

# Statistical Dissection of Cyto-Nuclear Epistasis Subject to Genomic Imprinting in Line Crosses

Tao He<sup>1</sup>, Jian Sa<sup>1,2</sup>, Ping-Shou Zhong<sup>1</sup>, Yuehua Cui<sup>1,2\*</sup>

**1** Department of Statistics and Probability, Michigan State University, East Lansing, Michigan, United States of America, **2** Division of Medical Statistics, School of Public Health, Shanxi Medical University, Taiyuan, Shanxi, China

## Abstract

Cytoplasm contains important metabolism reaction organelles such as mitochondria and chloroplast (in plant). In particular, mitochondria contains special DNA information which can be passed to offsprings through maternal gametes, and has been confirmed to play a pivotal role in nuclear activities. Experimental evidences have documented the importance of cyto-nuclear interactions in affecting important biological traits. While studies have also pointed out the role of interaction between imprinting nuclear DNA and cytoplasm, no statistical method has been developed to efficiently model such effect and further quantify its effect size. In this work, we developed an efficient statistical model for genome-wide estimating and testing the cytoplasmic effect, nuclear DNA imprinting effect as well as the interaction between them under reciprocal backcross and  $F_2$  designs derived from inbred lines. Parameters are estimated under maximum likelihood framework implemented with the EM algorithm. Extensive simulations show good performance in a variety of scenarios. The utility of the method is demonstrated by analyzing a published data set in an  $F_2$  family derived from C3H/HeJBir and C57BL/6 J mouse strains. Important cyto-nuclear interactions were identified. Our approach provides a quantitative framework for identifying and estimating cyto-nuclear interactions subject to genomic imprinting involved in the genetic control of complex traits.

**Citation:** He T, Sa J, Zhong P-S, Cui Y (2014) Statistical Dissection of Cyto-Nuclear Epistasis Subject to Genomic Imprinting in Line Crosses. PLoS ONE 9(3): e91702. doi:10.1371/journal.pone.0091702

**Editor:** Rongling Wu, Pennsylvania State University, United States of America

**Received:** November 26, 2013; **Accepted:** February 14, 2014; **Published:** March 18, 2014

**Copyright:** © 2014 He et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was partially supported by NSF grants DMS-1209112, IOS-1237969, and MCB-1121650, and by National Natural Science Foundation of China grant 31371336. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: cui@stt.msu.edu

## Introduction

One of the central foci in biological study is to unravel the genetic secrets of complex traits of agricultural, evolutionary and biomedical importance. Quantitative trait locus (QTL) mapping has been the major tool for this purpose over decades [1–3]. In QTL mapping, the identified QTLs are chromosome segments harboring potential genetic variants that could give rise to phenotypical manifestation. Large successes have been witnessed in the past [4]. However, there are still many phenomena that could not be explained by Mendelian genetics, leading to the new exploration of research focus on epigenetics [5].

Genomic imprinting, one of the major epigenetic phenomena, plays key roles in controlling embryonic growth and development [6,7]. Let subscript letter  $M$  and  $F$  denote the parental origin of two alleles in a diploid organism, then a locus with two alleles  $A$  and  $a$  is thought to be imprinted if two heterozygotes  $A_M a_F$  and  $a_M A_F$  have different expressions [8]. The malfunction of imprinted genes could potentially lead to abnormal characters such as cancers or other genetic disorders [9].

Genomic imprinting effect is considered as one type of parent-of-origin effect due to allelic effect with specific parental origin. In contrast to this, maternal effect or cytoplasmic effect is also considered as one type of parent-of-origin effect in which the offspring's expression is influenced by maternal parent. For example, a mother's genotype, even if not transmitted to her offspring, may influence in utero conditions and increase risk and/

or interact with genetic predisposition for particular diseases among those offspring [10]. For cytoplasm, it contains a wide variety of organelles such as mitochondria and chloroplast (in plant). Almost all the reactions of cellular metabolism take place in such an environment. It has been demonstrated that cytoplasm plays a central role in coordinating the activities of nuclear genetic materials [11–15]. Thus, the identification of cyto-nuclear interaction could shed novel insights into the genetic and epigenetic control of phenotypic variation. A number of empirical studies have documented the significant contribution of cyto-nuclear interaction to phenotypic variation in organisms such as wheat, rice, mice, yeast and *Drosophila* [16–19].

On the other hand, the existence of such parent-of-origin effects may lead to incorrect interpretation of the (marginal) effects of particular genes when performing genetic mapping studies, unless such effects are appropriately accounted for in the analysis [20]. Statistical methods for dissecting genomic imprinting effect has been extensively studied in literature (e.g., [21–26]). Tang et al. [27] developed a model to evaluate cyto-nuclear interaction effect based on experimental crosses. However, how the two types of parent-of-origin effects, one in nuclear level and one in cytoplasmic level, interact to influence phenotypic variation is largely unknown due to the lack of proper statistical models.

In this work, we discuss potential scenarios of parent-of-origin effects, and present a statistical method to dissect the cyto-nuclear interaction effects subject to genomic imprinting. The developed

framework is based on experimental crosses which can be realized through two different designs, reciprocal backcross and F<sub>2</sub> design. When an F<sub>2</sub> design is considered, sex-specific difference in recombination fractions, which is initially discovered in [28,29] and later observed in many species, is incorporated into our model to distinguish the genetic differences between two reciprocal heterozygote. Such information can be found in literature, such as the female-to-male recombination ratio of 1.6:1 for human [30], 1.4:1 for pig [31], 1.4:1 for dog [32], 1.25:1 for mouse [33] on average across the whole genome. A genome-wide scan for the identification of iQTL mapping cyto-nuclear epistasis is performed based on the interval mapping theory, and parameters are estimated based on the framework of maximum likelihood method implemented with the EM algorithm [34]. Extensive simulations are conducted to evaluate the performance of our model under different scenarios, such as different sample sizes, different heritability levels, and different gene effects. The utility of the method is illustrated by applying it to a genome-wide scan of four traits in an F<sub>2</sub> family derived from two inbred mouse strains.

### Statistical Methods

#### Genetic designs

Consider a design initiated with two inbred lines with two segregating alleles *A* and *a*. Let subscript letter *M* and *F* represent the parental origin of offspring alleles inherited from the mother and father, respectively. A complete dissection of the cyto-nuclear interaction subject to imprinting needs experimental designs that can distinguish the quantitative variation between two heterozygotes *A<sub>M</sub>a<sub>F</sub>* and *a<sub>M</sub>A<sub>F</sub>*, and also against the cytoplasmic effect. For this purpose, a reciprocal backcross or F<sub>2</sub> design is proposed so that variations and interactions can be fully introduced.

Figure 1 illustrates the reciprocal backcross design. Let the maternal parent carrying genotype *AA* (denoted as P2) has a cytoplasmic effect in contrast to the maternal line carrying genotype *aa* (denoted as P1) as the reference line. In the diagram, individuals that carry the maternal effect coming from the maternal parent with genotype *AA* are denoted by gray squares. Four possible backcrosses can be initiated as illustrated in Fig. 1. As shown in the figure, any backcross offsprings coming from the middle two designs carry the cytoplasmic maternal effect derived from the *AA* genotype.

