

郑珂晖, 黎哲镇, 叶景山, 等. 基于自然群体随机交配的单个印迹 QTL 定位的初步研究 [J]. 福建农业学报, 2019, 34 (12): 1364–1370.  
ZHENG K H, LI Z Z, YE J S, et al. A Preliminary Study on Single Imprinted QTL Mapping Based on Random Mating of Natural Populations [J].  
*Fujian Journal of Agricultural Sciences*, 2019, 34 (12): 1364–1370.

## 基于自然群体随机交配的单个印迹 QTL 定位的初步研究

郑珂晖<sup>1,2</sup>, 黎哲镇<sup>2</sup>, 叶景山<sup>2</sup>, 周富杰<sup>2</sup>, 温永仙<sup>1,2\*</sup>

(1. 福建农林大学生命科学学院, 福建 福州 350002; 2. 福建农林大学计算机与信息学院, 福建 福州 350002)

**摘要:**【目的】探索基于自然群体随机交配的单个印迹 QTL 的定位方法, 分析影响定位准确性的关键因素。

【方法】若印迹 QTL 决定的某一性状为数量性状, 假设该性状与标记之间的关系存在线性关系, 可以采用最小二乘法进行印迹 QTL 定位和遗传参数的估计。利用计算机模拟单点模拟标记、水稻真实自然群体标记进行印迹 QTL 定位, 比较在不同最小等位基因频率(Minor allele frequency, MAF)、不同遗传率、不同随机交配轮数下的统计功效与参数估计精度, 印迹 QTL 的显著性采用 *F* 检验和 *t* 检验。【结果】通过模拟研究, 证明该试验设计对于检测单个印迹 QTL 是有效的, 在 MAF 大于 5% 时, 印迹遗传率大于 10% 时, 定位与遗传参数估计趋于无偏。【结论】采用自然群体随机交配产生作图群体, 可以用来进行单个印迹 QTL 的定位, 定位的结果较好, 是一种有效的试验设计, 为下一步进行多个印迹 QTL 奠定了基础。

**关键词:**自然群体; 关联分析; *F* 检验; 印迹 QTL; 随机交配

中图分类号: Q 348

文献标志码: A

文章编号: 1008-0384 (2019) 12-1364-07

## A Preliminary Study on Single Imprinted QTL Mapping Based on Random Mating of Natural Populations

ZHENG Ke-hui<sup>1,2</sup>, LI Zhe-zhen<sup>2</sup>, YE Jing-shan<sup>2</sup>, ZHOU Fu-jie<sup>2</sup>, WEN Yong-xian<sup>1,2\*</sup>

(1. College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China; 2. College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China)

**Abstract:**【Objective】To explore the application of single imprinted QTL mapping based on random mating of natural populations and analyze key factors affecting the mapping accuracy.【Method】Assuming the trait determined by the imprinted QTL to be quantitative and a linear relationship between the trait and the markers existed, then the least squares method could be used to estimate the QTL mapping and genetic parameters. The imprinted QTL mapping was generated by computer simulating the single-point markers as well as a real data set of natural population markers. The statistical power and parameter estimation accuracy of different genotype frequencies, heritability rates, and random mating rounds were compared. The significance of imprinted QTLs was determined by using *F*- and *t*-tests.【Result】Through a simulation study, it was proven that the experimental design was effective in detecting a single imprinted QTL. When the MAF was greater than 5% and the imprinted heritability greater than 10%, the iQTLs positions and genetic parameters were presumably unbiased.【Conclusion】Using random mating in the nature to generate mapping populations could conceivably be used to locate single-imprinted QTLs with acceptable results of localization. It seemed to be an effective design for the next step for QTL mapping.

**Key words:** Natural population; association analysis; *F*-test; imprinted QTL; random mating

收稿日期: 2019-09-25 初稿; 2019-10-27 修改稿

作者简介: 郑珂晖 (1979-), 男, 博士, 讲师, 研究方向: 生物信息学 (E-mail: 38193754@qq.com)

\* 通信作者: 温永仙 (1966-), 女, 博士, 教授, 研究方向: 生物信息学 (E-mail: wenyx9681@fafu.edu.cn)

基金项目: 国家自然科学基金项目 (31571558); 福建省自然科学基金项目 (2017J01606); 福建农林大学科技创新专项基金 (KFA17028A)

