the MLEs can be estimated by using the inverse of the Fisher information matrix.

**Hypothesis tests**

Structural mapping is powerful for many biologically meaningful hypothesis tests. The first hypothesis is about the existence of any QTL affecting the impedance at a specific position on genome, which is formulated as:

$$\begin{cases} H_0: \theta_j = 0 \\ H_1: \text{at least one of the equalities above does not hold} \end{cases}$$  

where the $H_0$ corresponds to the reduced model, in which the data are fitted by assuming that no QTLs exist, and the $H_1$ corresponds to the full model (2).
Box 1 The derivatives.

The derivatives of the th element of R(μ) and l(μ) with respective to parameters R , R , τ , ψ , R , τ and ψ are obtained as:

\[
\frac{\partial R(\mu)}{\partial R} = 1,
\]

\[
\frac{\partial R(\mu)}{\partial R} = C_{ij} \left[ 1 + (\tau_{ij} \omega) \psi \cos \left( \frac{\nu_{ij} \pi}{2} \right) \right],
\]

\[
\frac{\partial R(\mu)}{\partial \tau} = -C_{ij} \frac{R_j \psi (\tau_{ij} \omega \psi)_{j-1}}{\left[ 2(\tau_{ij} \omega) \psi \ln(\tau_{ij} \omega) + \right.}
\]

\[
\left. \left( 1 + (\tau_{ij} \omega)^2 \psi \right) \cos \left( \frac{\nu_{ij} \pi}{2} \right) \ln(\tau_{ij} \omega) - \frac{\pi}{2} \left( 1 - (\tau_{ij} \omega)^2 \psi \right) \sin \left( \frac{\nu_{ij} \pi}{2} \right) \right],
\]

\[
\frac{\partial R(\mu)}{\partial \psi} = C_{ij} \left[ 1 + (\tau_{ij} \omega) \psi \cos \left( \frac{\nu_{ij} \pi}{2} \right) \right],
\]

\[
\frac{\partial R(\mu)}{\partial \tau} = -C_{ij} \frac{R_j \psi (\tau_{ij} \omega \psi)_{j-1}}{\left[ 2(\tau_{ij} \omega) \psi \ln(\tau_{ij} \omega) + \right.}
\]

\[
\left. \left( 1 + (\tau_{ij} \omega)^2 \psi \right) \cos \left( \frac{\nu_{ij} \pi}{2} \right) \ln(\tau_{ij} \omega) - \frac{\pi}{2} \left( 1 - (\tau_{ij} \omega)^2 \psi \right) \sin \left( \frac{\nu_{ij} \pi}{2} \right) \right],
\]

\[
\frac{\partial l(\mu)}{\partial R} = 0,
\]

\[
\frac{\partial l(\mu)}{\partial R} = -C_{ij} (\tau_{ij} \omega) \psi \sin \left( \frac{\nu_{ij} \pi}{2} \right),
\]

\[
\frac{\partial l(\mu)}{\partial \tau} = -C_{ij} \frac{R_j \psi (\tau_{ij} \omega \psi)_{j-1}}{\left[ 2(\tau_{ij} \omega) \psi \ln(\tau_{ij} \omega) + \right.}
\]

\[
\left. \left( 1 - (\tau_{ij} \omega)^2 \psi \right) \sin \left( \frac{\nu_{ij} \pi}{2} \right) \right],
\]

\[
\frac{\partial l(\mu)}{\partial \psi} = -C_{ij} (\tau_{ij} \omega) \psi \sin \left( \frac{\nu_{ij} \pi}{2} \right),
\]

\[
\frac{\partial l(\mu)}{\partial \tau} = -C_{ij} \frac{R_j \psi (\tau_{ij} \omega \psi)_{j-1}}{\left[ 2(\tau_{ij} \omega) \psi \ln(\tau_{ij} \omega) + \right.}
\]

\[
\left. \left( 1 - (\tau_{ij} \omega)^2 \psi \right) \sin \left( \frac{\nu_{ij} \pi}{2} \right) \right],
\]