Figure 2 shows the F<sub>2</sub> design. Denote the left one as design *S*<sub>1</sub>, and the right one as design *S*<sub>2</sub>. The cross of two F<sub>1</sub>'s generates four possible allele-specific F<sub>2</sub> genotypes. Assuming there is a cytoplasmic effect, F<sub>2</sub> offsprings may show different phenotypes depending on the genotype of the maternal parental lines. For example, if maternal cytoplasmic effect exists, the offspring phenotypic value for *AA* genotype may be different depending on whether it comes from the *S*<sub>1</sub> or *S*<sub>2</sub> design. In the F<sub>2</sub> design, the two reciprocal heterozygotes *A<sub>M</sub>a<sub>F</sub>* and *a<sub>M</sub>A<sub>F</sub>* cannot be distinguished in

general. Sex-specific recombination difference in male and female needs to be considered in order to distinguish the two (Cui et al. 2006) [25].

#### Statistical parameterization

For a particular cross, let *y<sub>j</sub>* (*j* = 1, ..., *n*) denote the phenotypic value of interest. Following Tang et al. [27], the one-QTL genetic model can be expressed as,

$$y_j = \mu + cx_{1j} + ax_{2j} + dx_{3j} + ix_{4j} + i_{ca}x_{5j} + i_{cd}x_{6j} + i_{ci}x_{7j} + e_j \quad (1)$$

where  $\mu$  is the overall mean; *c* is the cytoplasmic effect; *a*, *d* and *i* are the additive, dominance and imprinting effects of a QTL, respectively; and *i<sub>ca</sub>*, *i<sub>cd</sub>* and *i<sub>ci</sub>* are the cytoplasm by additive, cytoplasm by dominance and cytoplasm by imprinting interactions, respectively; *x<sub>1j</sub>* is an indicator variable, denoting *x<sub>1j</sub>* = 1 for the *AA* maternal cytoplasm and *x<sub>1j</sub>* = -1 for the *aa* maternal cytoplasm; *x<sub>2j</sub>*, *x<sub>3j</sub>*, ..., *x<sub>7j</sub>* are other indicators of various effects describing the additive, dominance and the interaction effect between the cytoplasm and genetic variables.

For the *S*<sub>1</sub> design in the F<sub>2</sub> population initiated with cross *AA* × *aa*, the mean genotypic values of four possible genotypes formed by different allelic combination from the two F<sub>1</sub> parents can be expressed as,

$$\begin{cases} \mu_1 = \mu + c + a + i_{ca}, & \text{for } A_M A_F, \\ \mu_2 = \mu + c + d + i + i_{cd} + i_{ci}, & \text{for } A_M a_F, \\ \mu_3 = \mu + c + d - i + i_{cd} - i_{ci}, & \text{for } a_M A_F, \\ \mu_4 = \mu + c - a - i_{ca}, & \text{for } a_M a_F. \end{cases} \quad (2)$$

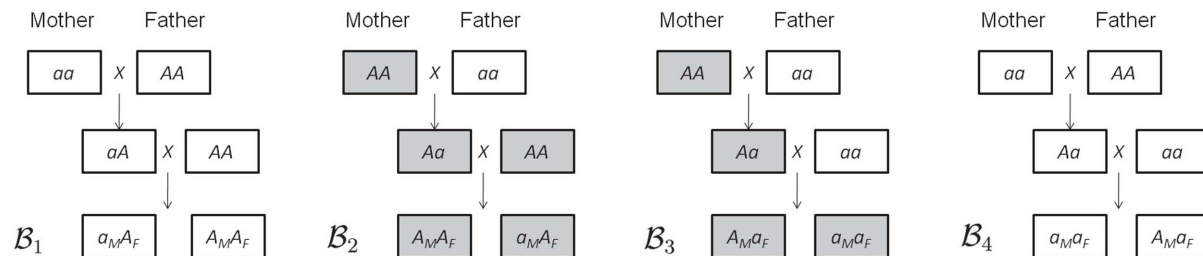
Similarly for the *S*<sub>2</sub> design initiated with cross *aa* × *AA*, the genetic model can be expressed as,

$$\begin{cases} \mu_5 = \mu - c + a - i_{ca}, & \text{for } A_M A_F, \\ \mu_6 = \mu - c + d + i - i_{cd} - i_{ci}, & \text{for } A_M a_F, \\ \mu_7 = \mu - c + d - i - i_{cd} + i_{ci}, & \text{for } a_M A_F, \\ \mu_8 = \mu - c - a + i_{ca}, & \text{for } a_M a_F. \end{cases} \quad (3)$$

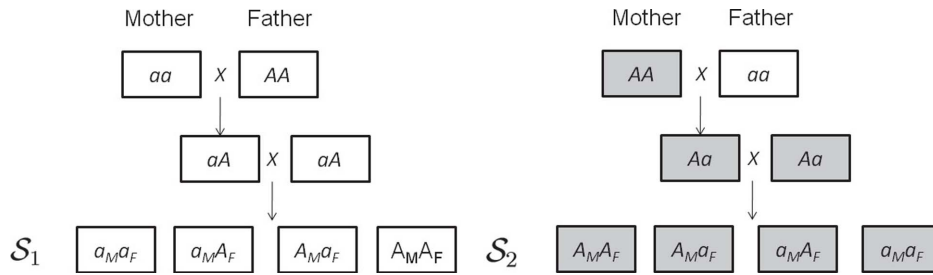
For the reciprocal backcross design which consists of *B*<sub>1</sub>, *B*<sub>2</sub>, *B*<sub>3</sub> and *B*<sub>4</sub>, the indicator variables in Eq. (1) describing different QTL genotypes are defined in Table 1.

For simplicity, we will use matrix form to rewrite the models. Let us denote

$$\beta = (\mu, c, a, d, i, i_{ca}, i_{cd}, i_{ci})^T, \mu = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8)^T.$$



**Figure 1. A reciprocal backcross design.**  
doi:10.1371/journal.pone.0091702.g001



**Figure 2. A reciprocal F<sub>2</sub> design.**  
doi:10.1371/journal.pone.0091702.g002

Then the relationship between the eight genetic means and the eight parameters can be written as

$$\mu = D\beta, \tag{4}$$

where

$$D = D^{BC} = \begin{pmatrix} 1 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 1 & -1 & 0 & -1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & -1 & 0 & 1 & -1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 1 & 0 & -1 & -1 \\ 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \tag{5}$$

for the reciprocal backcross design, and

$$D = D^{F_2} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & -1 & 0 & 1 & -1 \\ 1 & 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 1 & -1 & 0 & -1 & 1 \\ 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \tag{6}$$

for the F<sub>2</sub> design. For the purpose of illustration, the following

estimation and inference is demonstrated through the F<sub>2</sub> design. The same procedure applies to BC design too.

**The mixture model and the likelihood**

Statistical methods for QTL interval mapping based on a mixture model traced back to the work by Lander and Botstein [1]. In the mixture model, each observation *y* is modeled as a weighted mixture of *J* (known and finite) components, and each component, which corresponds to a certain genotype category depending on the underlying genetic design, follows a certain distribution *f<sub>j</sub>* with weight *π<sub>j</sub>*. Conditional on the marker genotype **M** and unknown parameters **φ** and **η**, the density of the observed *y* has the following expression

$$y \sim p(y|M, \phi, \eta) = \pi_1 f_1(y|M, \phi_1, \eta) + \dots + \pi_J f_J(y|M, \phi_J, \eta), \tag{7}$$

where **π** = (π<sub>1</sub>, ..., π<sub>J</sub>)<sup>T</sup> refers to the mixture proportions which are constrained nonnegative and  $\sum_{j=1}^J \pi_j = 1$ ; **φ** = (φ<sub>1</sub>, ..., φ<sub>J</sub>)<sup>T</sup> is a vector for the component-specific parameters, with φ<sub>*j*</sub> being specific to *j*th component; and **η** consists of parameters (i.e., residual variance) that are common to all components.

