

玉米产量及构成因子主效和上位性 QTL 的全基因组扫描分析

严建兵 汤华 黄益勤 郑用琰 Subhash Chander 李建生*

(中国农业大学国家玉米改良中心, 北京 100094; 华中农业大学作物遗传改良国家重点实验室, 武汉 430070.

* 联系人, E-mail: lijs@163bj.com)

摘要 在此前构建的含 174 个分子标记的连锁图基础上, 新增 31 个分子标记, 并利用统计分析的方法对基因型数据进行检测, 剔除基因型明显有误的数据, 构建了共包含 205 个分子标记, 平均间距为 11.2 cM, 覆盖玉米全基因组 10 条染色体的分子标记连锁图. 基于统计学的基因型数据检测, 能显著提高连锁图构建的准确性. 结合一年两点 6 个重复的田间实验, 利用软件 R/qtl 进行主效和上位性 QTL 分析, 同时利用多区间作图法(MIM)验证主效 QTL 和主效 QTL 互作的真实性和位置. 通过 MIM 方法, 检测到控制产量、行粒数、行数和百粒重的主效 QTL 分别为 5, 5, 7 和 5 个, 分别能解释 35.3%, 37.4%(含 1 个互作 QTL), 61.5%和 39.7%的遗传变异; 除行粒数的 2 个 QTL 之间存在显著性互作之外, 其他性状的主效 QTL 之间都没有检测到显著性互作存在. 同时, 还检测到 24 个上位性互作 QTL, 这些 QTL 涉及的位点几乎分布于所有 10 条染色体. 这些互作 QTL 以主效 QTL 与无显著效应基因座之间的互作居多, 对所有 4 个性状而言比例接近三分之二. 这也说明主效 QTL 除了单独起作用之外还可以通过与其他没有显著效应的基因座之间互作来影响性状表达. 结果表明, 上位性 QTL 在玉米产量性状的遗传中可能与主效 QTL 一样也起着重要的作用.

关键词 分子标记 上位性 QTL 数量性状遗传

玉米是世界上最重要的作物之一, 杂交玉米的推广和应用每年为世界多提供了 5500 万吨粮食^[1]. 随着世界人口的不断增长和耕地面积的持续减少, 人类对提高单位面积粮食产量的需求要求我们不断深入对产量性状遗传的研究. 过去 20 多年来, 分子标记技术以及 QTL 定位方法的快速发展为复杂数量性状的研究提供了强有力的手段. 对产量等复杂数量性状遗传机制的了解也越来越深入, 利用合适的分离群体构建覆盖全基因组的分子标记连锁图, 同时结合多年多点的田间实验, 定位了大量影响产量、农艺、品质、生物和非生物抗性等性状的 QTL 和基因. 迄今为止, 禾本科两个主要数据库(www.gramene.org; www.maizegdb.org)收录的水稻和玉米 QTL 分别超过 6000 和 2000 个. 对这些 QTL 的利用和研究, 一方面是通过对一些主效 QTL 进行分子标记辅助选择, 加速育种进程, 直接为生产实践服务^[2]; 另一方面是对一些重要的 QTL 进行克隆, 深入剖析和认识数量性状位点背后的遗传机制. 目前, 克隆的典型 QTL 有控制番茄果重的 QTL *fw2.2*^[3], 控制水稻开花期的 QTL *hd1* 及 *hd6*^[4,5], 控制产量的 QTL *Gn1a*^[6], 控制

耐盐的 QTL *SKC1*^[7] 以及控制玉米进化的重要 QTL *tb1*^[8] 等.

主效 QTL 的定位和克隆无疑有助于我们了解数量性状的遗传基础, 但这些都是基于遗传学的单基因理论模型. 基于生物化学和发育遗传学的认识, 基因产物之间应该存在大量的互作^[9]. 最近也有人借助大量转录水平的数据发现基因产物之间的二级、三级和高级互作广泛存在^[10]. 还有证据表明, 有着共同起源物种之间基因的互作是物种进化和物种形成的基础^[11,12]. 基于这些事实, 上位性应当是数量性状的遗传基础. 近年来, 通过分子标记作图和分离群体的田间实验, 不同研究小组在不同的物种中, 如豇豆、绿豆^[13] 及大豆^[14] 等, 都发现上位互作的普遍存在. 尤其对水稻的研究表明, 上位性对水稻产量性状的遗传和杂种优势的遗传基础起重要作用^[15-18]. 在玉米数量性状的遗传研究中, 前人有许多利用经典统计分析的数据支持上位性大量存在的假设^[11,19-21]. 也有研究表明, 控制玉米不同数量性状的 QTL 之间也存在显著的互作^[22-26]. 此前, 我们利用覆盖玉米全基因组的 174 个分子标记和双向方差分析也证明,

上位性在玉米产量性状的遗传和杂种优势的形成中起重要作用¹⁾。但基于标记互作的双向方差分析常常会高估上位性的数目和效应。本研究利用一个优良玉米杂交种所衍生的F_{2:3}分离群体和覆盖全基因组的分子标记连锁图及一个改进的分析软件R/qtl, 同时进行主效和上位性QTL分析, 以研究玉米产量性状的遗传基础。

1 材料与方法

() 供试材料. 本研究选用近几年来在生产上广泛推广的优良玉米杂交组合豫玉 22 (综 3×87-1) 为基础材料。经F₁自交得到F₂。在F₂群体中随机套袋自交获得 266 个单穗, 并以此发展的F_{2:3}家系作为本实验的供试群体材料。同时, 在幼株期采集相应F₂单株的叶片, 保存在-70 °C 冰箱中, 供提取DNA用。

() 田间实验和调查. 2000 年春, 温室和大棚育苗, 三叶期移栽至大田。田间实验分别在华中农业大学实验场和湖北襄樊正大农业开发公司实验场进行。实验采用随机区组设计, 单行区, 行长 5 m, 行距 0.70 m, 株距 0.25 m, 种植密度为 45000 株/hm²。田间管理同大田管理。在玉米成熟后, 从每行的第 3 株开始连续收获 10 株果穗, 并在自然条件下晾晒至含水量为 13% 后, 进行性状考察。考察的性状包括: (1) 产量: 一行玉米 10 个果穗籽粒的平均重量, 并转换成 t/hm²; (2) 行数: 一行玉米 10 个果穗行数的平均数; (3) 行粒数: 一行玉米 10 个果穗行粒数的平均数; (4) 百粒重: 一行玉米 10 个果穗的玉米籽粒混合后随机挑选 3 个 100 粒称重的平均数(g)。最后的数据分析是基于 4 个性状两个点 6 个重复的平均数进行的。