\[
\frac{\partial l(\mu)}{\partial \psi} = -C_{ij} \frac{R_j \psi (\tau_{ij} \omega \psi)_{j-1}}{\left[ 2(\tau_{ij} \omega) \psi \ln(\tau_{ij} \omega) + \right.}
\]

\[
\left. \left( 1 - (\tau_{ij} \omega)^2 \psi \right) \sin \left( \frac{\nu_{ij} \pi}{2} \right) \right].
\]
The log-likelihood ratio (LR) of the full model over the reduced model is applied to test the above hypotheses:

\[
LR = -2 \log \frac{L_0(\hat{\Omega})}{L_1(\hat{\Omega})}
\]

where \( \hat{\Omega} \) and \( \hat{\Omega} \) denote the MLEs of the unknown parameters under the \( H_0 \) and \( H_1 \), respectively. If a high peak of LR profiles exceeds a critical threshold, then a QTL that controls the curve of the electrical impedance on the complex plane is asserted to exist in a marker interval. The genome-wide critical threshold is determined by performing permutation tests [19].

The merit of the model also lies in the test of a number of physiologically meaningful properties of tissues. Each of the equivalent circuit EIS parameters contained within a distributed model of the tissues [1] has a particular biological meaning. For example, the relaxation times, \( \tau_1 \) and \( \tau_2 \), describe the peak of the high-frequency arc of the impedance spectrum, which are highly correlated with frost hardiness, and \( \psi_1 \) and \( \psi_2 \) are the distribution coefficients of the relaxation times.

At low frequencies, the current may not pass through the cell membranes but flows in the apoplastic space of tissues (for plants). Thus, the extracellular resistance \( (R_E) \) is calculated as:

\[
R_E = R + R_1 + R_2.
\]

But the high-frequency current may pass the cell membranes and, thus, flows in both the apoplastic and symplastic space, in which case the intracellular resistance \( (R_i) \) is calculated as:

\[
R_i = R \left(1 + \frac{R}{R_1 + R_2}\right).
\]

It is interesting to test how the QTLs detected control each of these parameters by formulating the null hypothesis, \( H_0 : (\tau_1, \tau_2) = (\bar{\tau}_1, \bar{\tau}_2) \), \( H_0 : (\psi_1, \psi_2) = (\bar{\psi}_1, \bar{\psi}_2) \), \( H_0 : R_E = R_E \) and \( H_0 : R_i = R_i \), respectively, for \( j = 1, \ldots, J \). The LR test statistics are calculated to compare against the critical values for the significance determination.

**RESULTS**

**Worked example**

We described a statistical model for structural mapping that capitalizes on the capacity of the EIS device to quantify the biological and physiochemical features of a tissue by an equivalent electrical circuit analysis [9]. To validate the ability of structural mapping to map a phenotypic trait from its structural underpinnings, we implemented the algorithm to analyze a mapping data collected from a controlled cross between two heterozygous parents in Chinese jujube (Zizyphus jujuba Mill.) [20]. A male tree Dongzao (Z. jujuba Mill, cv: Dongzao) was pollinated to a female tree Linyilizao (Z. jujuba Mill, cv: Linyilizao) to generate a full-sib family. A mapping population of 150 individuals from this family was established by genotyping 423 molecular markers to construct two parent-specific linkage maps. Thus, a pseudo-test backcross design is used for QTL mapping [21]. Part of the mapping population (86 hybrids) were measured for EIS at 31 different frequencies (300 Hz to 4 MHz) with a dual trace oscilloscope using shoots under a range of temperatures (–4 to 20°C). Electrical impedance was displayed as real (resistive component) and imaginative (capacitive reactance component) parts.

Figure 2 illustrates Cole–Cole impedance plots that describe the imaginary versus real complex plane for these hybrids. Here, the imaginary part of the complex impedance of the stems under study is plotted against the real part, each curve being a function of the frequency characterized by one hybrid. The imaginary part changes with the part par in a hyperbolic manner. The impedance spectra of each hybrid can be viewed as an equivalent circuit with two DCEs. The variation of Cole–Cole impedance plots among different hybrids increases dramatically with increasing frequency, implying a possible involvement of QTLs for electrical impedance spectra.