For the F<sub>2</sub> design we described above, there are four genotypes at each locus (A<sub>M</sub>A<sub>F</sub>, A<sub>M</sub>a<sub>F</sub>, a<sub>M</sub>A<sub>F</sub>, and a<sub>M</sub>a<sub>F</sub>). The genotype of the QTL is generally unobservable, but can be inferred by using the two flanking markers' information. Given the flanking marker genotypes of the *i*th individual, the conditional probabilities **π<sub>i</sub>** = (π<sub>A<sub>M</sub>A<sub>F</sub>|*i*</sub>, π<sub>A<sub>M</sub>a<sub>F</sub>|*i*</sub>, π<sub>a<sub>M</sub>A<sub>F</sub>|*i*</sub>, π<sub>a<sub>M</sub>a<sub>F</sub>|*i*</sub>) of the QTL genotype can be calculated. These conditional probabilities become the mixture proportions in the mixture model (7). Let us denote **π<sub>i</sub>** = (π<sub>1|*i*</sub>, π<sub>2|*i*</sub>, π<sub>3|*i*</sub>, π<sub>4|*i*</sub>) to simplify the notation. These conditional probabilities are expressed in terms of sex-specific recombination rates in *p* to distinguish the two reciprocal heterozygotes.

**Table 1. QTL genotypes and corresponding genetic components under different backcross designs.**

Backcross design	BC QTL genotype	<i>c</i>	<i>a</i>	<i>d</i>	<i>i</i>	<i>i<sub>ca</sub></i>	<i>i<sub>cd</sub></i>	<i>i<sub>ci</sub></i>
B <sub>1</sub>	A <sub>M</sub> A <sub>F</sub>	-1	1	0	0	-1	0	0
	a <sub>M</sub> A <sub>F</sub>	-1	0	1	-1	0	-1	1
B <sub>2</sub>	A <sub>M</sub> A <sub>F</sub>	1	1	0	0	1	0	0
	a <sub>M</sub> A <sub>F</sub>	1	0	1	-1	0	1	-1
B <sub>3</sub>	A <sub>M</sub> a <sub>F</sub>	1	0	1	1	0	1	1
	a <sub>M</sub> a <sub>F</sub>	1	-1	0	0	-1	0	0
B <sub>4</sub>	A <sub>M</sub> a <sub>F</sub>	-1	0	1	1	0	-1	-1
	a <sub>M</sub> a <sub>F</sub>	-1	-1	0	0	1	0	0

doi:10.1371/journal.pone.0091702.t001

Please refer to Cui et al. [25] for the conditional probabilities of QTL genotypes given marker genotypes in terms of sex-specific recombination fractions for an F<sub>2</sub> design.

Assume the total number of F<sub>2</sub> offspring for design S<sub>1</sub> and S<sub>2</sub> are n<sub>1</sub> and n<sub>2</sub> respectively, and let n = n<sub>1</sub> + n<sub>2</sub>. Phenotype data for a certain quantitative trait can be observed and recorded as a vector y = (y<sub>1</sub>, ..., y<sub>n</sub>), where y<sup>(1)</sup> = (y<sub>1</sub>, ..., y<sub>n<sub>1</sub></sub>) are from design S<sub>1</sub> and y<sup>(2)</sup> = (y<sub>n<sub>1</sub>+1</sub>, ..., y<sub>n</sub>) are from design S<sub>2</sub>. Marker information can be reorganized as matrix M = (M<sub>1</sub>; M<sub>2</sub>)<sup>T</sup>, where the j<sup>th</sup> (j = 1, ..., n<sub>1</sub>) row of M<sub>1</sub> (the j<sup>th</sup> row of matrix M) contains all the marker information of the j<sup>th</sup> individual under design S<sub>1</sub> and the k<sup>th</sup> (k = 1, ..., n<sub>2</sub>) row of M<sub>2</sub> (the (n<sub>1</sub> + k)<sup>th</sup> row of matrix M) contains all the marker information of the k<sup>th</sup> individual under design S<sub>2</sub>. Based on the mixture model (7), and with the independence assumption, the joint likelihood function for the F<sub>2</sub> family with total n individuals, constructed by combining design S<sub>1</sub> and S<sub>2</sub> together, can be formulated as,

$$L(\Theta|M, y) = \prod_{i=1}^{n_1} [\pi_{1|i}f_1(y_i) + \pi_{2|i}f_2(y_i) + \pi_{3|i}f_3(y_i) + \pi_{4|i}f_4(y_i)] \times \prod_{i=n_1+1}^n [\pi_{1|i}f_5(y_i) + \pi_{2|i}f_6(y_i) + \pi_{3|i}f_7(y_i) + \pi_{4|i}f_8(y_i)], \tag{8}$$

where the unknown vector Θ contains the QTL position, QTL effects and residual variance, and the density function f<sub>j</sub> (j = 1, ..., 8) is assumed to follow a normal distribution with mean μ<sub>j</sub> and common variance σ<sup>2</sup>. More specifically, parameter vector Θ can be divided into two subsets, Θ<sub>l</sub> and Θ<sub>g</sub>, where Θ<sub>l</sub> describes the location of QTL and Θ<sub>g</sub> contains all the genetic parameters, including QTL-effects vector β = (μ, c, a, d, i, i<sub>ca</sub>, i<sub>cd</sub>, i<sub>ci</sub>)<sup>T</sup> and residual variance σ<sup>2</sup> in our model.

**Parameter Estimation**

To estimate the unknown parameters Θ = (Θ<sub>l</sub>, Θ<sub>g</sub>), several algorithms could be implemented, such as Expectation-Maximization (EM), Newton Raphson and Fisher Scoring. Among all these methods, EM algorithm, which was initially developed by Dempster et al. [34], is most commonly used in QTL mapping study. In this paper, EM algorithm is applied to obtain the maximum likelihood estimates (MLEs). This procedure involves differentiating the log-likelihood function with respect to each unknown parameter, letting the derivatives equal to zero, and solving the log-likelihood equation for the corresponding parameter. Please read Appendix S1 for the detailed derivations of parameters estimation and the algorithm.

For the QTL position which is unknown in the model, we did not estimate parameters Θ<sub>l</sub> directly. As commonly treated in QTL mapping studies, we applied a grid search approach to estimate the putative QTL position via scanning the entire linkage genome by 1 or 2 cM increment flanked by two markers and did a hypothesis testing at each putative position. A likelihood ratio or LOD profile plot can be generated to graphically display the LR or LOD test statistic for a putative QTL at each testing position. The genomic position which corresponds to a peak in the profile plot is the MLE of the QTL location, given that the peak passes the genome-wide threshold determined by the permutation tests detailed below. Bootstrap methods can be applied to assess the confidence interval of the estimated position [35].