() 基因型分析和分子标记连锁图构建. 参照 Saghai Maroof 等人^[27]的方法提取亲本和群体单株总 DNA。从已发表的玉米分子标记图谱中挑选能较均匀覆盖全基因组的 479 个分子标记, 其中 SSR 标记 375 个, RFLP 标记 104 个, 进行多态性检测。SSR 分析根据 Senior 等人^[28]的方法进行; RFLP 分析中的酶切、Southern 印迹转移、分子杂交参照 Gardiner 等人^[29]的方法。在此前构建的含 174 个分子标记的连锁图^[30]的基础上, 又新添加了 27 个 SSR 标记和 4 个 RFLP 标记。用 R/qtl 软件^[31]的 cacl.genoprob 和 cacl.errorlod 命令检测每个单株的标记基因型, 当 LOD ≥ 2, 就认

为该基因型数据可能有误^[32], 并用缺失代替, 然后基于排除了错误基因型的标记数据, 利用 MAPMAKER/EXP3.0^[32] 构建了覆盖玉米 10 条染色体的分子标记连锁图 (LOD > 3.0)。

() 数据分析.

(1) 表型数据分析. 利用 SAS 软件^[33]和 F_{2:3} 家系的表型数据计算产量及产量构成因子亲本和群体的平均值及表型变异。产量性状 F_{2:3} 家系的遗传力和置信区间遵照 Knapp 等人^[34]的方法计算。遗传力计算公式为: $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{gl}^2/n + \sigma_e^2/nr)$ 。这里, σ_g^2 是遗传变异, σ_{gl}^2 是基因型与环境互作的变异, σ_e^2 是误差, r 是重复数, n 是地点数。假定地点随机的情况下, 利用 ANOVA 方法计算 σ_g^2 , σ_{gl}^2 和 σ_e^2 的值。

(2) QTL 分析. 利用 R/qtl 软件^[31]中自带的基于最大似然法的 EM 算法, 也即标准的区间作图法^[35]进行全基因组的主效 QTL 分析。在 R 语言下, 利用 scanone 命令进行主效 QTL 分析, 在 $P = 0.05$ 水平, 通过 1000 次模拟确定每个性状是否存在主效 QTL 的 LOD 阈值, 同时进一步利用多区间作图法 (MIM) 的 BIC: $g(n) = \ln(n)$ 模型验证检测到的主效 QTL 的真实性及位置^[36]。利用 R/qtl 软件^[31]中的 Haley-Knott 回归方法进行上位性 QTL 分析。该方法是基于多位点基因型概率的表型回归, 对于缺失数据利用 Sen 和 Churchill^[37]提出的基于 Pseudomarker 运算法则的多重填补 (Multiple Imputation) 方法处理 (重复 500 次)。首先用 cacl.genoprob 命令计算多位点的基因型概率, 然后用 scantwo 命令全基因组扫描可能的上位性 QTL。在 $P = 0.05$ 水平, 通过 500 次模拟确定存在上位性 QTL 的 Joint LOD 阈值。对互作涉及的两位点如 LOD 值同时 ≥ 2.5, 则认为互作发生在两个主效 QTL 之间。最后用 Windows QTL Cartographer 2.5^[38] 中附带的 MIM 方法计算主效 QTL 和主效 QTL 互作的贡献率及总的贡献率。

2 结果与分析

2.1 表型数据分析

产量及产量构成因子双亲、F₁ 和 F_{2:3} 家系的表型变异如表 1 所示。产量和行粒数具有较高的杂种优势, 分别达到 245.17% 和 100.60%; 而行数和百粒重的杂

1) Yan J B, Tang H, Huang Y Q, et al. QTL mapping and epistatic analysis for yield and yield components using molecular markers with an elite maize hybrid. *Euphytica* (in press)

表 1 两亲本及F_{2:3}群体表型变异及遗传力^{a)}

项目	产量/t · hm ⁻²	行数	行粒数	百粒重/g
综 3	2.37 ± 0.34	13.03 ± 0.51	21.00 ± 1.73	24.17 ± 6.81
87-1	2.81 ± 0.35	13.83 ± 0.49	15.44 ± 1.08	32.76 ± 7.15
F ₁	8.94 ± 0.50	16.20 ± 0.39	36.55 ± 1.27	34.55 ± 6.19
F ₃ 变化范围	2.4~6.2	11.8~18.6	18.6~31.2	24.6~35.5
H (%)	245.17	20.63	100.60	21.38
σ_g^2	0.30 ± 0.21	1.25 ± 0.32	4.83 ± 0.95	2.96 ± 0.97
σ_{gl}^2	0.11 ± 0.29	0.05 ± 0.45	1.50 ± 1.34	1.08 ± 1.37
σ_e^2	0.26	0.6	5.41	5.63
遗传力(h ²)	0.75	0.91	0.75	0.67
置信区间(90%)	0.71~0.79	0.89~0.92	0.70~0.79	0.60~0.72

a) $H = (F_1 - MP) / MP$; $MP = (P_1 + P_2) / 2$. H和MP分别代表中亲优势和中亲值

种优势较低, 分别为 20.63%和 21.38%. 但 4 个性状在F_{2:3}群体中都存在丰富的表型变异. 广义遗传力以行数最高达到 0.91, 而百粒重最低只有 0.67, 产量和行粒数居中都是 0.75. 4 个性状在P < 0.01 水平都检测到显著的基因型与环境的互作(数据未列出). 产量与 3 个构成因子中的行粒数的相关性最高, 说明在玉米育种中穗长对玉米单株产量贡献较大. 全基因组标记的杂合性与性状相关分析表明, 在P < 0.01 水平, 杂合性与杂种优势较高的性状如产量和行粒数具有显著的相关性, 但相关系数都比较小在 0.3 以下, 但与杂种优势较低的性状如行数和百粒重没有显著的相关性(数据未显示).