## 0 引言

**【研究意义】**自然群体往往不是严格的遗传群体, 如: 种质资源、野生种等, 它一般不用通过先构建亲本再进行交配产生 F<sub>2</sub> 代以后再用自交、回交等方法来产生。相对于结构群体, 它的遗传背景一般并不清晰, 往往采用关联分析来进行 QTL 定位。全基因组关联分析<sup>[1]</sup>主要基于基因的连锁不平衡, 用以判断表型性状与遗传标记之间的关联性。在全基因组上对逐点扫描遗传标记, 依次判断每个标记与相关表型性状的关联水平。关联分析方法出现以后, 有力地推动了数量遗传学的发展, 对于水稻<sup>[2-3]</sup>、玉米<sup>[4-5]</sup>等作物以及猪<sup>[6-7]</sup>、牛<sup>[8]</sup>等家畜的育种改良起了重要的推动作用。印迹效应是一种表观遗传效应, 指的是由于亲本产生的配子根据其来源父本或母本, 使得配子中某些基因呈现特异的表达或沉默, 造成子代体细胞中某些等位基因表达活性不同的现象。印迹在某些植物的胚乳形成<sup>[9]</sup>、哺乳动物的胚胎发育<sup>[10]</sup>、人类的一些罕见疾病<sup>[11]</sup>的形成中起着重要的作用。**【前人研究进展】**由于全基因组数据往往十分庞大, 关联分析对计算的要求较高。为了简化计算, 早期研究人员进行关联分析时往往将只考虑加性效应, 而忽略显性与印迹等效应。近些年来, 有部分研究人员将印迹效应引入到全基因组关联分析中。Marcos 等<sup>[12]</sup>采用 SNP 作为遗传标记, 采用回归方法来定位与猪背膘、日增重等性状相关联的 SNP, 并估计加性、显性和印迹 3 种遗传效应值。Hu 等<sup>[13]</sup>采用 GWAS 技术分析了印迹基因组对小鼠 BMI 指数的影响, 研究表明印迹可以世代传播, 对小鼠的 BMI 指数具有一定的影响。Steven 等<sup>[14]</sup>通过研究发现: 基因组印迹对在小鼠对于锥虫的抗性具有重要的作用, 并在小鼠染色体上定位到了与锥虫抗性相关的 3 个印迹 QTL。Wen 等<sup>[15]</sup>利用永久 F<sub>2</sub> 群体来进行印迹 QTL 定位, 可以无偏估计印迹 QTL 位置和加性、显性和印迹 3 种遗传效应。

**【本研究切入点】**自然群体的构建需要花费大量的人力与物力, 若将自然群体随机交配产生后代, 则相对可以用较低廉的价格获得遗传信息丰富的群体。目前, 基于自然群体随机交配的方法来定位印迹 QTL 目前尚未见到报道。由于基因组印迹具有亲本特异的特点, 所以, 在进行印迹 QTL 定位时, 需要知道某个株系的父本或母本。直接采用自然群体不能够直接进行印迹 QTL 的定位, 但可以将自然群体作为初始群体, 进行随机交配, 产生 F<sub>2</sub> 群体, 可以用较小的初始群体来获得一个大的 F<sub>2</sub> 群体, 该群

体包含的遗传重组事件较为丰富, 可以用作印迹 QTL 定位。**【拟解决的关键问题】**建立性状与加性、显性、印迹关系的遗传模型、给出单个印迹 QTL 定位与印迹 QTL 显著性检验的方法。

## 1 试验设计

假设有 1 个已知标记基因型的自然群体, 将自然群体植株间进行随机交配产生随机交配群体, 采用随机区组设计, 调查二倍体植株性状表型值, 生成随机交配群体如图 1 所示。

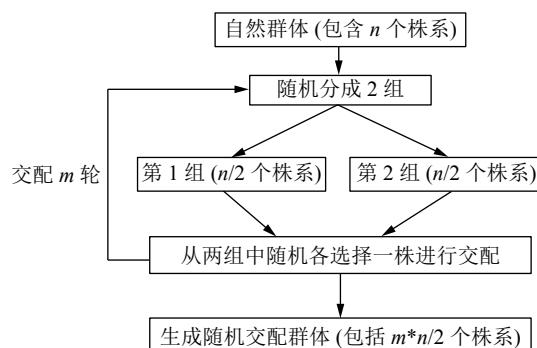


图 1 自然群体随机交配示意图

Fig. 1 Schematic diagram of random mating in natural populations

## 2 遗传模型与定位方法

假设基因组上某个位置上存在与表型性状关联的位点, 这个位点控制某个性状, 那么我们可以定义下列遗传模型:

$$y = \mu + ax + dz + ht + \varepsilon \quad (1)$$

式中,  $y$  为表型值,  $\mu$  为群体均值,  $a$  代表加性效应,  $d$  代表显性效应,  $h$  代表印迹效应,  $\varepsilon$  为随机误差, 服从正态分布  $N(0, \sigma^2_\varepsilon)$ ;  $x$ 、 $z$  和  $t$  为哑变量, 假设两个等位基因 A 和 a, A 的基因频率为  $p$ , a 的基因频率为  $q = 1 - p$ , 在随机交配群体中, 若考虑加性效应, 显性效应和印迹效应, 有 4 种基因型, 各自的频率如表 1 所示。