**Hypothesis Test**

Testing the overall QTL effect on the quantitative trait is the first step toward a complete dissection of different genetic

contributions to the trait. Once the MLEs of the parameters are obtained, the presence of QTL responsible for the variation of the quantitative phenotype can be tested by using the following hypotheses,

$$\begin{cases} H_0 : & a = d = i = i_{ca} = i_{cd} = i_{ci} = 0 \\ H_1 : & \text{not all equal zero} \end{cases}$$

The test statistic for testing the above hypotheses is calculated as the log-likelihood ratio test statistic (LR) of the full model (H<sub>1</sub>) over the reduced model (H<sub>0</sub>):

$$LR = -2 \log[L(\tilde{\Theta}) - L(\hat{\Theta})] \tag{9}$$

where  $\tilde{\Theta}$  and  $\hat{\Theta}$  denote the MLEs of the unknown parameters under H<sub>0</sub> and H<sub>1</sub>, respectively. Since a genome-wide scan involves multiple correlated tests, we use the permutation tests proposed by Churchill and Doerge [36] to find the threshold value.

If there is a QTL, a number of other hypothesis tests can also be performed to test the property of the detected QTL. To test the imprinting effect, we can simply formulate the hypothesis as H<sub>0</sub> : i = 0 vs H<sub>1</sub> : i ≠ 0 to assess the mean difference of the two reciprocal heterozygotes. To test the cytoplasmic maternal effect, the hypothesis can be stated as H<sub>0</sub> : c = 0 vs H<sub>1</sub> : c ≠ 0. The epistatic effects of all interaction terms can also be tested as

$$\begin{cases} H_0 : & i_{ca} = i_{cd} = i_{ci} = 0 \\ H_1 : & \text{not all equal zero} \end{cases}$$

Similarly, additive and dominance effects can be tested as

$$\begin{cases} H_0 : & a = d = 0 \\ H_1 : & \text{not all equal zero} \end{cases}$$

If specific interest is focused, for instance, the interaction of imprinting and maternal effect, the hypothesis can be formulated as H<sub>0</sub> : i<sub>ci</sub> = 0 vs H<sub>1</sub> : i<sub>ci</sub> ≠ 0. All the above tests can be done by applying the likelihood ratio test in which the test statistic asymptotically follows a χ<sup>2</sup> distribution with degrees of freedom equal to the difference of the parameters under the null and the alternative hypotheses. For example, when testing H<sub>0</sub> : a = d = 0, the LR test statistics is compared with the χ<sup>2</sup> cutoff with 2 degrees of freedom.

**Results**

**Monte Carlo simulation**

Monte Carlo simulations were performed to investigate the statistical behavior of our model. We simulated an F<sub>2</sub> population, with one half of the population coming from design S<sub>1</sub> and the other half coming from design S<sub>2</sub>. A genome with 100 cM long linkage group, composed of 6 equidistant markers, was constructed. The position of QTL was assumed to be located at 48 cM away from the first marker on the linkage group. The marker genotypes in the F<sub>2</sub> population were simulated by mimicking sex-specific recombination fractions in mice, i.e., r<sub>M</sub> = 1.25r<sub>P</sub> [33]. A series of simulation study with different sample size (n = 400 vs n = 800) and different heritability levels (H<sup>2</sup> = 0.1, 0.25, 0.4) was conducted to examine the impacts of parameter spaces on parameter estimation and testing power. These simulation designs, which were aimed to give a better understanding of model

performance under different situations, can provide biologists some empirical evidences to design their experiments.

In the simulation study, the residual variance was calculated under different heritability levels. For the  $F_2$  design, the genetic variances of various terms can be calculated as follows:  $\sigma_a^2 = a^2/2$ ,  $\sigma_d^2 = d^2/4$ ,  $\sigma_i^2 = i^2/2$ ,  $\sigma_c^2 = c^2$ ,  $\sigma_G^2 = \sigma_a^2 + \sigma_d^2 + \sigma_i^2 + i_{ca}^2/2 + i_{cd}^2/2 + i_{ci}^2/2 + ci_{cd}$ , and the broad sense heritability level can be expressed as

$$H^2 = \sigma_G^2 / (\sigma_G^2 + \sigma_e^2).$$

For given genetic parameters and the heritability level, the residual variance can be calculated as  $\sigma_e^2 = (\sigma_Q^2 + \sigma_c^2)(1/H^2 - 1) - \sigma_c^2$ , from which the phenotype data can be generated.

The MLEs of the QTL position and effect parameters, based on 200 simulation replicates under different heritability levels and sample sizes, are displayed in Table 2. The square root of mean squared error (RMSE) of parameter estimates are given in parenthesis to show the estimation accuracy. As we expected, the accuracy of parameter estimates increases as the sample size and heritability level increase. For instance, the RMSE of estimated QTL position decreases from 13.28 to 3.81, an 71% increase in accuracy when the sample size increases from 400 to 800 under a fixed heritability level 0.1. The other parameter estimates show the same pattern. If we increase the heritability level when the sample size is fixed, a clear reduction in RMSE can be observed. For example, with 400 samples, the RMSE of estimated QTL position decreases from 13.28 to 4.99, then to 2.77 as  $H^2$  gradually increases from 0.1 to 0.4. From the decreasing RMSEs of the parameter estimates, we observed that simply increasing sample size is less efficient than increasing heritability level in order to increase the precision of parameter estimation. Since high heritability corresponds to small environmental variability [37], reducing environmental variation is of more practically important than just simply increasing sample size.

Note that we did not list the estimation of the imprinting effect  $i$  and cyto-imprinting interaction  $i_{ci}$ , which are not estimable under the  $F_2$  design. The reason is that the imprinting direction cannot

be inferred from the  $F_2$  design [25]. However, we can still conduct hypothesis test to infer the imprinting effect as well as its interaction with cytoplasmic effect. To further investigate the testing performance of cytoplasmic and imprinting effects, we introduced two proportions, namely  $\eta_c$  and  $\eta_i$ , where  $\eta_c = \sigma_c^2 / \sigma_G^2$  and  $\eta_i = \sigma_i^2 / \sigma_G^2$ . We can evaluate the test power under different cytoplasmic and imprinting effect sizes. Given all other genetic parameter values fixed (as shown in Table 2), simple algebra shows that the cytoplasmic effect  $c$  and imprinting effect  $i$  can be calculated for a given value of  $\eta_c$  or  $\eta_i$ , i.e.,

$$c = \frac{i_{cd}\eta_c + \sqrt{(i_{cd}\eta_c)^2 + 4\eta_c Q_1}}{2}, \text{ and } i = \sqrt{\frac{2\eta_i Q_2}{1 - \eta_i}}$$

where  $Q_1 = \sigma_a^2 + \sigma_d^2 + \sigma_i^2 + i_{ca}^2/2 + i_{cd}^2/2 + i_{ci}^2/2$ , and  $Q_2 = \sigma_a^2 + \sigma_d^2 + i_{ca}^2/2 + i_{cd}^2/2 + i_{ci}^2/2 + ci_{cd}$ .