2.2 分子标记连锁图的构建

此前, 基于同一群体, 我们构建了包含 174 个分子标记的分子标记连锁图^[30], 覆盖玉米 10 条染色体, 分为 12 个连锁群, 总长度为 2531.6 cM, 平均间距为 14.5 cM. 在此连锁图的基础上, 增添了 27 个SSR标记和 4 个RFLP标记, 这 205 个分子标记被分为 11 个连锁群, 总长度为 2395.5 cM, 平均间距为 11.7 cM. 基于R/qtl软件的基因型错误检测功能, 在LOD ≥ 2.0 的水平, 共有 123 个标记基因型数据有误, 涉及 462 个位点, 分布在所有 10 条染色体(图 1), 占有基因型数据的 0.936%. 最少的是第 9 染色体, 涉及 7 个标记 13 个位点; 最多的是第 8 染色体, 涉及 16 个标记 100 个位点(表 2). 大部分基因型有误标记涉及到的位点都在 3 以下, 只有 6 个标记涉及到的位点超过 10(数据未给出). 这些基因型可能有误的 462 个位点全部当缺失数据处理后重新构建了分子标记连锁图, 发现分子标记连锁图标记的位置几乎没有什么变化,

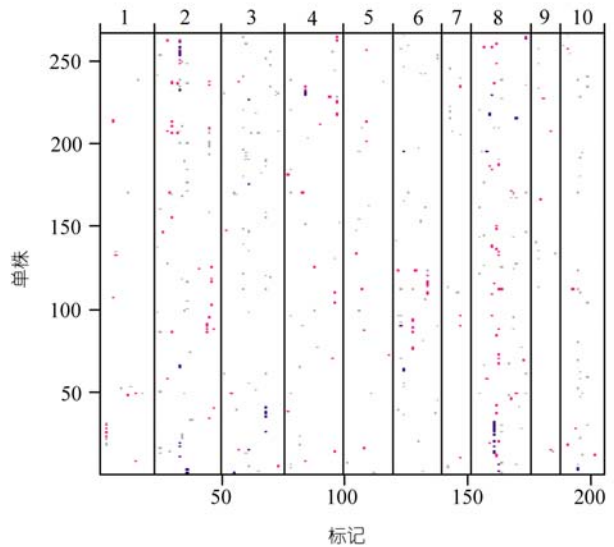


图 1 全基因组扫描基因型数据检测到可能有误的标记及涉及到的位点

横轴是 1~205 个标记, 纵轴是 266 个单株; 灰色的点表示 2 < LOD < 3, 粉红色的点表示 3 < LOD < 4.5, 紫色的点表示 LOD > 4.5

但长度与原始图谱相比缩短为 2305.4 cM, 平均间距由 11.7 cM 变为 11.2 cM, 影响最大的是第二染色体, 距离缩短近 27 cM(图 2). 用去掉基因型错误的数据进行下一步的 QTL 分析.

2.3 QTL 分析

() 主效 QTL 分析. 利用 R/qtl 软件中自带的基于最大似然法的 EM 算法进行主效 QTL 分析, 1000 次模拟之后, 在 P = 0.05 水平, LOD 阈值最大的为产量 3.1, 最小为行粒数的 2.9, 最后统一取 3.1 为确定所有 4 个性状是否存在一个 QTL 的临界值(对个别 LOD 值在

表2 涉及标记基因型错误的染色体和位点及检测前后各染色体遗传距离的比较

染色体	1	2	3	4	5	6	7	8	9	10	总计	平均
涉及标记/个	9	18	18	14	11	14	5	16	7	10	122	12.2
涉及位点/个	23	97	67	43	18	51	16	100	13	34	462	46.2
检测前遗传距离/cM	302.1	312.2	284.1	268.4	231.6	283.5	176.6	193.5	120	223.5	2395.5	11.7
检测后遗传距离/cM	297.7	285.5	271.9	260.1	228.1	273.3	172.1	184.6	112.9	219.2	2305.4	11.2