表 1 自然群体中基因型的频率与哑变量取值

Table 1 Frequency of genotypes and value of dummy variables in natural populations

基因型 Genotype	频率 Frequency			
		$a$	$d$	$h$
A/A	$p^2$	1	0	0
A/a	$pq$	0	1	1
a/A	$pq$	0	1	-1
a/a	$q^2$	-1	0	0

我们可以推导出，它的理论均值为：

$$M = a(p - q) + 2pqd \quad (2)$$

则其遗传方差为：

$$\begin{aligned} \sigma_G^2 &= [p^2a^2 + pq(d+h)^2 + pq(d-h)^2 + q^2a^2] - \\ &[a(p-q) + 2dpq]^2 = 2pq[a+d(q-p)]^2 + \\ &(2pqd)^2 + 2pqh^2 \end{aligned} \quad (3)$$

等位基因替代效应的平均效应 (the average effect of allele substitution) 可以通过式 4 计算得到：

$$A = a + d(q - p) \quad (4)$$

则总遗传方差为  $\sigma_G^2 = \sigma_A^2 + \sigma_d^2 + \sigma_h^2$ ，其中根据文献 [16-17]，加性方差、显性方差和印迹方差可以采用下列公式，加性方差为：

$$\sigma_A^2 = 2pq[a + d(q - p)]^2 \quad (5)$$

显性方差为：

$$\sigma_d^2 = (2pqd)^2 \quad (6)$$

印迹方差为：

$$\sigma_h^2 = 2pqh^2 \quad (7)$$

设  $\sigma_{Ad}^2 = \sigma_A^2 + \sigma_d^2$ ，若设印迹效应的遗传率为  $h_i^2$ ，则

$$h_i^2 = \frac{2pqh^2}{(\sigma_{Ad}^2 + 2pqh^2 + \sigma_e^2)} \quad (8)$$

若不考虑印迹效应，则其遗传方差为  $\sigma_G^2 = 2pq[a + d(q - p)]^2 + (2pqd)^2 = \sigma_{Ad}^2$ 。

### 3 单标记印迹 QTL 定位方法

假设有一个自然群体进行随机交配后，产生了个体数目为  $n$  的群体，为此我们建立了以下线性模型

$$y_i = \mu + ax_i + dz_i + ht_i + \varepsilon_i \quad i = 1, 2, \dots, n \quad (9)$$

式中， $y_i$  为表型值， $\mu$  为群体均值， $a$  代表加性效应， $d$  代表显性效应， $h$  代表印迹效应， $\varepsilon_i$  为随机误差，服从正态分布  $N(0, \sigma_e^2)$ ； $x_i$ 、 $z_i$  和  $t_i$  为哑变量，其值的确定见表 1。

为了简单起见，本研究采取单标记关联分析，采用最小二乘法回归分析，对基因组上的标记进行逐点扫描，对于式 9，我们提出假设  $H_0 : a = d = h = 0$ ，其备择假设  $H_1 : a, d$  和  $h$  不全为零。若仅检验印迹效应，我们提出假设  $H_0 : h = 0$ ，其备择假设  $H_1 : h \neq 0$ ，我们采用  $F$  检验来检测标记的显著性，因为是单个标记的检验，直接用 0.05 或 0.01 下的  $F$  临界值进行判断。由于这里有 3 种遗传效应，那么显然的，如果：

对于给定的显著性水平，可以构建统计量，该统计量服从  $F(p, n-p-1)$  的  $F$  分布，则有式 10。

$$\frac{RSS/p}{ESS/(n-p-1)} > F_\alpha(p, n-p-1) \quad (10)$$

式中， $RSS$  代表随机残差平方和， $ESS$  为回归平方和， $n$  为样本数， $p$  为变量个数，此处取值为 3。如果方程总体通过了显著性检验以后，再逐个对  $x$ 、 $z$ 、 $t$  做  $t$  检验，检验各变量是否显著。 $F$  统计量与  $t$  统计量计算简单，使用计算机运算速度较快，可以逐点计算全基因组上每个位点的  $F$  值，并对每个位点的 3 种效应估计值进行  $t$  检验，将  $F$  最大（超过临界值，且通过  $t$  检验）的点作为可能的印迹 QTL 位点。