Based on 1000 simulation runs under different heritabilities, sample sizes and variation proportions, the power of cytoplasmic effect test, imprinting effect test, interaction effects test and additive/dominance effects test are listed in Table 3. As we expect, the test power increases with the increasing of sample size and heritability level. For example, the cytoplasmic testing power increases from 0.663 to 0.905 as sample size increases from 400 to 800, a 36.5% increase in power for fixed  $\eta_c = 0.1$  and  $H^2 = 0.25$ . The same pattern is observed for the imprinting test. As the proportion of variance explained by the cytoplasmic and imprinting effect increases, the power increases accordingly. Noted that when both  $\eta_c$  and  $\eta_i$  are zeros, the testing power corresponds to the type I error rate for the corresponding factor. From the table we can see that the size of imprinting test is well controlled under different sample sizes and heritability levels. For the cytoplasmic effect, the size is a little inflated under  $n = 400$ , but it gets close to the nominal 5% level as sample size increases to 800, especially under large heritability (e.g.,  $H^2 = 0.4$ ). In sum, the simulation evidences show that the model performs reasonably well in both parameter estimation and testing.

**Table 2.** The MLEs of the QTL position and effect parameters based on 200 simulation replicates under different heritabilities and sample sizes.

$H^2$	$n$	Position 48 cM	$\mu = 10$	$c = 1$	$a = 1$	$d = 0.8$	$i = 0.6$	$i_{ca} = 0.6$	$i_{cd} = 0.5$	$i_{ci} = 0.4$	$\sigma^2$
0.1	400	47.93 (13.28)	10.00 (0.33)	0.99 (0.35)	0.96 (0.35)	0.78 (0.55)	-	0.59 (0.32)	0.49 (0.53)	-	3.679 (0.22)
	800	47.35 (3.81)	10.01 (0.23)	1.02 (0.21)	0.98 (0.24)	0.77 (0.39)	-	0.60 (0.22)	0.45 (0.33)	-	3.73 (0.15)
0.25	400	47.04 (4.99)	10.02 (0.17)	1.00 (0.18)	0.96 (0.16)	0.78 (0.26)	-	0.60 (0.17)	0.50 (0.26)	-	1.99 (0.11)
	800	47.65 (2.29)	10.00 (0.12)	1.01 (0.12)	0.98 (0.12)	0.81 (0.20)	-	0.59 (0.10)	0.50 (0.18)	-	2.01 (0.08)
0.4	400	48.02 (2.77)	9.99 (0.10)	1.00 (0.11)	0.99 (0.09)	0.81 (0.16)	-	0.60 (0.10)	0.50 (0.17)	-	1.23 (0.06)
	800	47.95 (1.78)	9.99 (0.07)	0.99 (0.07)	0.99 (0.08)	0.81 (0.12)	-	0.60 (0.07)	0.51 (0.11)	-	1.24 (0.04)

The squared roots of the mean squared errors (RMSE) of the MLEs are given in parentheses.

The locations of the QTL is described by the map distances (in cM) from the first marker of the linkage group. The hypothesized  $\sigma^2$  value is 3.81 for  $H^2 = 0.1$ , 2.04 for  $H^2 = 0.25$  and 1.26 for  $H^2 = 0.4$ .

doi:10.1371/journal.pone.0091702.t002

**Table 3.** The power of four hypothesis tests based on 1000 samplings under different heritabilities, sample sizes and variation proportions.

<i>n</i>	$H^2$	Power <sup>1</sup>			Power <sup>2</sup>			Power <sup>3</sup>	Power <sup>4</sup>
		$\eta_c = 0\%$	$\eta_c = 10\%$	$\eta_c = 20\%$	$\eta_i = 0\%$	$\eta_i = 10\%$	$\eta_i = 20\%$		
400	0.10	0.088	0.353	0.528	0.066	0.085	0.089	0.654	0.678
	0.25	0.081	0.663	0.928	0.051	0.263	0.432	0.990	0.992
	0.40	0.064	0.930	0.998	0.050	0.807	0.942	1.000	1.000
800	0.10	0.073	0.464	0.753	0.059	0.100	0.126	0.898	0.927
	0.25	0.066	0.905	0.998	0.047	0.473	0.713	1.000	1.000
	0.40	0.049	0.997	1.000	0.050	0.980	0.997	1.000	1.000

Power<sup>k</sup> ( $k = 1, 2, 3, 4$ ) refer to the powers for testing 1)  $H_0 : c = 0$  vs  $H_1 : c \neq 0$ ; 2)  $H_0 : i = 0$  vs  $H_1 : i \neq 0$ ; 3)  $H_0 : i_{ca} = i_{cd} = i_{ci} = 0$  vs  $H_1 : \text{not all equal to } 0$ ; and 4)  $H_0 : a = d = 0$  vs  $H_1 : \text{not all equal to } 0$ , respectively. For a given  $\eta_c$ , all other effect values are fixed as 0.8 except for  $c$ , which can be calculated in terms of  $\eta_c$  and other parameters. The hypothesized  $c$  value is 0 for  $\eta_c = 0$ , 0.461 for  $\eta_c = 0.1$  and 0.679 for  $\eta_c = 0.2$ . Similarly, the value of  $i$ , which depends on imprinting effect variation proportion  $\eta_i$ , is 0 for  $\eta_i = 0$ , 0.680 for  $\eta_i = 0.1$  and 1.020 for  $\eta_i = 0.2$ .

doi:10.1371/journal.pone.0091702.t003

### A case study

We applied the model to a published  $F_2$  cross data set based on design  $S_1$  and design  $S_2$  aimed to find QTLs that contribute to variation in quantitative traits related to colitis severity in 1L-10-deficient mice [38]. The data contain 411  $F_2$  mice derived from inbred strains, where 203 mice are from design  $S_1$  and 208 mice are from design  $S_2$ . Ninety-one markers were obtained with an average length of  $\sim 15$  cM spanning across the 19 autosome chromosomes. For more information about the data, the readers are referred to the original paper [38].

It has been reported that on average the female chromosome is 25% longer in genetic distance between homologous loci than the male in mice [38]. The sex-specific recombination fraction, expressed as  $r_M = 1.25r_P$ , was reconstructed based on the marker information (see Cui et al. for details [25]). The method was applied to four phenotypes, cecum total score (CTS), spleen/body weight ratio (SBWR), mesenteric lymph node (MLN)/body weight ratio (MBWR), and secretory IgA (SIgA) level, where cecum total score was graded by using colitis-related criteria, including severity, hyperplasia, ulceration and the percentage of area involved. Box-Cox transformation was applied to all traits before fitting the Gaussian-mixture model. The genome-wide LOD profile plots for the four phenotypes are shown in Figure 3, where the solid blue curves correspond to the LOD values and the dashed red lines correspond to the 5% genome-wide threshold values out of 1000 permutations. The LOD score is calculated as  $\log_{10} LR$ , where LR is obtained from equation (9) to test the null hypothesis:  $H_0 : a = d = i = i_{ca} = i_{cd} = i_{ci} = 0$ .

As shown in Figure 3, one QTL on chromosome 3 is detected for cecum total score trait, two QTLs on chromosomes 3 and one on chromosome 1 are detected for spleen/body weight ratio trait, three QTLs on chromosome 3 are detected for MLN/body weight ratio trait, and two QTLs located on chromosome 3 are detected for SIgA trait. The QTL located at 60.6 cM on chromosome 3 is common to three traits. The one located at 52.6 cM for SIgA trait is very close to it. It is highly possible that it is the same QTL that controls the four traits. Such a pleiotropic effect needs to be further evaluated. It should be mentioned that the QTL detected in the original paper for the four traits is located at 61.8 cM on chromosome 3 [38], which is 1.2 cM away from the one we found. Such a difference may arise because of the capitalization of sex-specific recombination rates and different models fitted.