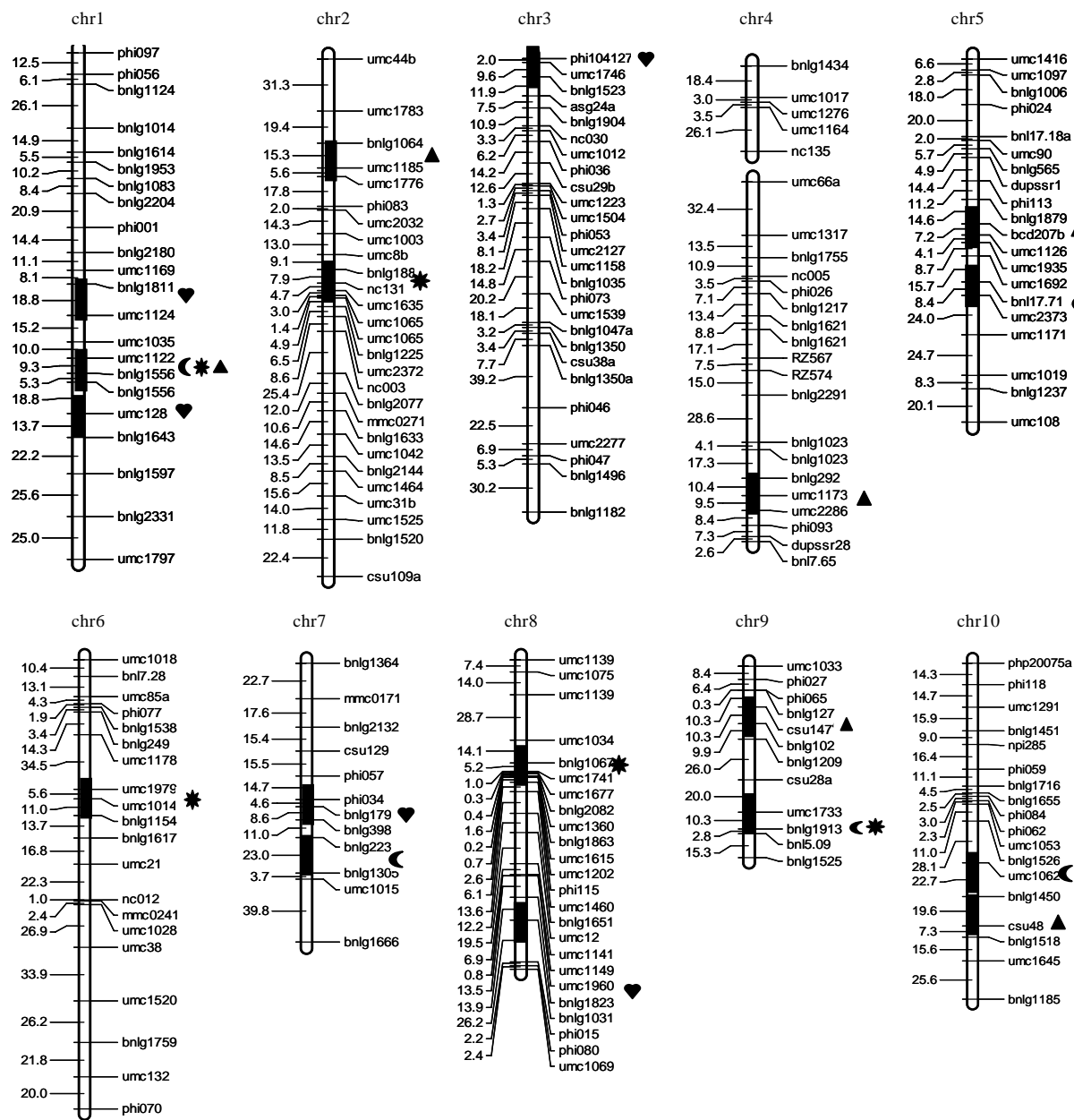


图2 影响产量及构成因子的单位点 QTL 在染色体上的位置

○, 产量; ★, 行粒数; ▲, 行数; ♥, 百粒重

3.1 之下, 但 $P < 0.01$ 的 QTL 也列出来). 利用检测到的 QTL 为基础模型, 在 $P = 0.05$ 水平下, 进一步利用 MIM 方法验证 R/qrtl 的分析结果. 如表 3 所示, 基于 R/qrtl 方法, 检测到的产量、行粒数、行数的 QTL 分别为 5, 5, 7 和 5 个主效 QTL, 其中行数和粒数两个主效 QTL 之间存在显著的互作. 通过 MIM 方法这 4 个性状分别有 4, 4, 6 和 4 个主效 QTL 得到验证, 有 4 个 QTL 被认为可能是假阳性 QTL, 分别为 *gy6*, *kpr2a*, *rn8* 和 *kw4*. 同时这 4 个性状还各发现一个新的 QTL, 分别为 *gy10*, *kpr2a*, *rn2* 和 *kw7b*.

以下仅分析 MIM 方法得到的结果. 产量一共检测到 5 个主效 QTL, 分别位于第 1, 5, 7, 9 和 10 染色体上, *gy1* 和 *gy5* 由来自 87-1 的基因型起增效作用, 而其他 3 个 QTL 由来自综 3 的等位基因起增效作用, 单

个 QTL 解释的遗传变异从 3.0% 到 9.2% 不等, 这 5 个 QTL 共能解释 35.3% 的遗传变异. 行粒数一共检测到 5 个 QTL, 分别位于第 1, 2, 6, 8 和 9 条染色体, 除 *kpr2b* 外, 其他 4 个 QTL 都由来自综 3 的等位基因起增效作用. 同时还发现 QTL *kpr1* 和 *kpr8* 之间存在显著的互作, 能解释 3.4% 的遗传变异, 这 6 个 QTL (含一个上位性 QTL) 共能解释 37.4% 的遗传变异. 一共检测到 7 个影响行数的 QTL, 分布于除第 3, 7 和 8 染色体之外的其他 7 条染色体上, 其中 *rn2*, *rn9* 和 *rn10* 由来自综 3 的等位基因起增效作用, 其他 4 个 QTL 由来自 87-1 的等位基因起增效作用, 这 7 个 QTL 共解释 61.5% 的遗传变异. 影响百粒重的 QTL 有 5 个, 分布于第 1 (2 个 QTL), 3 和 7 (2 个 QTL) 条染色体上, *kw1b* 和 *kw3* 由来自 87-1 的等位基因起增效作用, 其

表 3 影响产量及构成因子的单位点 QTL

性状	QTL ^{a)}	位置/cM ^{b)}	LOD ^{c)}	A	D	贡献率(%)
产量	<i>gy1</i>	192(190)	3.9(5.7)	-0.15	0.34	9.2
	<i>gy5</i>	154(152)	2.3(3.6)	-0.16	0.17	5.0
	<i>gy6^{d)}</i>	(235)	(3.3)			
	<i>gy7</i>	133(132)	3.8(9.5)	0.25	0.08	8.8
	<i>gy9</i>	104(98)	2.1(4.6)	0.07	0.25	5.3
	<i>gy10^{e)}</i>	122	1.9	0.04	0.32	7.0
						35.3
行粒数	<i>kpr1</i>	199(198)	3.5(4.8)	0.70	1.20	12.9
	<i>kpr2a^{d)}</i>	(78)	(2.4)			
	<i>kpr2b^{e)}</i>	149	2.0	-0.47	0.88	3.0
	<i>kpr6</i>	86(84)	2.0(3.9)	0.66	0.53	5.7
	<i>kpr8</i>	66(64)	2.9(6.0)	0.50	1.22	7.9
	<i>kpr9</i>	104(98)	1.6(3.6)	0.08	0.98	4.5
	<i>kpr1:kpr8</i>	198:64	1.8(2.5)	0.76 ^{f)}	-1.57 ^{g)}	3.4
						37.4
行数	<i>rn1</i>	195(194)	4.7(4.5)	-0.52	0.07	9.4
	<i>rn2^{e)}</i>	58	4	0.48	-0.20	5.8
	<i>rn4</i>	242(240)	3.0(3.7)	-0.47	-0.12	8.3
	<i>rn5</i>	120(112)	5.5(12.2)	-0.58	0.00	9.8
	<i>rn6</i>	103(95)	4.9(4.3)	-0.45	-0.30	9.0
	<i>rn8^{d)}</i>	(152)	(6.2)			
	<i>rn9</i>	19(18)	6.3(10.1)	0.66	-0.10	13.2
	<i>rn10</i>	148(162)	2.8(2.7)	0.39	-0.10	6.0
						61.5
百粒重	<i>kw1a</i>	151(150)	1.8(3.4)	0.04	1.30	9.6
	<i>kw1b</i>	218(212)	1.9(5.7)	-0.80	0.06	5.9
	<i>kw3</i>	2(1)	1.8(3.1)	-0.54	0.43	4.4
	<i>Kw4^{d)}</i>	(260)	(3.5)			
	<i>kw7a</i>	86(86)	2.8(3.3)	0.78	-0.22	7.2
	<i>kw7b^{e)}</i>	166	2.1	0.71	0.97	12.6

