

植物育种模拟的原理和应用

王建康¹, Wolfgang H. Pfeiffer²

¹中国农业科学院作物科学研究所, 国家农作物基因资源与基因改良重大科学工程, CIMMYT 中国办事处, 北京 100081;

²国际热带农业研究中心, 哥伦比亚)

摘要: 传统植物育种很大程度上依赖于表型选择和育种家的经验, 分子生物学积累的大量遗传数据使得在基因水平上进行目标性状的选择成为可能, 但是由于缺乏必要的工具, 大量的遗传信息尚未在育种中得以有效利用。模拟方法可以利用各类遗传信息, 在育种家进行田间试验之前, 对杂交组合的表现、后代选择效果以及整个育种过程进行模拟, 提出最佳的亲本选配、杂交和后代选择策略, 从而提高常规育种的预见性和效率。本文介绍育种模拟的一些基本原理, 包括遗传模型的构建、杂交类型和选择方法的定义等, 概述育种模拟工具 QuLine 的基本功能及其在育种方法的比较、利用已知基因信息的亲本选配和设计育种等方面的应用。模拟方法通过定义复杂的遗传模型和育种方法实现对育种过程的模拟, 在更真实的情形下比较不同育种方法的优劣、预测不同亲本杂交后代的表现, 从而为育种家有效利用大量的分子信息和遗传数据提供了手段。

关键词: 育种模拟; 遗传模型; 育种策略; 设计育种; 综述

Simulation Approach and Its Applications in Plant Breeding

WANG Jian-kang¹, Wolfgang H. Pfeiffer²

¹*Institute of Crop Science, The National Key Facility for Crop Gene Resources and Genetic Improvement, and CIMMYT China Office, Chinese Academy of Agricultural Sciences, Beijing 100081;* ²*HarvestPlus, c/o The International Center for Tropical Agriculture (CIAT), A.A. 6713, Cali, Colombia)*

Abstract: Conventional plant breeding largely depends on phenotypic selection and breeding experience, therefore breeding efficiency is low and predictions are inaccurate. Along with the fast development in molecular biology and biotechnology, a large amount of biological data is available for genetic studies of important breeding traits in plants, which in turn allows conducting genotypic selection in the breeding process. However, gene information has not been effectively used in crop improvement due to the lack of appropriate tools. The simulation approach can use vast and diverse genetic information, predict cross performance, and compare different selection methods. Hence, the best performing crosses and effective breeding strategies can be identified. QuLine is a computer tool capable of defining a range from simple to complex genetic models and simulating breeding processes for developing final advanced lines. Based on the results from simulation experiments, breeders can optimize their breeding methodology and greatly improve the breeding efficiency. In this paper, we first introduce the underlying principles of simulation modeling in crop enhancement, and then summarize several applications of QuLine in comparing different selection strategies, precision parental selection using known gene information, and the design approach in breeding. Breeding simulation allows the definition of complicated genetic models consisting of multiple alleles, pleiotropy, epistasis and gene by environment interaction, and provides a useful tool to efficiently use the wide spectrum of genetic data and information available to breeders.

Key words: Breeding simulation; Genetic model; Breeding strategy; Design breeding; Review

收稿日期: 2006-05-15; 接受日期: 2006-08-16

基金项目: 国际 HarvestPlus 挑战项目 Funded by the HarvestPlus Biofortification Challenge Program (www.harvestplus.org) of the Consultative Group on International Agricultural Research (CGIAR). HarvestPlus operates under the auspices of the International Center for Tropical Agriculture (CIAT) and the International Food Policy Research Institute (IFPRI)

作者简介: 王建康 (1965-), 男, 河南内乡人, 研究员, 研究方向为数量遗传、育种模拟、设计育种和 QTL 作图。Tel: 010-62133890; E-mail: wangjk@caas.net.cn or jkwang@cgiar.org