### 4 模拟研究

为了验证该方法的可行性，我们采用计算机来模拟验证，由于本文仅是基于自然群体随机交配的进行印迹 QTL 定位方法的初步研究，目的在探讨其可行性，所以进行 2 个模拟：第一是采用单点模拟，假设已知 QTL 位置，仅考察其参数估计的准确度和精度，及其是否达到显著性；第二是采用实际的水稻的基因型数据进行模拟，验证其可行性。

#### 4.1 单点模拟研究

由于自然群体有别于结构群体，自然群体的遗传结构是不清晰的，在群体中有大量的进化事件掺杂其中，群体结构较为复杂。在群体中，等位基因  $A$ ， $a$  的基因频率往往不是 1:1，甚至有可能是很极端的，如可能在 0.05 以下，就是罕见变异。那么自然群体中在不同基因频率下，基于自然群体随机交配产生的群体，我们所建立的遗传模型和提出相应地印迹 QTL 检测方法与效应估计方法是否还能有效，我们通过模拟来进行检验。在这里，仅考察所建立的遗传模型在不同的 MAF 下相应地参数估计和显著检验方法是否可行。

我们假定仅有单点标记，在遗传模型中

$$y_i = \mu + ax_i + dz_i + ht_i + \varepsilon_i \quad i = 1, 2, \dots, n \quad (11)$$

进行参数估计，在此基础上，首先我们对方程进行  $F$  检验，看其是否存在 QTL ( $a, d$  和  $h$  不全为零)，在此基础上进一步对印迹效应做显著性检验 ( $h \neq 0$ )。

根据式 9 可以计算出印迹效应的大小如式 12 所示。

$$h = \sqrt{\frac{h_i^2(\sigma_{Ad}^2 + \sigma_e^2)}{(1 - h_i^2)2pq}} \quad (12)$$

若令  $C = \sigma_{Ad}^2 + \sigma_e^2 = \sigma_A^2 + \sigma_d^2 + \sigma_e^2$ ，在模拟中可以采用如下设置： $C$  取一常数，给定印迹遗传率，从而计算出印迹效应，如式 13 所示：

$$h = \sqrt{\frac{Ch_i^2}{(1-h_i^2)2pq}} \quad (13)$$

再计算环境方差, 最终产生表型值, 这样就可以考虑遗传率的不同, 随机交配次数不同的功效, 参数估计等。

文献 [18] 根据印迹与加性、显性的关系将它们分成 5 种印迹类型, 如表 2 所示。

对这 5 种类型的 iQTL 进行模拟定位并对模拟结果进行讨论, 由于在 GWAS 研究中, 研究人员关心的往往是 MAF 大于 0.05 的这部分标记, 所以, 将 MAF, 即  $p$  设置 0.05、0.1 两种情况, 设  $q=1-p$ , 常数  $C$  取 4、10 两种情况, iQTL 的遗传率设置为 5%、10%、15% 三个水平, 印迹效应根据式 12 进行推算。根据表 2 中的 5 种印迹类型得出  $a$ 、 $d$  的值。假设自然群体作为初始群体, 包含 100 个株系, 利用该自然群体进行随机交配 6、20 轮产生 F2 群体。每

表 2 印迹类型与参数设置与缩写

Table 2 Imprint types, parameter settings, and abbreviations

印迹类型 Imprinting type	简写 abbreviation	参数 parameter
来源于父本的表达 Parental expression, Paternal	PEP	$d=0$ 且 $a=h$
来源于母本的表达 Parental expression, Maternal	PEM	$d=0$ 且 $a=-h$
显性印迹(双相) Dominance imprinting, Bipolar	DIB	$a=0$ 且 $d=0$
显性印迹(单相)超显性 Dominance imprinting, Polar, Over-dominance	DIPOD	$a=0$ 且 $d=h$
显性印迹(单相)低于显性 Dominance imprinting, Polar, Under dominance	DIPUD	$a=0$ 且 $d=-h$

种情况重复 100 次, 100 次参数估计的标准误差提供了参数估计误差标准, 显著性水平取为 0.05, 限于篇幅, 此处仅列出 C=10, 交配 20 轮的结果, 结果见表 3。