a) QTL 后面的数字表示 QTL 所在的染色体, 如果同一条染色体有多个 QTL 则在数字后面用字母 a, b... 表示. b) QTL 扫描峰值所在的染色体位置, 括弧外边是 MIM 验证结果, 括弧里面是 R/qrtl 扫描结果. c) 括弧里面是 R/qrtl 计算得到的 LOD 值, 括弧外面是 MIM 验证和扫描得到的 LOD 值. d) R/qrtl 扫描找到, 但 MIM 没能验证的 QTL. e) R/qrtl 扫描没有找到, 但 MIM 重新检测到的 QTL. f) AA 互作的表型值. g) DD 互作的表型值

他3个QTL由来自综3的等位基因起增效作用,这5个QTL共解释39.7%的遗传变异. 综上,共检测到22个影响产量等4个性状的主效QTL,同时发现影响行粒数的两个主效QTL之间存在显著的互作,单个QTL解释的遗传变异在2.9%到12.9%之间,主效QTL(包括主效QTL之间的互作)能解释总的遗传变异从产量的35.3%到行数的61.5%不等(表3). 从表3和图2还可以看出,有些影响不同性状的QTL落在同一染色体区间,如影响产量、行粒数和行数的3个QTL *gyl*, *kpr1* 和 *rn1* 都位于第一染色体的192~199 cM之间. 又如,影响产量和行粒数的QTL *gyl9*和 *kpr9* 的LOD扫描峰值都位于104 cM. 这些QTL有可能是一因多效也可能是不同QTL紧密连锁,但这些QTL为下一步的QTL分子标记辅助选择或者精细定位和克隆提供了有益信息.

() 上位性互作QTL分析. 利用R/qtl软件进行全基因组上位性QTL检测,通过500次模拟,在 $P = 0.05$ 水平,4个性状Joint LOD阈值在6.9~7.1之间变化,最后以Joint LOD ≥ 7.1 和上位性互作 LOD \geq

3.0为临界值确定是否存在上位性QTL. 按照互作QTL涉及的2个基因座是否存在显著效应,把互作QTL分为3种类型:互作两基因座都存在显著效应,也即发生在两主效QTL之间的互作,称为QQ类型;互作两基因座一个存在显著效应另一个不存在,也即发生在主效QTL与无显著效应基因座之间的互作,称为QN类型;互作两基因座都不存在显著效应,也即发生在两无显著效应基因座之间的互作,称为NN类型. 产量共检测到6个互作QTL,涉及到的基因座分布于除第3和8条染色体之外的其他8条染色体,其中4个属于QN类型,2个属于NN类型,没有发现QQ类型的互作. 行粒数检测到7对互作QTL,涉及到6条染色体上的13个基因座位,7个互作QTL中包括1个QQ类型,4个QN类型和2个NN类型. 检测到5个影响行数的互作分布在第3,4,6,7,8和10条染色体上,包括4个QN类型和1个NN类型. 5个影响百粒重的互作QTL分布在6条染色体的8个不同基因座位上,其中有3个属于QN类型的互作,其他2个是NN类型的互作(表4). 图3在两维平面上

表4 影响产量及构成因子的上位性互作QTL

性状	染色体 ^{a)}	QTL _i ^{b)}	染色体 ^{c)}	QTL _j ^{d)}	LOD _{jnt} ^{e)}	LOD _{int} ^{f)}	LOD _{qi} ^{g)}	LOD _{qj} ^{h)}
产量	1	140	1	200	8.3	3.2		3.6
	1	188	2	128	9.7	3.8	5.6	
	4	176	7	100	7.7	6.3		
	5	4	6	240	7.5	3.1		2.7
	7	96	7	112	7.3	5.1		
	7	124	10	16	8.8	4	4.5	
	9	60	10	116	7.6	4		
行粒数	1	200	2	132	12.2	3.1	7.4	
	1	172	7	112	11.3	4.8	6.2	
	1	200	8	64	13.2	2.8	7.4	3.9
	2	16	8	72	9.6	5.4		3.0
	2	112	10	120	7.3	3.3	2.5	
	4	80	5	136	7.1	5.1		
	8	40	8	80	7.3	4.1		
行数	3	208	8	152	7.3	3.3		3.8
	4	228	7	0	7.4	3.1	4.2	
	6	148	10	160	8.1	4.3		
	3	256	4	220	8.3	3.3		3.9
	4	240	3	96	11.5	3.3	4.7	
百粒重	1	96	10	96	7.1	5		
	3	0	4	20	8.0	5.3		
	4	260	5	212	9.0	5.2	2.7	
	4	20	7	148	9.8	3.5		6.0
	7	148	10	36	10.6	3.4	6.0	

a)和c)互作两QTL所在染色体. b)和d)互作两QTL所在染色体的具体染色体位置(cM). e)联合LOD值. f)互作QTL LOD值. g)和h)在上位性模型下两互作QTL基因座分别进行主效QTL检测的条件LOD值