0 引言

生物个体的表型是基因型和环境共同作用的结果,植物育种的主要任务是寻找控制目标性状的基因,研究这些基因在不同目标环境下的表达形式,聚合存在于不同亲本材料中的有利基因,从而为农业生产提供适宜的品种^[1,2]。一个常规育种项目一般每年要配置数百甚至上千杂交组合,然而最终只有 1%~2% 的组合可以选育出符合育种目标的品种,大量组合在不同育种世代的选择过程中被淘汰,育种在很大程度上仍然依赖表型选择和育种家的经验,传统育种中的周期长、效率低和预见性差等问题未得到根本解决。分子生物学和生物技术 in 植物育种中的广泛应用,使得对育种性状的认识和研究手段发生了根本的变化,目前已实现对数量性状基因(即 QTL, quantitative trait locus)在染色体上的定位和作图^[3],对 QTL 的主效应、QTL 之间的互作效应、QTL 与遗传背景和环境间的互作效应等已有大量研究^[4-7],在此基础上,对控制数量性状的基因作单基因分解、精细定位^[8],有的甚至已达到图位克隆^[9]。但是,如果没有适当的工具和手段,育种家将难以把这些遗传信息有效地应用于育种实践。依据传统数量遗传学的原理和方法,如遗传力和配合力的估计等,可以对一些简单的选择方法在简化的遗传模型下进行初步的分析和预测^[10-14],但是这些分析方法建立在许多假设的基础上,如无基因型和环境互作、无上位性效应、无一因多效等,而实际情况往往不符合这些假设^[15,16]。模拟方法可以建立较真实的遗传模型,在育种家的田间试验之前,对育种程序中的各种因素进行模拟筛选和优化,提出最佳的亲本选配和后代选择策略,从而提高育种过程中的预见性。关于计算机模拟在育种上的应用,Frish 等比较了不同标记辅助选择的效果^[17,18],Podlich 等在考虑基因与环境互作的情况下研究不同轮回选择方法和多环境试验^[19,20],但这些研究多针对育种过程中的某一个侧面,缺乏对整个育种过程的模拟研究。建立在 QU-GENE^[19] 基础上的遗传育种模拟工具 QuLine (又称 QuCim) 则实现了对植物育种整个过程的模拟^[21-23]。本文首先介绍育种模拟的一些基本原理,然后简述 QuLine 的建立和一些育种应用。

1 植物育种模拟方法的原理

要模拟一个育种程序,首先要把育种过程用计算机能够理解的语言描述出来。模拟方法利用经典遗传

学、数量遗传学和群体遗传学的基本原理^[1,13,14],结合各种遗传研究结果定义育种性状的遗传模型,一个遗传模型所包含的基本内容有:控制性状的基因有多少?它们在染色体上的什么地方?每个基因位点上的等位基因数有多少?同一位点上等位基因间的作用方式以及不同位点上基因间的作用方式等内容。模拟方法利用育种学的基本原理定义育种方法,如系谱选择、混合选择、一粒传、标记辅助选择、群体内轮回选择、以及各种修饰选择方法等^[2,21,22]。模拟过程中,计算机根据孟德尔遗传规律配置杂交组合并产生育种后代材料、根据所定义的遗传模型产生育种后代的表现型、根据育种策略中的选择方法对后代材料进行选择,从而实现复杂育种过程的模拟。

1.1 育种策略的定义

广义地说,育种策略 (breeding strategy) 包含一个完整的育种周期 (breeding cycle) 中有关杂交、种子繁殖方式、田间试验设计、选择信息等各方面的内容,一个育种周期从配置杂交组合开始,到产生的新自交系用作下一轮杂交亲本为止。这里笔者用 CIMMYT (国际玉米小麦改良中心) 面包小麦育种中的选择混合方法 (selected bulk selection method) 为例说明育种策略所包含的内容^[21,22,24],选择混合方法的流程如图 1 所示。

1.1.1 选择混合方法中所包含的世代数以及每个世代的选择次数 在图 1 中,经过 F_1 到 F_{10} 这些世代的选择,剩余的家系将在国际育种试验中分发和测试,也将成为下一个育种周期的杂交亲本,因此一个育种周期中包含 10 个世代。每个世代所选择的性状和选择强度是不同的,可以利用各世代所包括的选择次数、种子来源、田间种植方式和选择细节等 (表 1 和表 2) 来定义一个育种世代。一些世代只经历一次选择,如 F_1 到 F_6 ,然而有些世代要经历多次选择,如 F_7 到 F_9 (图 1 和表 1)。对于只有一次选择的世代,种子来源只能来自前一个世代;对于包含多次选择的世代,如 F_7 包含 4 次选择,即 F_7 、 F_8 (T)、 F_8 (B) 和 F_8 (YT) (表 1),种子来源可以指明 F_8 (T)、 F_8 (B) 和 F_8 (YT) 的种子均来自 F_7 ,还是 F_8 (T) 的种子来自 F_7 、 F_8 (B) 的种子来自 F_8 (T)、 F_8 (YT) 的种子来自 F_8 (B)。这两种种子来源在育种实践中都是存在的。