表 3 单点模拟 5 种类型的 iQTL 定位结果

Table 3 Single-point simulation of 5 types of iQTL mapping results

MAF/%	H/%	DIPOD				DIPUD				DIB				PEM				PEP			
		<i>a</i>	<i>d</i>	<i>h</i>	P/%	<i>a</i>	<i>d</i>	<i>h</i>	P/%	<i>a</i>	<i>d</i>	<i>h</i>	P/%	<i>a</i>	<i>d</i>	<i>h</i>	P/%	<i>a</i>	<i>d</i>	<i>h</i>	P/%
RV		0	2.35	2.35		0	2.35	-2.35		0	0	2.35		2.35	0	-2.35		2.35	0	2.35	
5	5	0.18 (1.17)	2.2 (1.22)	2.4 (0.31)	85	0.07 (1.22)	2.25 (1.23)	-2.4 (0.33)	93	0.12 (1.19)	-0.1 (1.24)	2.35 (0.37)	88	2.41 (1.16)	-0.06 (1.22)	-2.35 (0.31)	91	2.16 (1.04)	0.24 (1.13)	2.35 (0.29)	87
RV		0	3.42	3.42		0	3.42	-3.42		0	0	3.42		3.42	0	-3.42		3.42	0	3.42	
5	10	0.09 (1.19)	3.31 (1.21)	3.45 (0.35)	83	0 (1.01)	3.47 (1.06)	-3.38 (0.31)	82	-0.02 (1.19)	-0.02 (1.25)	3.43 (0.35)	86	3.46 (1.34)	0.03 (1.35)	-3.45 (0.35)	92	3.37 (1.03)	0.1 (1.05)	3.38 (0.29)	86
RV		0	4.31	4.31		0	4.31	-4.31		0	0	4.31		4.31	0	-4.31		4.31	0	4.31	
5	15	0.07 (1.2)	4.28 (1.25)	4.34 (0.27)	90	-0.01 (1.18)	4.31 (1.26)	-4.35 (0.29)	90	0.03 (1.24)	-0.05 (1.24)	4.25 (0.33)	90	4.36 (1.03)	-0.05 (1.03)	-4.31 (0.3)	88	4.42 (1.03)	-0.08 (1.08)	4.33 (0.32)	85
RV		0	1.71	1.71		0	1.71	-1.71		0	0	1.71		1.71	0	-1.71		1.71	0	1.71	
10	5	-0.02 (0.53)	1.75 (0.6)	1.72 (0.21)	100	-0.04 (0.53)	1.75 (0.55)	-1.73 (0.21)	100	-0.04 (0.57)	0.06 (0.63)	1.72 (0.23)	100	1.64 (0.47)	0.06 (0.58)	-1.67 (0.24)	100	1.69 (0.57)	-0.02 (0.63)	1.69 (0.22)	100
RV		0	2.48	2.48		0	2.48	-2.48		0	0	2.48		2.48	0	-2.48		2.48	0	2.48	
10	10	-0.05 (0.55)	2.52 (0.61)	2.5 (0.23)	100	-0.05 (0.57)	2.53 (0.63)	-2.51 (0.25)	100	-0.02 (0.61)	0.03 (0.7)	2.51 (0.26)	100	2.54 (0.51)	-0.04 (0.56)	-2.51 (0.2)	100	2.59 (0.6)	-0.06 (0.62)	2.52 (0.23)	100
RV		0	3.13	3.13		0	3.13	-3.13		0	0	3.13		3.13	0	-3.13		3.13	0	3.13	
10	15	-0.04 (0.49)	3.18 (0.57)	3.12 (0.23)	100	-0.08 (0.52)	3.22 (0.57)	-3.13 (0.23)	100	-0.08 (0.62)	0.06 (0.65)	3.13 (0.26)	100	3.24 (0.47)	-0.05 (0.53)	-3.12 (0.22)	100	3.18 (0.55)	0.01 (0.59)	3.12 (0.24)	100

注: 表中括号内为标准差, 此处列出的是 C=10, 交配轮数为 20 的结果, 表中 H 代表遗传率 (Heritability), RV 代表真值 (Real value), P 代表统计功效 (Power)

Note: The number in parentheses in the table is the standard deviation, C=10 and cross rounds is 20, H represents heritability, RV means real value and p means statistical power.