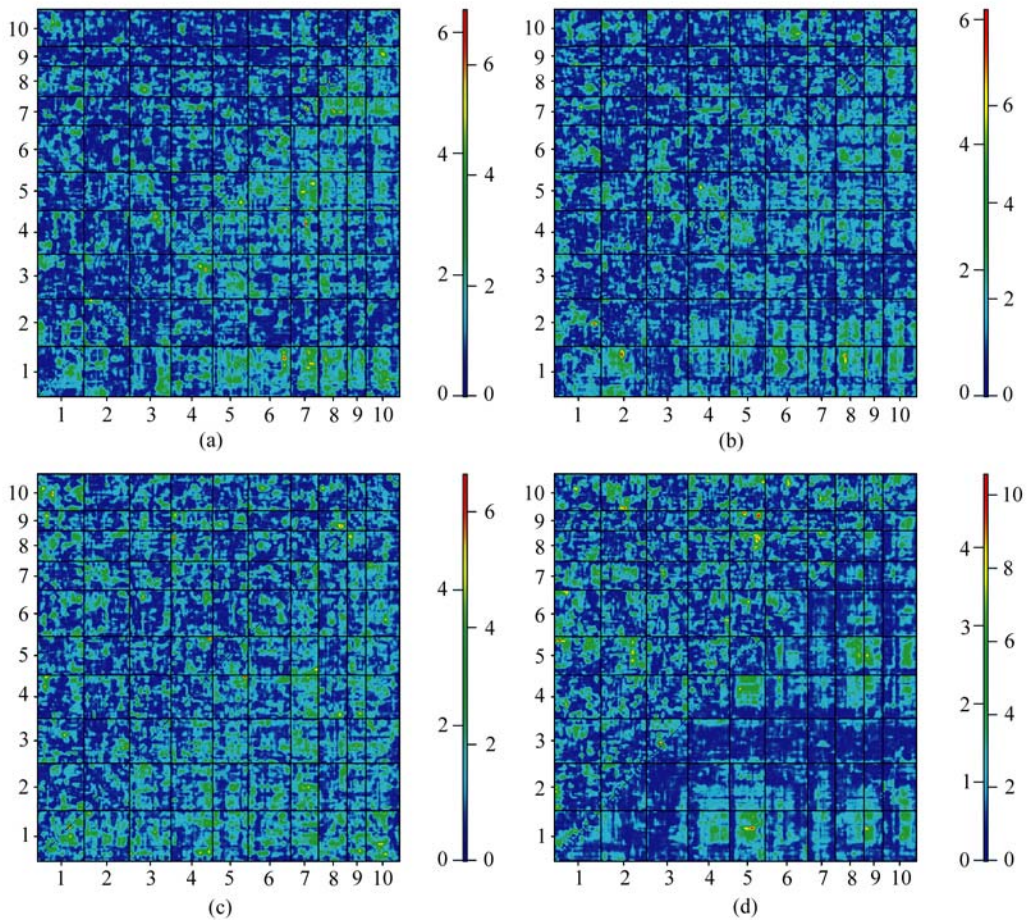


图 3 产量及构成因子上位性互作 QTL 两维平面扫描结果

(a) 产量; (b) 行粒数; (c) 行数; (d) 百粒重. 横轴和纵轴均为按顺序排列的 10 条染色体上的标记, 从蓝色到红色表明联合 LOD 值逐渐增大

直观显示了各个性状在全基因组水平上互作的扫描情况, 不同的颜色代表不同的 LOD 值. 从图 3 可以看出, 影响不同性状的互作在全基因组水平上有不同也有相同. 相同的是, 影响各个性状的互作都是大量存在并涉及到几乎所有染色体; 不同的是, 影响不同性状的互作还存在不同的热点分布区域. 从图 4(a)可以看出, 除了第 2 条染色体外其他染色体之间都有大量的互作影响产量的性状表现, 尤其第 1 染色体几乎与所有其他 9 条染色体有显著性的互作. 而影响行数的互作 QTL(图 4(c))几乎是平均分布于所有 10 条染色体. 而影响百粒重的互作 QTL 却是有几个明显的热点区域, 如在第 1, 4, 5 三条染色体及 5, 8, 9 三条染色体之间. 所有在上位性 QTL 中有显著效应的基因座(主效 QTL)大部分(11/15)与区间作图方法检测到的主效 QTL 一致, 而其中没有能检测到的部分可能与在区间作图分析中采取较严格的标准有关.

3 讨论

3.1 分子标记连锁图构建的准确性

从 1986 年第一张包含 116 个位点的玉米分子标记连锁图诞生^[39]到 2004 年包含 5863 个位点的玉米高密度整合图谱公布(<http://www.maizegdb.org/map.php#rep>), 过去不到 20 年的时间里, 分子标记技术发展极为迅速, 几乎所有重要作物都构建了高密度的分子标记连锁图谱, 并利用这些图谱开展了一些重要的应用和基础研究. 但分子标记操作过程包括单株 DNA 提取, DNA 条带的读取, 标记数据的获得大都由手工完成, 对包含数百个单株和标记多达数万条带信息量的数据而言, 其中出现少量错误在所难免, 但在实际的连锁图构建中却很少考虑这一因素, 即使考虑也很难有效排除. 最近, 有人对 IBM 图谱中的 77 个偏分离标记重新进行了检测, 发现有 1.86% 的标记位点可能存在错误^[40], 基于这个错误概率, 通

过 1000 次模拟, 他们估计图谱(IBM_Gnp2004)的遗传距离比实际值可能有平均 15.3%的扩张. 本研究利用 R/qtl 软件提供的检测基因型数据是否存在错误的功能, 对构建图谱的基因型数据进行了检测, 在 $LOD \geq 2.0$ 水平, 发现有 123 个标记共涉及 462 个位点的基因型数据可能有误, 占有所有基因型数据的大约 1%. 在剔除这些数据后, 各条染色体的遗传距离都有一定程度的减少, 整个连锁图的长度由 2395.5 cM 缩短为 2305.4 cM. 如果把显著性水平放宽到 0.05, 检测到的错误基因型数据将会更多. 在无法对整个基因型数据进行重新确认的情况下, 通过统计的方法排除其中一些明显的错误数据, 无疑对提高分子标记连锁图的精度和 QTL 定位的准确性有帮助.