1.1.2 种子繁殖类型 种子繁殖类型描述在中选的家系中植株通过何种方式产生下一世代的种子,为方便起见,笔者把各种控制杂交也视为不同的种子繁殖

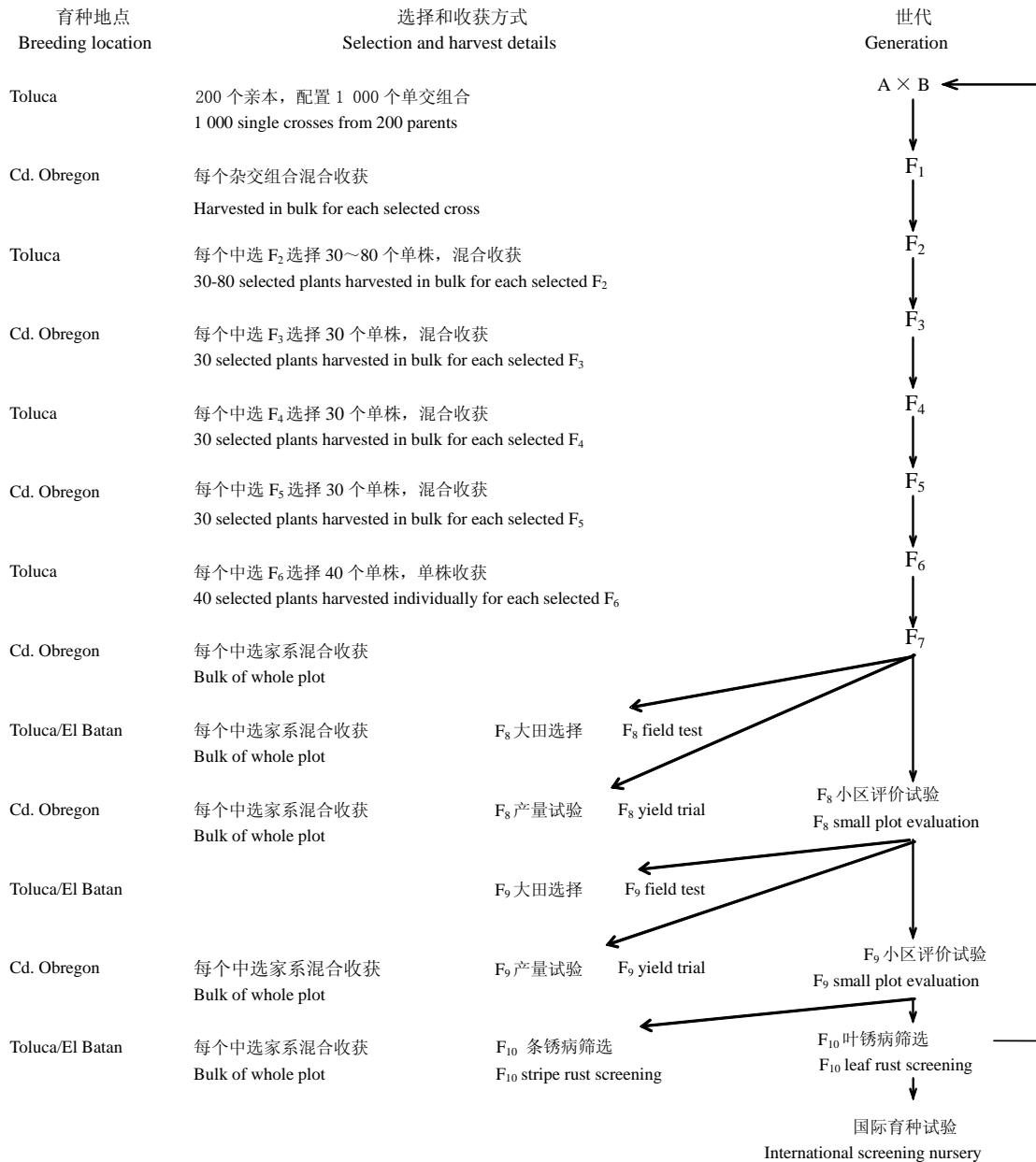


图 1 CIMMYT 面包小麦选育流程图 (图中所示育种方法称为选择混合 SELBLK)

Fig. 1 Germplasm flow in CIMMYT's wheat breeding program (The breeding strategy described was called selected bulk selection method)

类型, 这样常见的种子繁殖类型有以下 9 种, 即无性系繁殖 (clone)、加倍单倍体 (doubled haploids)、自交 (self-pollination)、单交 (single cross)、回交 (backcross)、顶交 (或三交) (topcross)、双交 (doublecross)、随机交配 (random mating) 和排除自交的随机交配 (self-pollination excluded random mating)。通过定义种子繁殖类型这一参数, 自花授粉作物的大多数繁殖方式和杂交方式都可以进行模