结果表明：在 MAF 较小，为 0.05 时，即使 iQTL 的遗传率达到 15%，其检测功效和参数估计的标准误都较高，当 MAF 达到 0.1 以上时，其检测功效就较高了；随着 MAF 的增大，遗传率的增加，其估计的标准误逐渐减小。在和遗传率相同的条件下，随着 C 值的增加，其参数估计的标准误也增加，其原因还有待于进一步的研究。即使在 MAF 高于 0.05 时，当遗传率为 5% 时，效应值估计有一定的误差，检测功效也较低，当遗传率在 10% 以上的时候，效应估计就较为精准。所以可以看出基因频率和遗传率是影响 iQTL 定位效果的因素。此外，随着交配次数的增加，iQTL 的检测功效随之增加，参数估计的标准误随之减少，从 5 种 iQTL 来看，PEP、PEM、DIPOD、DIPUD 在 MAF、遗传率、C 值相同的情况下，其 iQTL 检测效果相似，然而 DIB 的 iQTL 检测效果略差些，这说明加性效应和显性效应影响了 iQTL 定位。

综上所述，如果某个性状的表型值由一个 iQTL 确定，那么不论它是上述五种类型中的哪种类型，只要在 MAF 大于 0.1 且遗传率大于 10% 时，其检测功效均较好，说明我们的提出的方法在自然群体随机交配产生的 F2 群体上定位单个 iQTL 是有效的。

#### 4.2 基于水稻基因型数据单个位点的模拟研究

为了验证该方法在实际中的实施情况，我们采用韩斌院士课题组发布的水稻自然群体的数据集<sup>[19]</sup>，该数据集收集了 960 多种水稻种质资源。为了简化研究的问题，对原始数据进行了预处理，把其中粳稻的 254 个株系取出，并将 MAF 小于 0.05 的 SNP 位点去掉，并去除了所有的缺失位点，最终得到了一个包含 16 382 个 SNP，在本模拟中，假定这些 SNP 是等间距分布在一条染色体上，每个 SNP 间的距离为 1 kb，254 个株系的数据集作为原始群体。原始群体采用随机交配的方式生成随机交配群体，在进行模拟前，统计各个位点的 SNP 频率。从图 2 可知，在自然群体中，每个位置上的 SNP 的 MAF 并不是均衡的。

假定一个位于 470 kb 位置的这个 SNP 与某个表型性状关联，群体均值为  $\mu=10$ 。模拟前，先计算该位点的 SNP 频率  $p$  和  $q$ ，印迹效应根据公式 12 进行推算（C 值取 2, 4, 6 三种组合）。根据表 2 中的 5 种印迹类型得出  $a, d$  的值，根据前面的模拟结果，iQTL 的遗传力水平设置为 10% 和 15%，利用自然群体进行 3、5、7 次随机交配，产生随机交配群体。每种模拟情况重复 100 次，100 次参数估计的标准误偏差提供了参数估计误差标准，阈值可用 Bonferroni

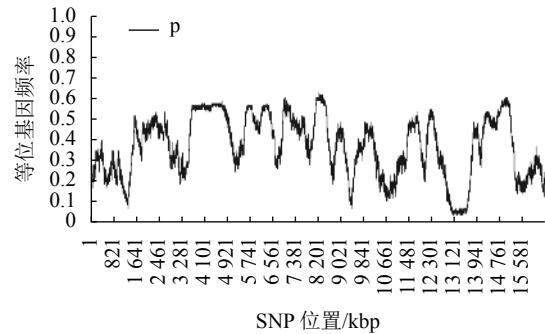


图 2 水稻自然群体随机交配产生的群体 SNP 等位基因频率 (p) 示例 (包含 254 个不同株系)

Fig. 2 Examples of SNP allele frequencies (p) in populations from random mating of natural populations of rice (including 254 different lines)

校正，显著性水平为 0.05。此处，限于篇幅，仅列出 C 值取 2、6，交配轮数为 7 次的结果，见表 4，交配以后得到的随机交配群体中 470 kb 处的 SNP 的频率  $p$  为 0.54，则  $q$  为 0.46，接近于 1:1。由于在模拟时选择的 iQTL 位置上的 MAF 接近于 0.5，从真实数据的模拟结果来看：当遗传率为 10%、15% 时，5 种类型的 iQTL 定位都具有较好的结果，iQTL 所在位置的估计较为准确，各参数估计的准确度和精度达到较好的效果，而且检测功效也较高，在 15% 时所有类型的 iQTL 检测功效均为 100%。结果表明，随着交配次数的增加，参数估计的精度也得到提高，这符合我们的构想，当样本量增加时，其估计的准确度和精度得到提高；对于常数 C 值，从表可以看出，随着 C 值的增加，其参数估计的精度逐渐下降，其原因还有待于进一步研究；从 5 种类型的印迹 QTL 来看，在相同的条件下，DIB 的检测功效较差，而其余 4 种 DIPOD, DIPUD, PEM, PEP 检测的功效相近，由此可见：加性效应和显性效应对 iQTL 的检测具有一定的影响。