3.2 QTL 定位的准确性

对大多数遗传研究而言, 由两个自交系组配而来的 F_2 群体理论上可以提供最多和最丰富的遗传信息^[41]. 但用 F_2 群体很难进行复杂数量性状分析, 因为每一个基因型只有一个单株, 无法进行多重实验. 利用 F_3 家系的平均值作为 F_2 个体基因型的估计值, 以减少实验误差, 提高实验精度, 部分解决了这个问题, 类似研究在水稻^[15]等作物中也被广泛采用. 但同时, 利用 F_3 家系代替 F_2 单株, 通过一代自交其显性效应被低估; 另外当 F_2 个体基因型是杂合的时候, 其 F_3 家系剩余误差(residual error)的分布将变成一个多种分布的混合体, 在目前的 QTL 定位中我们并没有考虑这种情况, 因此估计可能是有偏的^[42]. 在下一步研究中, 我们将利用“永久 F_2 ”群体来解决这些问题. “永久 F_2 ”群体这一设想首先在水稻^[43]中提出, 该群体不但具有 F_2 群体一切的优点同时还可以进行多年多点的重复实验. 最近, 在老鼠中也开始利用这一设计来研究复杂数量性状, 结果表明, 该群体能显著提高 QTL 检测的能力, 以及复杂上位性互作及基因型与环境互作检测的能力, 以及复杂上位性互作和基因型与环境互作检测能力^[44].

3.3 上位性对玉米产量性状遗传也起重要作用

上位性在复杂数量性状的遗传中起着重要的作用, 这几乎逐渐成为共识, 但正如 Carlborg 和 Haley^[45]提到的一样, 因为技术发展的限制和相关统计分析方法的缺乏, 上位性在复杂数量性状遗传中的重要性常常被忽视. 本研究中, 我们基于一个可以

同时进行主效 QTL 和上位性 QTL 联合分析的软件 R/qtl 对产量及其 3 个构成因子在全基因组水平进行了分析, 同时利用 MIM 方法对主效 QTL 和主效 QTL 之间的互作进行了验证. 结果表明, 4 个性状单位点 QTL 能够解释的遗传变异分别为: 产量 35.3%, 行粒数 37.4%, 行数 61.5% 和百粒重 39.7% (表 3); 而这 4 个性状对应的广义遗传力则分别为产量 75%, 行粒数 75%, 行数 91% 和百粒重 67% (表 1). 显然主效 QTL (包括主效 QTL 之间的互作) 解释的遗传变异不能完全解释这 4 个数量性状全部的遗传变异. 与此同时, 检测到数量与主效 QTL 相当的上位性互作 QTL, 其中行数和百粒重的互作 QTL 分别为 5 个, 产量和行粒数的互作 QTL 分别为 7 个 (表 4). 因为方法的限制, 暂时还无法准确估计这些互作 QTL 所能解释的遗传变异. 但从互作的类型来看, 主效 QTL 之间的互作很少, 只有行粒数检测到一个, 没有显著效应基因座位之间的互作占三分之一 (8/24); 大部分互作属于主效 QTL 与没有显著效应的基因座位之间的互作 (15/24) (表 3), 这也说明主效 QTL 除了单独起作用之外还可以通过没有显著效应的基因座位之间的互作来影响性状的表达. 这一结果与我们此前的研究结果一致¹⁾ 及大量其他的研究结果相吻合^[15~18]. 通常情况下, 我们是选择一个 LOD 阈值来判定两基因座位之间是否存在互作, 这个阈值如果偏高则犯 Type I 类型错误的概率就增大, 阈值偏低则犯 Type II 类型错误的概率就增大, 如何合理的确定阈值现在仍然是一个挑战. 本研究通过二维扫描图形直观地把所有可能概率水平的互作 QTL 的结果都显示出来 (图 3), 从图中可以看出尽管不同性状之间的互作涉及的基因座位和方式可能不尽相同, 但有一个共同的特点就是, 显著而丰富的互作在所有 4 个性状中都是广泛存在的. 这也暗示上位性 QTL 在玉米产量性状的遗传中可能有着与主效 QTL 一样重要的作用. 如何更有效揭示和利用这些上位性 QTL 则是我们下一步要深入考虑的问题.

致谢 UMC 玉米 RFLP 实验室 Ed Coe 博士馈赠 RFLP 探针. 本工作为国家自然科学基金 (批准号: 39893350) 和国家重点基础研究发展计划 (批准号: 2001CB1088) 资助项目.