拟。选择混合育种方法只包含自交和单交这 2 种种子繁殖类型 (表 1)。

1.1.3 世代递进方式 世代递进方式描述中选家系内中选单株上种子的收获方式, 种子收获方式有两种, 用系谱 (pedigree) 和混合 (bulk) 表示。系谱意味着种子单株收获, 单株保存, 每个植株上的种子在下一个世代种植成一个家系; 混合意味着同一家系内所有中选单株上的种子混合收获, 混合种子在下一个世代

种植成一个家系。世代递进方式这一参数使得不仅能够模拟传统的系谱选择和混合选择, 而且可以模拟各种修饰系谱和混合选择方法, 如 F_2 和 F_3 世代作系谱选择, F_4 至 F_6 作混合选择, F_7 世代作系谱选择, 以后世代再作混合选择等。在选择混合育种方法中, 除 F_6 世代采用系谱递进方式外, 其它世代均采用混合递进方式, 这样从 F_1 到 F_6 , 家系数实际上等于保留下来的杂交组合数, 经过 F_6 系谱递进, F_7 世代的家系数等于 F_6 世代中所有中选家系中的中选单株数的总和。

1.1.4 田间设计 田间设计包括每个家系的测试地点数、重复数、小区内的植株数等。如表 1, F_2 的种植地点为 Toluca, 每个 F_2 群体种植 1 000 个分离单株, 1 次重复; 世代 F_9 (YT) 的种植地点为 Cd. Obregon, 每个家系种植 100 个单株, 2 次重复。

1.1.5 家系间选择和家系内选择 “优中选优” 其实概括了植物育种中的家系间和家系内两种选择类型, 在进行这两种类型的选择时, 可以对相同的性状也可以对不同的性状进行选择, 所选择的性状数可以相等也可以不等 (表 2)。对于每个被选择性状, 可以采用不同的选择类型和选择比例, 育种中常用的选择类型有 4 种, 即选择高表型值的个体或家系 (如产量和产量构成因素)、选择低表型值的个体或家系 (如倒伏性)、选择中间表型值的个体或家系 (如株高和生育期) 以及随机选择 (表 2)。在进行选择时, 育种家有时按照一定的比例进行选择, 有时按照指定的阈值进行选择, 当某一世代对多个性状进行选择时, 育种家有时采用独立水平选择 (independent culling levels) [13], 有时采用指数选择 (index selection) [13]。例如在选择混合育种方法的 F_5 世代中, 考虑 3 个不同性状作家系间选择, 即倒伏性、叶锈和分蘖数, 采用独立水平选择, 3 个性状的选择比例依次为 99%、60% 和 95% (表 2); 家系内选择时则考虑 7 个不同性状, 即倒伏性、叶锈、株高、抽穗期、分蘖数、穗粒数和千粒重, 选择比例依次为 90%、70%、90%、90%、80%、20% 和 60% (表 2)。

1.2 育种性状遗传模型的定义

在育种模拟过程中, 需要产生个体在不同育种性状上的表现型, 进而根据表现型作选择。在数量遗传中, 表现型定义为基因型加上环境效应和随机误差 [12-14]。要产生基因型值, 就需要定义控制育种性状的所有基因在不同环境下的遗传效应, 然后根据性状遗传力的大小来产生随机环境误差。常见的数量性状遗传模型包括加性-显性效应模型、加性-显性-上位性

位性效应模型、连锁、复等位基因 (multiple alleles)、一因多效 (pleiotropy)、显性和共显性分子标记等等 [19, 21, 22], 当然基因的定义也包括在染色体上的位置。对于选择混合育种方法, 共有 10 个性状用于选择, 包括 7 个农艺性状和 3 个锈病性状。控制每个性状的基因数和基因的效应经过和 CIMMYT 育种家讨论并参考相关遗传研究结果后确定, 表 3 列出 59 个控制性状的独立遗传基因及其在 Cd. Obregon 环境下的遗传效应。除产量性状基因, 基因效应被认为是固定效应, 考虑一因多效, 以解释性状间的相关 (表 3)。

对产量来说, 除受其它基因多效应的影响外 (表 3), 还假定有 20 个产量基因, 每个位点包含 2 个等位基因。对这 20 个产量基因我们考虑 4 种遗传模型, 用 AD0、AD1、AD2 和 ADE 表示。模型 AD0 为纯加性模型, 即对于基因效应有 $Aa = (AA+aa) / 2$ (即显性效应 $d=0$), A 和 a 为 2 个等位基因; 模型 AD1 为部分显性模型, 即对于基因效应有 $Aa \neq (AA+aa) / 2$, 但 Aa 的表现介于 AA 和 aa 之间; 模型 AD2 可能包含加性、部分显性或超显性, 即 3 种基因型的效应没有任何限制; 模型 ADE 考虑 2 个基因位点间的上位性。