In addition to the QTL identified in our analysis and the original paper, some other major QTLs which are not detected in

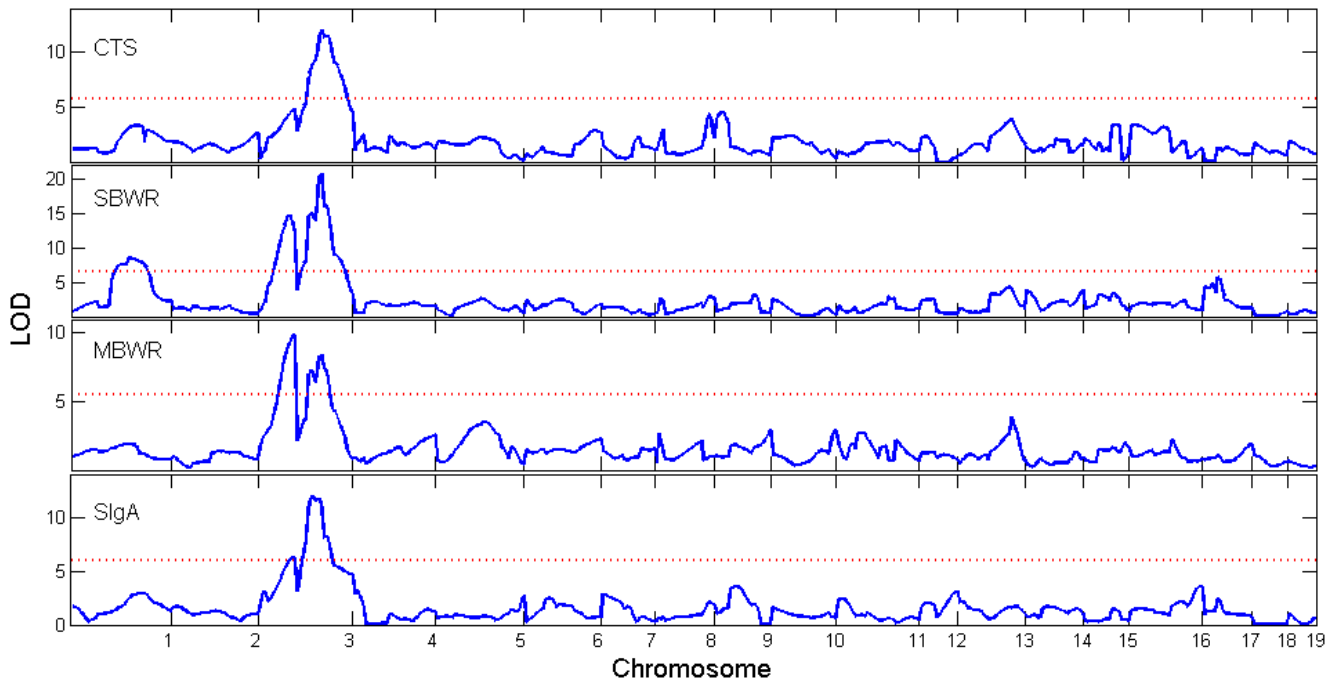
Farmer et al. [14], such as those at 52.6 cM and 38.6 cM on chromosome 3, stand out in our model and therefore need further investigation. Almost all the QTLs on chromosome 3 are clustered together, whose local LOD profiles of four traits are shown in Figure 4.

The marker interval for each QTL is listed in Table 4, which also tabulates the p-values under four different tests for each estimated QTL using permutation tests. From the test results, it can be seen that most QTLs have strong additive and dominance genetic effect, except for the spleen/body weight ratio QTL located at D1Mit156+31.1 cM on chromosome 1. This QTL shows evidence of cytoplasmic, imprinting as well as cyto-nuclear interaction effects (p-values for the three tests are 0.0282, 0.0296 and 0.0073, respectively), but shows no sign of additive and dominance effect. In addition, the MLN/body weight ratio QTL located at D3Mit78+5.6 cM shows evidence of cytoplasmic effect. In summary, we identified one QTL on chromosome 1 with evidence of cyto-nuclear interaction effect and this QTL also shows evidence of cytoplasmic and imprinting effect. Further functional validation is needed to confirm the results.