To ensure its homoscedasticity over frequency, both the real and imaginary data are log-transformed. By scanning for the existence of QTLs in the genome, we calculated LRs at every two 2 cM through the linkage map, obtaining an LR profile against the genome position. To simplify calculations, we used a single-DCE model which takes the first two terms of model (1). Several significant QTLs were detected to affect electrical impedance spectra at the chromosome- and genome-wide level, but only one at the genome-wide level detected on the female-based linkage map is reported here for the demonstration of the new model (Figure 3). At this QTL, two different genotypes each are described by a group of EIS parameters \( (R, R_i, \tau_1, \psi_1) \). As shown in Figure 4, the two QTL genotypes display different
patterns of EIS curves, suggesting the impact of this QTL on structural components of shoots.

The EIS parameters through recalculation can be used to describe extracellular and intracellular resistances of a tissue which are related to its physiological properties such as frost hardness. We test whether the QTL detected using our model affects extracellular and intracellular resistances of the stems of Chinese jujube. This QTL is found to display highly significant effects on both types of cellular resistance.
According to previous studies in tree physiology [8, 9], these two types of resistance are highly associated with frost hardness. Our mapping population allows the distinction of two QTL-genotype groups of Chinese jujube hybrids based on the posterior probabilities of QTL genotypes given marker genotypes [18]. It is found that these two genotypes differ highly significantly in frost hardness of jujube hybrids ($P < 0.001$), suggesting that results of QTL mapping with EIS can be used to diagnose the performance of trees to tolerate and resist to coldness.

**Simulation**

We performed a simulation study to evaluate the precision of parameter estimates, the power of structural mapping and its false positive rates using EIS data. The simulation mimics the real example analyzed above by assuming a 100-cM long linkage group with 11 equidistant markers in a backcross population. A putative QTL is assumed to be located at 46 cM from the first marker of the group. Different heritabilities ($H^2 = 0.1, 0.4$) and various sample sizes ($n = 100, 200$) are considered. Estimated values of the parameters from the new model are tabulated in Table 1, in a comparison with their true values used to simulate the data. It can be seen that the model provides reasonably good estimates of the model parameters, even with a modest heritability (0.1) and sample size (100). When the sample size increases to 200, the model can surely well estimate all the parameters. At a small heritability and sample size, the model displays reasonably high power for QTL detection (0.80) and low FRP (0.06). These features can improve further when a heritability and/or sample size increases.

**DISCUSSION**

The electrical properties of tissues in a range of frequencies are associated with the cellular components and the dimensions, internal structure and arrangements of the constituent cells [3]. Therefore, tissues with different cellular structures will produce impedance spectra characteristic of these issues. By analyzing electrical properties of biological tissues, we can better understand their structure and function. More recently, EIS has been increasingly used to study the electrical impedance of various tissues over a frequency range and determine their frequency-dependent electric and dielectric behavior [3, 6–11], providing scientific guidance for prognosis, diagnosis and prevention. The use of impedance spectroscopy electrical characterization is a novel approach to comprehend the genetic architecture of phenotypic traits underlying pathology and complex diseases.

We described a computational model for structural mapping by taking into account the structural components of phenotypic traits using EIS.
information. Structural mapping incorporates the electrical properties of phenotypic formation into a genetic mapping framework by which the genetic control of complex traits can be studied from a mechanistic perspective. Structural mapping was used to analyze a mapping data collected for an experimental cross of outcrossing tree species, Chinese jujube. Significant QTLs were detected to affect the complex impedance of shoots, some of which were also detected by mapping frost hardiness [20]. Because there is a strong correlation between intracellular resistance and frost hardiness [8, 9], the detection of pleiotropic QTLs for these two traits can well confirm the validation of our model. The simulation study by mimicking the real example provides a justification of statistical properties of the model. The EIS-based structural mapping model can be expanded to study the comprehensive genetic architecture of complex traits using genome-wide association study. More specifically, structural mapping (2) should involve the impacts of genetic interactions [22], genetic imprinting [23], epigenetic marks [24] and copy number variation [25] on phenotypic variation. Furthermore, structural mapping can also be used to study phenotypic response to different environmental factors, such as temperature [8, 9] through studying the genetic architecture of phenotypic plasticity to a range of environments. By constructing a logistic sigmoid function, we can determine the biological properties of a tissue, such as frost hardiness, based on extracellular resistance over a series of temperatures. This model is expressed as:

\[ y = \frac{A}{1 + e^{B(C-x)}} + D \]

where \( y \) and \( x \) are the specific extracellular resistance and the exposure temperature, respectively, \( A \) and \( D \) define asymptotes of the function, and \( B \) is the slope at the inflection point \( C \). The frost hardness is estimated as the inflection point of the above logistic function [9]. Thus, by incorporating the above logistic function, structural mapping can directly detect specific QTLs associated with frost hardness through structural dissection.

**Table 1:** Means of the QTL position and MLEs of the parameters for two QTL genotypes in a pseudo-test backcross design of different sample sizes (\( n \)) under different heritabilities (\( H^2 \)), calculated from 200 simulation replicates. The root of mean squares errors of the estimate of each parameter is in parentheses.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>( n = 100 )</th>
<th>( n = 200 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H^2 = 0.1 )</td>
<td>( H^2 = 0.4 )</td>
<td>( H^2 = 0.1 )</td>
<td>( H^2 = 0.4 )</td>
</tr>
<tr>
<td>QTL position</td>
<td>46</td>
<td>45.95 (1.67)</td>
<td>45.89 (1.64)</td>
</tr>
<tr>
<td>( R_1 )</td>
<td>14 288.2</td>
<td>14 371 (1468)</td>
<td>14 150 (598)</td>
</tr>
<tr>
<td>( R_{11} )</td>
<td>132 114</td>
<td>131 453 (9543)</td>
<td>131 654 (3775)</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>3.137</td>
<td>3.185 (0.135)</td>
<td>3.148 (0.063)</td>
</tr>
<tr>
<td>( \psi_1 )</td>
<td>0.5256</td>
<td>0.526 (0.008)</td>
<td>0.525 (0.004)</td>
</tr>
<tr>
<td>( R_2 )</td>
<td>14 109</td>
<td>13 089 (1334)</td>
<td>13 963 (557)</td>
</tr>
<tr>
<td>( R_{12} )</td>
<td>181 502</td>
<td>182 249 (13212)</td>
<td>182 735 (5410)</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>6.862</td>
<td>7.225 (0.287)</td>
<td>6.946 (0.117)</td>
</tr>
<tr>
<td>( \psi_2 )</td>
<td>0.5489</td>
<td>0.539 (0.007)</td>
<td>0.547 (0.003)</td>
</tr>
<tr>
<td>( r_1 )</td>
<td>0.9994</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>( r_2 )</td>
<td>0.9930</td>
<td>0.993</td>
<td>0.993</td>
</tr>
<tr>
<td>( \sigma_1^2 )</td>
<td>0.229</td>
<td>0.223 (0.033)</td>
<td>0.224 (0.025)</td>
</tr>
<tr>
<td>( \sigma_2^2 )</td>
<td>0.263</td>
<td>0.279 (0.046)</td>
<td>0.279 (0.038)</td>
</tr>
<tr>
<td>( s_1^2 )</td>
<td>0.046</td>
<td>0.045 (0.006)</td>
<td>0.045 (0.006)</td>
</tr>
<tr>
<td>( s_2^2 )</td>
<td>0.053</td>
<td>0.053 (0.007)</td>
<td>0.053 (0.007)</td>
</tr>
</tbody>
</table>

The parameters include three types: (i) the position of QTL, (ii) QTL effects specified by genotype-specific DCE parameters and (iii) AR(1) parameters.
FUNDING
This work is partially supported by Special Fund for Forestry Scientific Research in the Public Interest (No. 201004017); NSF/IOS-0923975; Changjiang Scholars Award and ‘Thousand-Person Plan’ Award.

References