1.3 育种模拟工具 QuLine 的建立

QU-GENE 由澳大利亚昆士兰大学研制 [19], QuLine 由 CIMMYT 研制 [21, 22], 二者的结合实现了在复杂的遗传模型下对纯系育种过程的模拟 (图 2)。使用 QuLine 之前, 用户需要建立 2 个输入文件, 第一个输入文件包含育种目标环境群体 (TPE, target population of environments) 的构成、育种性状的遗传模型和初始育种亲本群体等信息, QU-GENE 利用这些信息产生 2 个输出文件供 QuLine 调用, 一个输出文件用于定义模拟过程中所需的基因和环境系统 (gene and environment system), 另一个用于定义育种起始亲本群体; 第二个需要用户建立的文件包含所要模拟的育种策略, 这个文件中可以包含多个育种策略, 以保证在同样的遗传模型和亲本群体下比较育种策略间的差异。

QuLine 利用 QU-GENE 的 2 个输出文件和用户创建的育种策略文件 (图 2), 按照育种策略所指定的组合数配置杂交组合, 按照种子繁殖方式产生育种后代材料, 按照田间设计方案评价育种材料的表现, 按照家系间和家系内选择信息选择后代材料, 按照指定的世代递进方法产生下一世代的育种材料 [23]。QuLine 重复这一过程直到完成一个完整的育种周期为至, 如果用户指定的模拟周期数大于 1, QuLine 则在结束一

个育种周期后，把保留下来的家系作为下一个周期的亲本，然后开始下一个育种周期的模拟。QuLine 结束一个育种周期的模拟后，把终选群体的各种遗传参数写入不同的输出文件中，图 2 给出其中的 3 个输出文件，终选群体在不同育种性状上的表现可以用于遗传进度的计算从而比较不同选择方法的育种效果，终选

群体中的基因频率可以用于研究选择前后等位基因和群体遗传多样性的变化，每个世代中来自不同杂交组合的家系和单株数可以用于比较杂交组合的优劣。其它输出信息包括群体遗传方差、家系选择史、基因固定和丢失等等，这样用户可以根据不同的研究目的使用不同的输出信息。

表 3 模拟试验中所考虑的育种性状及在 Cd. Obregon 环境条件下的遗传模型

Table 3 Number of segregating genes and their genetic effects in the Cd. Obregon environment type

基因分类 Gene classification	基因数 Number of genes	影响性状 Traits effected	单个基因效应 Individual gene effects		
			AA	Aa	aa
产量 Yield	20	产量 Yield (t/ha)	产量基因的 4 个模型: AD0 (纯加性 pure additive), AD1 (部分显性 partial dominance), AD2 (超显性 overdominance), ADE (2 个位点间的上位性 digenic epistasis)		
倒伏性 Lodging	3	倒伏性 Lodging (%)	0.00	5.00	10.00
		产量 Yield (t/ha)	0.00	-0.40	-0.80
茎锈 Stem rust	5	秆锈 Stem rust (%)	0.00	0.50	1.00
		产量 Yield (t/ha)	0.00	-0.25	-0.50
叶锈 Leaf rust	5	粒重 Kernel weight (g)	0.00	-0.75	-1.50
		叶锈 Leaf rust (%)	0.00	5.00	10.00
		产量 Yield (t/ha)	0.00	-0.25	-0.50
株高 Height	3	粒重 Kernel weight (g)	0.00	-0.75	-1.50
		株高 Height (cm)	40.00	30.00	20.00
		倒伏性 Lodging (%)	5.00	2.50	0.00
抽穗期 Maturity	5	抽穗期 Maturity (day)	20.00	16.00	12.00
		粒重 Kernel weight (g)	-1.00	-0.50	0.00
分蘖数 Tillering	3	分蘖数 Tillering (no.)	5.00	3.00	1.00
		倒伏性 Lodging	2.00	1.00	0.00
		抽穗期 Maturity (day)	1.00	0.50	0.00
穗粒数 Grains per ear	5	穗粒数 Grains per ear	-1.00	-0.50	0.00
		粒重 Kernel weight (g)	-1.50	-0.75	0.00
		穗粒数 Gains per ear	14.00	10.00	6.00
		倒伏性 Lodging (%)	2.00	1.00	0.00
粒重 Kernel weight	5	粒重 Kernel weight (g)	-1.00	-0.50	0.00
		粒重 Kernel weight (g)	12.00	8.50	5.00
		产量 Yield (t/ha)	1.00	0.50	0.00
		倒伏性 Lodging (%)	2.00	1.00	0.00