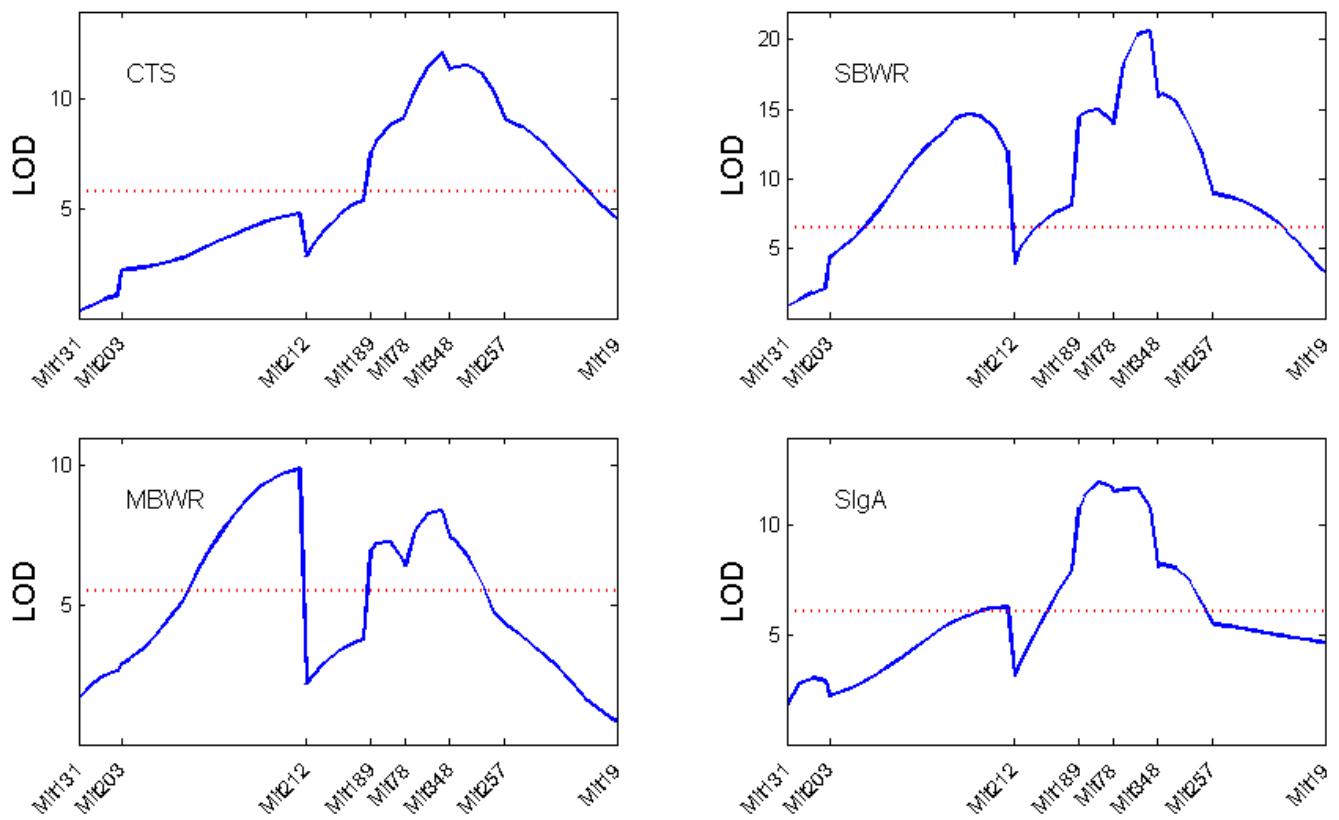
### Discussion

The cytoplasmic environment influences the expression of nuclear information in a very complicated way, which is still an unexplained mystery to many organisms. For example, Burgess and Husband have demonstrated great maternal contributions to the fitness of mulberry hybrids [39]. While it is an important parent-of-origin effect affecting offspring fitness, genomic imprinting, another source of parent-of-origin effect can also lead to phenotypic variation. Increasing evidence from cytoplasmic substitution and cell fusion experiments suggests that weakness of hybrids may connect with the interactions between cytoplasm and nuclear [40,41], and the evidences about interaction between cytoplasm and imprinting have also been observed [42,43]. As the source of genetic variation for many traits, these two types of parent-of-origin effects are often confounded, making it difficult to distinguish without proper statistical dissection. Although the role of cross-talk between the two sets of factors on phenotypic variation has been recognized, which genes are involved in the process and in what form they respond to the cytoplasmic changes are still unclear.

In this paper, we developed a statistical model to evaluate the cytoplasmic environment and nuclear gene interaction subject to imprinting effect within the framework of QTL interval mapping.



**Figure 3. The LOD profiles of the four traits across the 19 chromosomes using the linkage map constructed from microsatellite markers [38].** The genomic positions corresponding to the peak of the curve are the MLEs of the QTL locations.  
doi:10.1371/journal.pone.0091702.g003



**Figure 4. The local LOD profiles of the four traits across chromosomes 3.**  
doi:10.1371/journal.pone.0091702.g004

**Table 4.** List of QTL positions, corresponding marker intervals and p-values under different tests for the four traits.

Trait	Ch	QTL position	Marker interval	p-value <sup>1</sup>	p-value <sup>2</sup>	p-value <sup>3</sup>	p-value <sup>4</sup>
CTS	3	60.6 cM	D3Mit78-D3Mit348	0.6680	0.3046	0.5005	<10 <sup>-11*</sup>
			(D3Mit78+5.6 cM)				
SBWR	3	60.6 cM	D3Mit78-D3Mit348	0.4322	0.6107	0.4502	<10 <sup>-16*</sup>
			(D3Mit78+5.6 cM)				
	3	32.6 cM	D3Mit203-D3Mit212	0.2003	0.9502	0.2700	<10 <sup>-13*</sup>
			(D3Mit203+11.4 cM)				
	1	63.9 cM	D3Mit156-D1Mit17	0.0282*	0.0296*	0.0073*	0.8064
			(D1Mit156+31.1 cM)				
MBWR	3	38.6 cM	D3Mit203-D3Mit212	0.3574	0.2651	0.1658	<10 <sup>-9*</sup>
			(D3Mit203+27.4 cM)				
	3	60.6 cM	D3Mit78-D3Mit348	0.0102*	0.9708	0.4831	<10 <sup>-8*</sup>
			(D3Mit78+5.6 cM)				
	3	52.6 cM	D3Mit189-D3Mit78	0.2120	0.6421	0.7095	<10 <sup>-8*</sup>
			(D3Mit189+2.9 cM)				
SlgA	3	52.6 cM	D3Mit189-D3Mit78	0.1016	0.6488	0.3195	<10 <sup>-9*</sup>
			(D3Mit189+2.9 cM)				
	3	38.6 cM	D3Mit203-D3Mit212	0.7174	0.4051	0.8020	<10 <sup>-7*</sup>
			(D3Mit203+27.4 cM)				

p-value<sup>k</sup> ( $k = 1, 2, 3, 4$ ) refer to the p-values obtained with the likelihood ratio tests for testing 1)  $H_0 : c = 0$  vs  $H_1 : c \neq 0$ ; 2)  $H_0 : i = 0$  vs  $H_1 : i \neq 0$ ; 3)  $H_0 : i_{ca} = i_{cd} = i_{ci} = 0$  vs  $H_1$  : not all equal to 0; and 4)  $H_0 : a = d = 0$  vs  $H_1$  : not all equal to 0, respectively. Significant test results are indicated with the "\*" sign.  
doi:10.1371/journal.pone.0091702.t004

The model that considers eight genetic factors which measure the degree of imprinting, cytoplasmic, additive, dominance effects as well as the interaction effects among them, provides a complete dissection of cyto-nuclear epistasis subject to imprinting effect. A number of hypothesis tests can be performed not only to assess major genetic effect(s) responsible for phenotypic variation, but also to find the statistical evidence for the existence of imprinting, cytoplasmic effect as well as the cyto-nuclear interactions. Simulation study showed relative good performance of the model under the  $F_2$  design, in which parameters are estimated efficiently with modest heritability and sample size. Low heritability level ( $H^2 = 0.1$ ) and small sample size ( $n = 400$ ) result in large mean squared error of parameter estimation. This result is valuable in practice as we need to be careful about the interpretation of genetic effects obtained in real data analysis when the proportion of variance explained by the QTL is small (i.e., low heritability). Although our model cannot estimate the imprinting effect (so the cyto-imprinting interaction effect) due to the nature of the  $F_2$  design, existence test of imprinting (or cyto-imprinting interaction effect) can be achieved. Nevertheless, such imprinting estimation problem can be solved under the reciprocal backcross design illustrated in Figure 1.

In the real data analysis, one QTL located on chromosome 1 were found to have significant cytoplasmic, imprinting effect and cyto-nuclear interaction effects for spleen/body weight ratio. Other than that, no imprinting effect was found for all other QTLs, and only one on chromosome 3 that shows cytoplasmic effect for the MLN/body weigh ratio trait. It is worth mentioning that several QTL on chromosome 3 are detected by our model, but they are located relatively close to each other, as shown in Figure 3. For the sake of cautiousness, we reported all of them. However, these detected clustered QTLs may be due to the limitation of interval mapping, which can be overcome by fitting a

composite interval mapping model as following,

$$y_j = \mu + cx_{1j} + ax_{2j} + dx_{3j} + ix_{4j} + i_{ca}x_{5j} + i_{cd}x_{6j} + i_{ci}x_{7j} + \sum_{k=1}^K (a_k x_{2jk} + d_k x_{3jk} + i_k x_{4jk} + i_{ca_k} x_{5jk} + i_{cd_k} x_{6jk} + i_{ci_k} x_{7jk}) + e_j$$

where  $x_{2jk}, \dots, x_{7jk}$  are corresponding variables for the  $k$ th marker, assuming total  $K$  markers are selected for controlling background genetic effect. Although more variables are introduced in the model, theoretically some dimension-reduction techniques such as LASSO, can be applied to implement the variable selection for each marker before fitting them into the final model [44]. The composite interval mapping is know for its improved resolution in QTL detection. Regardless of the potential limitations mentioned above, the integration of cyto-nuclear interactions into the QTL mapping framework provides a testable platform with feasible experimental design for biologists to test the existence of cytoplasmic and imprinting effects, as well as the interactions of interested. The proposed model will have important biological implications with potentials to lift a corner of the great veil of the genetic system.