参 考 文 献

- 1) Duvick D N. Heterosis: Feeding people and protecting natural

1) 同 1414 页脚注

- resources. In: Coors J G, Pandey S, eds. *The Genetics And Exploitation Of Heterosis In Crops*. American Society of Agronomy, Inc, Crop Science Society of America, Inc, Soil Science Society of America, Inc, Madison, WI. 1999. 19—29
- 2 Ribaut J M, Hoisington D. Marker-assisted selection: new tools and strategies. *Trends Plant Sci*, 1998, 3(6): 236—239[DOI]
 - 3 Frary A, Nesbitt T C, Grandillo S, et al. *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science*, 2000, 289: 85—88[DOI]
 - 4 Yano M, Katayose Y, Ashikari M, et al. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the arabidopsis flowering time gene *CONSTANS*. *Plant Cell*, 12: 2473—2484
 - 5 Takahashi Y, Shomura A, Sasaki T, et al. *Hd6*, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. *Proc Natl Acad Sci USA*, 2001, 98: 7922—7927[DOI]
 - 6 Ashikari M, Sakakibara H, Lin S Y, et al. Cytokinin oxidase regulates rice grain production. *Science*, 2005, 309: 741—745[DOI]
 - 7 Ren Z H, Gao J P, Li L G, et al. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nature Genet*, 2005, 37: 1141—1146[DOI]
 - 8 Wang R L, Stec A, Hey J, et al. The limits of selection during maize domestication. *Nature*, 1999, 398: 236—239[DOI]
 - 9 Wright S. Genic and organismic selection. *Evolution*, 1980, 34: 825—843[DOI]
 - 10 Brem R B, Storey J D, Whittle J, et al. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature*, 2005, 436: 701—703[DOI]
 - 11 Allard R W. Genetic basis of the evolution of adaptedness in plants. *Euphytica*, 1996, 92: 1—11[DOI]
 - 12 Rieseberg L H, Sinervo B, Linder C R, et al. Role of gene interactions in hybrid speciation: evidence from ancient and experimental hybrids. *Science*, 1996, 272: 741—745
 - 13 Fatokun C A, Menancio-Hautea D I, Danesh D, et al. Young evidence for orthologous seed weight genes in cowpea and mung bean based on RFLP mapping. *Genetics*, 1992, 132: 841—846
 - 14 Maughan P J, Saghai Maroof M A, Buss G R. Molecular-marker analysis of seed-weight: Genomic locations, gene action, and evidence for orthologous evolution among three legume species. *Theor Appl Genet*, 1996, 93: 574—579
 - 15 Yu S B, Li J X, Xu C G, et al. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA*, 1997, 94: 9226—9231[DOI]
 - 16 Li Z K, Luo L J, Mei H W, et al. Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics*, 2001, 158: 1737—1753
 - 17 Li Z K, Pinson S R, Park W D, et al. Epistasis for three grain yield components in rice. *Genetics*, 1997, 145: 453—465
 - 18 Luo L J, Li Z K, Mei H W, et al. Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics*, 2001, 158: 1755—1771
 - 19 Lamkey K R, Schnicker B J, Melchinger A E. Epistasis in an elite maize hybrid and choice of generation for inbred line development. *Crop Sci*, 1995, 35: 1272—1281
 - 20 Gardiner J M, Coe E H, Melia-Hancock S, et al. Development of a core RFLP map in maize using an immortalized F₂ population. *Genetics*, 1993, 134(7): 917—917
 - 21 Eta-Ndu J T, Openshaw S J. Epistasis for grain yield in two F₂ populations of maize. *Crop Sci*, 1999, 39: 346—352
 - 22 Doebley J, Stec A, Gustus C. Teosinte branched-1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*, 1995, 141: 333—346
 - 23 Lukens L N, Doebley J. Epistatic and environmental interactions for quantitative trait loci involved in maize evolution. *Genet Res*, 1999, 74: 291—302[DOI]
 - 24 Barriere Y, Gibelin C, Argillier O, et al. Genetic analysis in recombinant inbred lines of early dent forage maize. QTL mapping for yield, earliness, starch and crude protein contents from per se value and top cross experiments. *Maydica*, 2001, 46: 253—266
 - 25 McMullen M D, Snook M, Lee E A, et al. The biological basis of epistasis between quantitative trait loci for flavone and 3-deoxyanthocyanin synthesis in maize (*Zea mays* L.). *Genome*, 2001, 44: 667—676[DOI]
 - 26 Moreau L, Charcosset A, Gallais A. Experimental evaluation of several cycles of marker-assisted selection in maize. *Euphytica*, 2004, 137: 111—118[DOI]
 - 27 Saghai Maroof M A, Soliman K M, Jorgensen R A, et al. Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population, and population dynamics. *Proc Natl Acad Sci USA*, 1984, 81: 8014—8018[DOI]
 - 28 Senior M L, Heun M, Manfred H. Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. *Genome*, 1993, 36: 884—889
 - 29 Gardiner J M, Coe E H, Melia-Hancock S, et al. Development of a core RFLP map in maize using an immortalized F₂ population. *Genetics*, 1993, 134(7): 917—917
 - 30 严建兵, 汤华, 黄益勤, 等. 不同发育时期玉米株高 QTL 的动态分析. *科学通报*, 2003, 48(18): 1959—1964
 - 31 Broman K W, Wu H, Sen S, et al. R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, 2003, 19: 889—890
 - 32 Lincoln S, Daly M, Lander E. *Mapping Genetic Mapping With MAPMAKER/EXP3.0*. Cambridge: Whitehead institute Technical Report, 1992
 - 33 SAS Institute, Inc. *SAS Users Guide: Statistic*. Cary, NC: SAS Institute, 1996
 - 34 Knapp S J, Stroup W W, Ross W M. Exact confidence intervals for heritability on a progeny mean basis. *Crop Sci*, 1985, 25: 192—194
 - 35 Lander E S, Botstein S. Mapping Mendelian factors underlying quantitative traits using TFLP linkage maps. *Genetics*, 1989, 121: 185—199
 - 36 Kao C H, Zeng Z B, Teasdale R D. Multiple interval mapping for quantitative trait loci. *Genetics*, 1999, 152: 1203—1216
 - 37 Sen S, Churchill G A. A statistical framework for quantitative trait mapping. *Genetics*, 2001, 159: 371—387
 - 38 Wang S C, Basten C J, Zeng Z B. *Windows QTL cartographer 2.5 user manual*. Dissertation for the Doctoral Degrel. North Carolina state: North Carolina State University. 2005
 - 39 Helentjaris T, Slocum M, Wright S, et al. Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. *Theor Appl Genet*, 1986, 72: 761—769
 - 40 Falque M, Decousset L, Dervins D, et al. Linkage mapping of 1454 new maize candidate gene loci. *Genetics*, 2005, 170: 1957—1966[DOI]
 - 41 Allard R W. Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia*, 1956, 24: 235—278
 - 42 Zhang Y M, Xu S Z. Mapping quantitative trait loci in f₂ incorporating phenotypes of F₃ progeny. *Genetics*, 2004, 166: 1981—1993[DOI]
 - 43 Hua J P, Xing Y Z, Xu C G, et al. Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. *Genetics*, 2002, 162: 1885—1895
 - 44 Zou F, Gelfond A L, Airey D C, et al. Quantitative trait locus analysis using recombinant inbred intercrosses: Theoretical and empirical considerations. *Genetics*, 2005, 170: 1299—1311[DOI]
 - 45 Carlborg Ö, Haley C S. Epistasis: Too often neglected in complex trait studies? *Nature Rev Genet*, 2004, 5: 618—625[DOI]

(2005-11-21 收稿, 2006-03-24 接受)