2 模拟工具 QuLine 的育种应用

2.1 修饰系谱和选择混合两种育种方法的比较

修饰系谱法 (MODPED, modified pedigree method) 和选择混合法 (SELBLK, selected bulk method) 是 CIMMYT 小麦育种中经常采用的 2 种育种方法^[21, 24]。图 1 给出的是选择混合法的流程, 这两种方法的主要区别在于: 修饰系谱法在 F₂ 代进行单株系谱选择, F₃~F₅ 进行混合选择, F₆ 又采用单株系谱选择, 而选择混合法从 F₂~F₅ 进行混合选择, 只有 F₆ 采用单株系谱选择 (表 1, 图 1)。根据表 1、2 和 3 定义的选择方法和遗传模型对这两种育种方法进行模拟。首先利用 QU-GENE^[19]产生一个群体大小为 200

的初始群体做亲本, 群体的基因频率均设定为 0.5; 然后 QuLine 对两种方法分别配置 1 000 个杂交组合, 经过 10 个世代的选择后, 得到 258 个高代近交系, 利用这 258 个近交系组成的群体来比较这种选择方法的遗传进度。为了在不同的模型间比较, 采用校正的性状值, 假定 TG_H 和 TG_L 分别是最高目标基因型值和最低目标基因型值, y 为个体的基因型值, 则校正基因型值 y_{ad} 定义为:

$$y_{ad} = \frac{y - TG_L}{TG_H - TG_L} \times 100\%$$

初始亲本群体中, 由于基因频率为 0.5, 群体平均产量位于最低和最高目标基因型的中间, 产量的校正值为 50% (图 3)。经过一个育种周期后, 修饰系谱

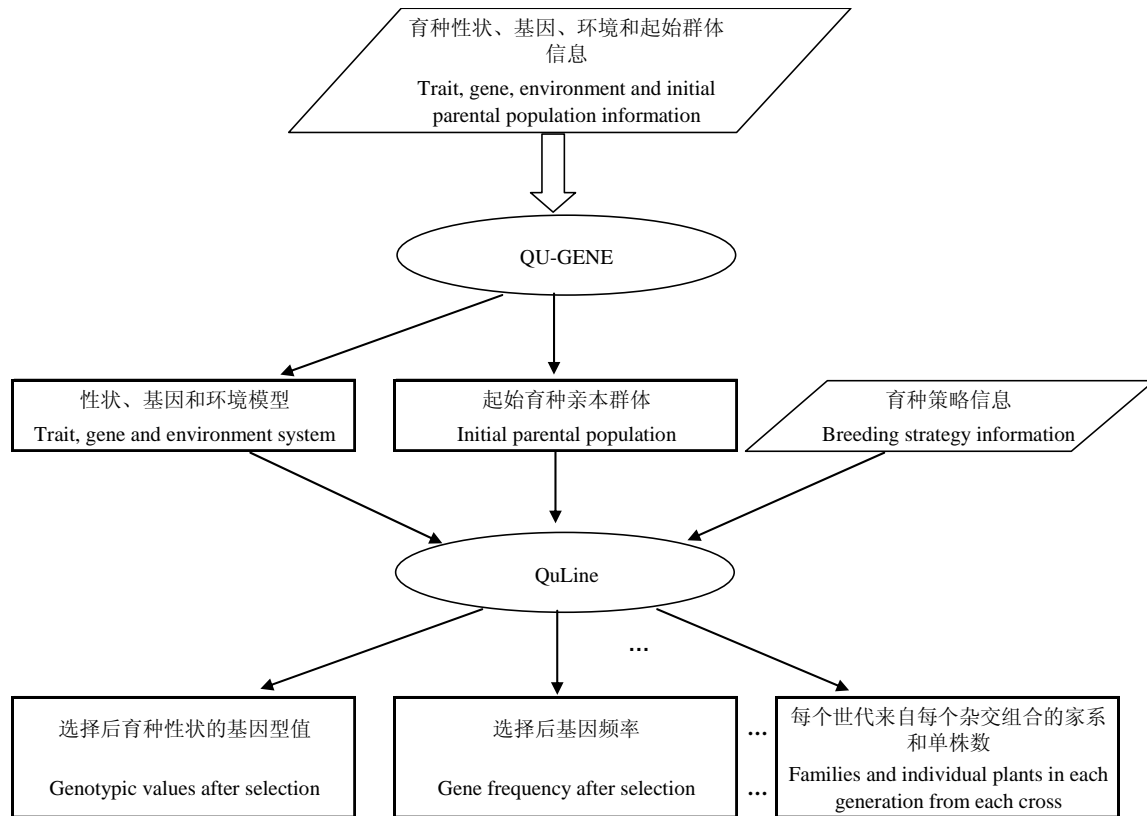
法把群体的产量基因型值提高到 55.83%，选择混合法把群体的产量基因型值提高到 56.02%（图 3-A），因此从产量性状的遗传增益上看，选择混合育种方法要略优于修饰系谱育种方法。两种方法在 F₁ 代的杂交数均为 1 000，F₁ 代选择后淘汰约 30% 的组合，经过 10 个世代的选择后，在中选的 258 个近交系中，修饰系谱平均保留了 118 个组合，而选择混合平均保留了 148 个组合，在中选群体中，选择混合育种方法要比修饰系谱育种方法保留的组合多 30%（图 3-B），较多的组合数，意味着较高的遗传多样性，因此从中选群体的遗传多样性上看，选择混合育种方法要明显优于修饰系谱育种方法。从两种方法分别产生的家系数和种植的单株数来看，从 F₁ 至 F₈，选择混合育种方法产生的家系数只是修饰系谱育种方法的 40%，选择混合育种方法种植的植株数只是修饰系谱育种方法的三分之