## 5 讨论

我们提出了基于自然群体随机交配产生的 F2 群体进行 iQTL 定位的思想，构建其初步的定位方法，通过单点模拟，全基因组的单点模拟说明基于自然群体随机交配产生的群体进行 iQTL 定位是可行的，当然这里没有考虑其他因素（如群体结构等）的干扰。模拟表明，在不考虑其他因素的影响下，该 iQTL 定位方法的检测功效，以及参数估计的准确度和精度都较好。

虽然自然群体的群体遗传结构不清晰，但在 GWAS 时代，测序成本不断降低，甚至可以直接对个体进行测序。所以，基因型频率可以看成已知

表4 不同C值下5种印迹类型的iQTL定位结果  
Table 4 iQTL mapping results of 5 imprint types under varied C values

H C	DIPOD				DIPUD				DIB				PEM				PEP					
	POS/kbp	a	d	h	POS/kbp	a	d	h	P%													
RV	470	0	0.73	0.73	470	0	0.73	-0.73	470	0	0	0.73	470	0.73	0	-0.73	470	0.73	0	0.73		
10 2	470	-0.01	0.74	0.73	470	0.01	0.74	-0.72	470	0	0	0.74	470	0.73	0	-0.73	470	-0.01	0.74	0.73	100	
	0	(0.1)	(0.13)	(0.07)	0	(0.09)	(0.13)	(0.07)	0	(0.1)	(0.14)	(0.08)	0	(0.1)	(0.12)	(0.08)	0	(0.1)	(0.1)	(0.1)		
10 6	RV	470	0	1.26	1.26	470	0	1.26	-1.26	470	0	0	1.26	470	1.26	0	-1.26	470	1.26	0	1.26	
	470	0.02	1.25	1.28	470	-0.04	1.25	-1.28	470.02	0.01	0.01	1.26	470	1.27	0	-1.25	470	0.02	1.25	1.28	100	
15 2	RV	470	0	0.93	0.93	470	0	0.93	-0.93	470	0	0	0.93	470	0.93	0	-0.93	470	0.93	0	0.93	
	470	-0.01	0.92	0.93	470	0	0.9	-0.93	470	-0.01	0	0.92	470	0.95	0	-0.92	470	-0.01	0.92	0.93	100	
15 6	RV	470	0	1.6	1.6	470	0	1.6	-1.6	470	0	0	1.6	470	1.6	0	-1.6	470	1.3	0	1.6	
	470	0	1.59	1.59	470	0	1.62	-1.63	470	0.01	0.01	1.6	470	1.59	-0.03	-1.62	470	0	1.59	1.59	100	
	0	(0.17)	(0.2)	(0.13)	0	(0.18)	(0.22)	(0.13)	0	(0.16)	(0.19)	(0.12)	0	(0.17)	(0.2)	(0.12)	0	(0.17)	(0.2)	(0.1)		

注: 表中括号内为标准差, 此处列出C=2、6, 交配轮数为7次的结果, 表中H代表遗传率, H代表遗传率(Heritability), C代表C值(C value), RV代表真值(Real value), P代表统计功效(Power), POS代表iQTL在染色体上的位置(iQTL position)。

Note: The number in parentheses in the table is the standard deviation, the results listed here are the C values are 2 and 4, the number of mating rounds is 7, C represents C-value, RV represents real value, P represents statistical power and POS means the iQTL's position on chromosome.

的, 基于自然群体随机交配产生的群体, 只要交配次数足够, 就可以产生较大的样本; 从初步研究模拟也可以看出, 随着交配次数的增加, 其检测功效, 参数估计的精度也增加。

在全基因组关联分析(GWAS)中, 是以长期重组后保留下来的基因或位点间连锁不平衡为基础, 在获得群体表型数据和基因型数据后, 采用统计方法检测标记与性状之间关联的分析方法, 一般以自然群体或种质资源为研究材料, 遗传结构较为复杂。在进行GWAS时, 要求遗传背景一致或者相似的群体, 但在未知群体结构的关联分析研究中, 由不同亚群引起的MAF的差异会造成假关联。通常采用基因组控制、基于家系的关联分析检验和主成分分析等方法控制。一般而言: 基于自然群体随机交配产生的群体会产生结构分层, 所以在这种情况下, 必须要控制因为结构分层造成的假关联, 在后续的工作中, 可以利用混合线性模型, 传递不平衡分析等其他方法来进行印迹QTL定位方法的研究。