## Supporting Information

**Appendix S1 Detailed derivation of the EM algorithm.** (PDF)

## Author Contributions

Conceived and designed the experiments: YC. Analyzed the data: TH. Wrote the paper: TH YC. Participated in analysis: JS PSZ.



## References

- Lander E, Bostein E (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.
- Zeng Z (1994) Precision mapping of quantitative trait loci. *Genetics* 136: 228–235.
- Wu R, Ma C, Gallo-Meagher M, Littell R, Casella G (2002) Statistical methods for dissecting triploid endosperm traits using molecular markers: an autogamous model. *Genetics* 162: 875–892.
- Mackay T (2001) Quantitative trait loci in *Drosophila*. *Nat Rev Genet* 2: 11–20.
- Jirtle R, Skinner M (2007) Environmental epigenomics and disease susceptibility. *Nat Rev Genet* 8: 253–262.
- Isles A, Holland A (2005) Imprinted genes and mother-offspring interactions. *Early Hum Dev* 81: 73–77.
- Tycko B, Morison I (2002) Physiological functions of imprinted genes. *J Cell Physiol* 192: 245–258.
- Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2: 21–32.
- Falls J, Pulford D, Wylie A, Jirtle R (1999) Genomic imprinting: implications for human disease. *Am J Pathol* 154: 635–647.
- Clerget-Darpoux F, Babron MC, Deschamps I, Hors J (1991) Complementation and maternal effect in insulin-dependent diabetes. *Am J Hum Genet* 49: 42–48.
- Nanney D (1953) Nucleo-cytoplasmic interaction during conjugation in *Tetrahymena*. *Biol Bull* 105: 133–148.
- Haga N (1995) Elucidation of nucleus-cytoplasm interaction: change in ability of the nucleus to express sexuality according to clonal age in *Paramecium*. *J Cell Sci* 108: 3671–3676.
- Elfgang C, Rosorius O, Hofer L, Jaksche H, Hauber J, et al. (1999) Evidence for specific nucleocytoplasmic transport pathways used by leucine-rich nuclear export signals. *Proc Natl Acad Sci USA* 96: 6229–6234.
- Chinnery P (2003) Searching for nuclear-mitochondrial genes. *Trends Genet* 19: 60–62.
- Rand D, Haney R, Fry A (2004) Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol* 19: 645–653.
- Roubertoux P, Sluyter F, Carlier M, Marcet B, Maaroufveray F (2003) Mitochondrial DNA modifies cognition in interaction with the nuclear genome and age in mice. *Nat Genet* 35: 65–69.
- Tao D, Hu F, Yang J, Yang G, Yang Y, et al. (2004) Cytoplasm and cytoplasm-nucleus interactions affect agronomic traits in japonica rice. *Euphytica* 135: 129–134.
- Zeyl C, Andreson B, Weninck E (2005) Nuclear-mitochondrial epistasis for fitness in *Saccharomyces cerevisiae*. *Evolution* 59: 910–914.
- Rand D, Fry A, Sheldahl L (2006) Nuclear-mitochondrial epistasis and *Drosophila* aging: introgression of *Drosophila simulans* mtDNA modifies longevity in *D. melanogaster* nuclear backgrounds. *Genetics* 172: 329–341.
- Li S, Lu Q, Fu W, Romero R, Cui Y (2009) A regularized regression approach for dissecting genetic conflicts that increase disease risk in pregnancy. *Stat Appl Genet Mol Biol* 8: 45.
- Hanson R, Kobes S, Lindsay R, Knowler W (2001) Assessment of parent-of-origin effects in linkage analysis of quantitative traits. *Am J Hum Genet* 68: 951–962.
- de Koning DJ, Bovenhuis H, van Arendonk J (2002) On the detection of imprinted quantitative trait loci in experimental crosses of outbred species. *Genetics* 161: 931–938.
- Cui Y (2007) A statistical framework for genome-wide scanning and testing imprinted quantitative trait loci. *J Theo Biol* 244: 115–126.
- Cui Y, Cheverud J, Wu R (2007) A statistical model for dissecting genomic imprinting through genetic mapping. *Genetica* 130: 227–239.
- Cui Y, Lu Q, Cheverud J, Littell R, Wu R (2006) Model for mapping imprinted quantitative trait loci in an inbred F2 design. *Genomics* 87: 543–551.
- Li G, Cui Y (2010) A general statistical framework for dissecting parent-of-origin effects underlying triploid endosperm traits in flowering plants. *Ann Appl Stat* 4: 1214–1233.
- Tang Z, Wang X, Hu Z, Yang Z, Xu C (2007) Genetic Dissection of Cytonuclear Epistasis in Line Crosses. *Genetics* 177: 669–672.
- Haldane J (1922) The part played by recurrent mutation in evolution. *Am Nat* 67: 5–9.
- Huxley J (1928) Sexual difference of linkage *Grammarus chereuxi*. *J Genet* 20: 145–156.
- Dib C, Fauer S, Fizames C, Samson S, Drouot N, et al. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380: 152–154.
- Marklund L, Moller M, Hoyheim B, Davies W, Fredholm M, et al. (1996) A comprehensive linkage map of the pig based on a wild pig-Large White intercross. *Anim Genet* 27: 255–269.
- Neff M, Broman K, Mellers C, Ray K, Acland GM, et al. (1999) A second-generation genetic linkage map of the domestic dog, *Canis familiaris*. *Genetics* 151: 803–820.
- Dietrich W, Miller J, Steen R, Merchant M, Damron-Boles D, et al. (1996) A comprehensive genetic map of the mouse genome. *Nature* 380: 149–152.
- Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B* 39: 1–38.
- Manichaikul A, Dupuis J, Sen S, Broman K (2006) Poor Performance of Bootstrap Confidence Intervals for the Location of a Quantitative Trait Locus. *Genetics* 174.
- Churchill G, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963–971.
- Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates, Inc.
- Farmer M, Sundberg J, Bristol I, Churchill G, Li R, et al. (2001) A major quantitative trait locus on chromosome 3 controls colitis severity in IL-10-deficient mice. *PNAS* 98: 13820–13825.
- Burgess K, Husband B (2004) Maternal and paternal contributions to the fitness of hybrids between red and white mulberry. *Am J Bot* 91: 1802–1808.
- Levin D (2003) The cytoplasmic factor in plant speciation. *Syst Bot* 28: 5–11.
- Rhode J, Cruzan M (2005) Contributions of heterosis and epistasis to hybrid fitness. *Am Nat* 166: E124–E139.
- Moore T, Haig D (1991) Genomic imprinting in mammalian development: a parental tug-of-war. *Tren Genet* 7: 45–49.
- Cardoso M, Leonhardt H (1999) DNA Methyltransferase Is Actively Retained in the Cytoplasm during Early Development. *J Cell Bio* 147: 25–32.
- Li Z, Sillanpää M (2012) Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. *Theor Appl Genet* 125: 419–435.