二（图 3-C 和图 3-D），因此选择混合育种方法花费较少的劳力、占用较少的土地资源，从经济的角度看，选择混合育种方法也明显优于修饰系谱育种方法。

2.2 利用已知基因信息的亲本选配

选择合适的亲本配置杂交组合是育种成败的关键，然而由于缺乏对育种目标性状遗传的了解，杂交方案多依据亲本材料的表型性状来确定，因此育种实践中在看似理想的杂交组合往往得不到理想的后代材料。例如在图 1 的育种程序中，经过 F₁ 到 F₈ 的选择要淘汰掉 70% 左右的杂交组合，然后经过 2 次产量试验，只有不到 15% 的组合保留下来。遗传研究的深入使我们获得了许多重要农艺性状的遗传信息，育种家希望能够利用这些信息对不同杂交组合的后代表现进行预测，从而根据后代的表现选择最佳的杂交组合。

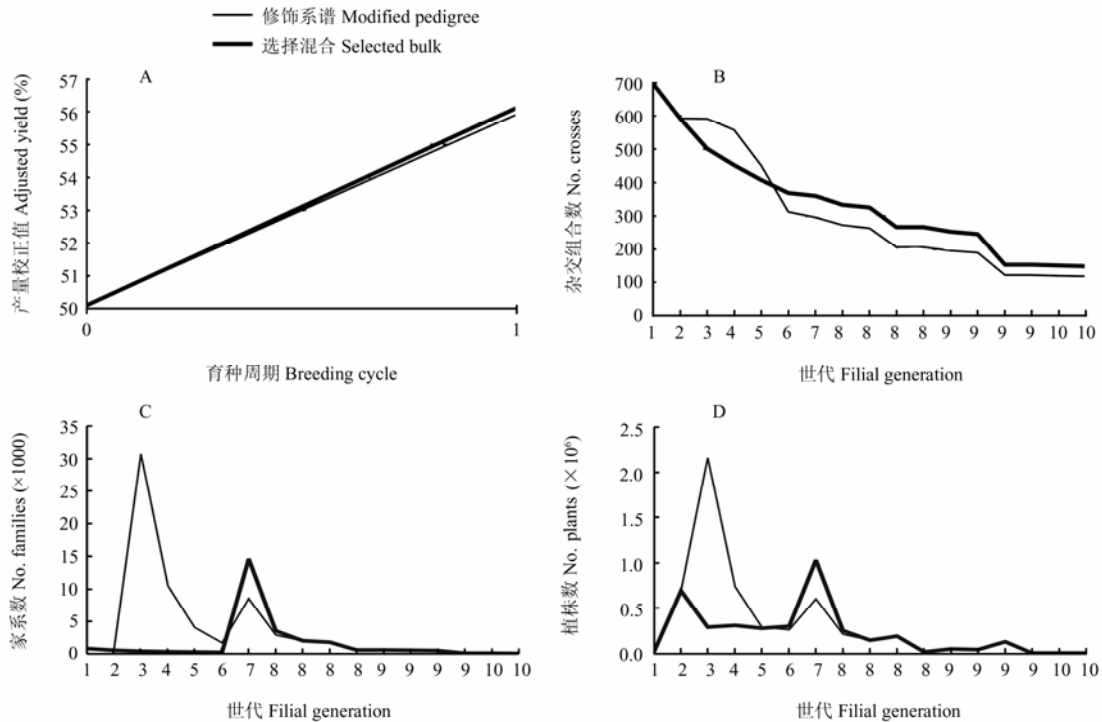
QuLine 可以利用这些已知的遗传信息，通过模拟试验



椭圆表示计算机程序 QU-GENE 和 QuLine，平行四边形表示用户输入信息，长方形表示输出结果
The two ellipses represent the two computer programs, i.e. QU-GENE and QuLine, rectangles represent inputs for QU-GENE and QuLine, and parallelograms represent outputs from QU-GENE and QuLine