## 参考文献:

- [1] 涂欣, 石立松, 汪樊, 等. 全基因组关联分析的进展与反思[J]. 生理科学进展, 2010, 41 (2): 87–94.  
TU X, SHI LS, WANG F, WANG Q. Genomewide Association Study:

Advances, Challenges and Deliberation [J]. *Progress in Physiological Sciences*, 2010, 41 (2) : 87–94. (in Chinese)

- [2] LI F M, XIE J Y, ZHU X Y, et al. Genetic basis underlying correlations among growth duration and yield traits revealed by GWAS in rice(*Oryza sativa* L.) [J]. *Frontiers in Plant Science*, 2018, 9: 650.  
[3] WANG H R, XU X, VIEIRA F G, et al. The power of inbreeding: NGS-based GWAS of rice reveals convergent evolution during rice domestication [J]. *Molecular Plant*, 2016, 9 (7) : 975–985.  
[4] 田润苗, 张雪海, 汤继华, 等. 玉米种子萌发相关性状的全基因组关联分析[J]. *作物学报*, 2018, 44 (5) : 672–685.  
TIAN R M, ZHANG X H, TANG J H, et al. Genome-wide association studies of seed germination related traits in maize [J]. *Acta Agronomica Sinica*, 2018, 44 (5) : 672–685. (in Chinese)  
[5] 邵晓宇, 宋希云, 潘顺祥, 等. 玉米穗粗性状的全基因组关联分析及QTL元分析 [J]. *植物生理学报*, 2017, 53 (12) : 2091–2102.  
SHAO XY, SONG XY, PAN S X, et al. Genome-wide association study and Meta-QTL analysis of ear diameter trait in maize [J]. *Plant Physiology Communications*, 2017, 53 (12) : 2091–2102. (in Chinese)  
[6] WANG K J, LIU Y F, XU Q, et al. A post-GWAS confirming GPAT3 gene associated with pig growth and a significant SNP influencing its promoter activity [J]. *Animal Genetics*, 2017, 48 (4) : 478–482.  
[7] UZZAMAN M R, PARK J E, LEE K T, et al. A genome-wide association study of reproductive traits in a Yorkshire pig population [J]. *Livestock Science*, 2018, 209: 67–72.

- [8] IBEAGHA-AWEMU E M, PETERS S O, AKWANJI K A, et al. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits [J]. *Scientific Reports*, 2016, 6: 31109.
- [9] JESSICA A R, DANIEL Z. Evolution and function of genomic imprinting in plants [J]. *Genes & Development*, 2015, 29 ( 24) : 2517–2531.
- [10] PILVAR D, REIMAN M, PILVAR A, et al. Parent-of-origin-specific allelic expression in the human placenta is limited to established imprinted loci and it is stably maintained across pregnancy [J]. *Clinical Epigenetics*, 2019, 11: 94.
- [11] PETERS J. The role of genomic imprinting in biology and disease: an expanding view [J]. *Nature Reviews Genetics*, 2014, 15 ( 8) : 517–530.
- [12] LOPES M S, BASTIAANSEN J W M, JANSS L, et al. Estimation of additive, dominance, and imprinting genetic variance using genomic data [J]. *G3-Genes Genomes Genetics*, 2015, 5 ( 12) : 2629–2637.
- [13] HU Y D, ROSA G J, GIANOLA D. A GWAS assessment of the contribution of genomic imprinting to the variation of body mass index in mice [J]. *BMC Genomics*, 2015, 16: 576.
- [14] CLAPCOTT S J, TEALE A J, KEMP S J. Evidence for genomic imprinting of the Major QTL controlling susceptibility to trypanosomiasis in mice [J]. *Parasite Immunology*, 2000, 22 ( 5) : 259–263.
- [15] WEN Y X, WU W R. Mapping of imprinted quantitative trait loci using immortalized F2 populations [J]. *PLoS One*, 2014, 9 ( 3) : e92989.
- [16] DE KONING D J, BOVENHUIS H, VAN ARENDONK J A. On the detection of imprinted 513 quantitative trait loci in experimental crosses of outbred species [J]. *Genetics*, 2002, 161 ( 2) : 931–938.
- [17] FALCONER D S, MACKAY T F C. An Introduction to Quantitative Genetics[M]. Ed. 4. Longman Group, Essex, UK, 1996:12-28.
- [18] CHEVERUD J M, HAGER R, ROSEMAN C, et al. Genomic imprinting effects on adult body composition in mice [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105 ( 11) : 4253–4258.
- [19] HUANG X H, ZHAO Y, WEI X H, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm [J]. *Nature Genetics*, 2012, 44 ( 1) : 32–39.

(责任编辑：林海清)