图 2 利用 QuLine 开展育种模拟的流程图

Fig. 2 Flowchart of the breeding simulation tool QuLine



A: 产量性状一个育种周期后的遗传进度。B: 每个世代选择结束后保留下来的杂交组合数。C: 每个世代所种植的家系数。D: 每个世代所种植的单株数
A, Adjusted genetic gain after one breeding cycle across all experimental sets. B, Number of crosses after each generation's selection across all experimental sets. C, Number of families in each generation in one breeding cycle. D, Number of individual plants in each generation in one breeding cycle

图3 修饰系谱和选择混合2种育种方法的模拟比较

Fig. 3 Comparison of modified pedigree and selected bulk from the simulation experiment

来预测杂交后代的表现,为杂交方案的确定提供依据。这里我们以小麦品质育种为例说明 QuLine 在利用已知基因信息进行亲本选配中的应用^[25]。

研究发现3个决定低分子量谷蛋白亚基的基因位点 *Glu-A3*、*Glu-B3* 和 *Glu-D3*, 以及3个决定高分子量谷蛋白亚基的基因位点 *Glu-A1*、*Glu-B1* 和 *Glu-D1* 对面包小麦的2个重要品质性状,即面团最大抗延阻力(Rmax, maximum dough resistance)和面团延展性(Dough extensibility)有较大的影响^[25-27]。这6个基因位点上均存在复等位基因,并且已知这6个位点在染色体上的位置以及每个等位基因的遗传效应^[26,27]。Silverstar 是一个澳大利亚面包小麦品种,是 CIMMYT 小麦 Pavon 和澳大利亚小麦 TM56 的杂交后代, Silverstar 有很多姊妹系,这里选其中的8个用 Silverstar 1 到 Silverstar 8 表示。在位点 *Glu-A1*, *Glu-B3* 和 *Glu-D3* 上,它们有相同的基因;在位点 *Glu-B1* 上,它们有2个不同的等位基因 *b* 和 *i*,但这2个基因在最大抗延阻力和延展性上有近似的遗传效应。8个

Silverstar 姊妹系在 *Glu-A3* 位点上的等位基因 *b* 和 *c* 中,等位基因 *b* 比 *c* 对最大抗延阻力性状有较大的效应;在 *Glu-D1* 位点上的等位基因 *a* 和 *d* 中,等位基因 *a* 比 *d* 对最大抗延阻力性状有较大的效应。一个实际的问题是“这些姊妹系在作为杂交亲本时是否有显著差异?”

在模拟试验中,选用其他4个澳大利亚小麦品种 Westonia、Krichauff、Machete 和 Diamondbird 与8个 Silverstar 姊妹系作杂交。Westonia 在 *Glu-A3* 位点上没有等位基因 *b*,在 *Glu-D1* 位点上没有等位基因 *a*; Krichauff 在 *Glu-A3* 位点上没有等位基因 *b*,但在 *Glu-D1* 位点上有等位基因 *a*; Machete 在 *Glu-A3* 位点上有等位基因 *b*,但在 *Glu-D1* 位点上没有等位基因 *a*; Diamondbird 在 *Glu-A3* 位点上有等位基因 *b*,在 *Glu-D1* 位点上有等位基因 *a*。利用 QuLine,在选用的4个品种和8个 Silverstar 姊妹系间配置相关杂交组合,每个杂交组合采用一粒传的方法产生1000个 F_8 近交家系,采用4种不同的选择方法,每种方法最终选择40

个家系, 选择比例为 4% (表 4)。

不同育种目标下, 能把待改良亲本的目标性状提高 5% 以上的 Silverstar 姊妹系列于表 4, 在与 Westonia 的杂交组合中, 当最大抗延阻力被选择时, Silverstar 3 和 7 表现出最大的遗传增益。这些姊妹系也能提高 Westonia 的延展性, 尤其是延展性也是被选择的性状。当高最大抗延阻力和高延展性是期望的品质性状, 但高最大抗延阻力更重要时, Silverstar 3 和 7 都是理想的杂交亲本, 但 Silverstar 3 更佳。在与 Krichauff 的杂交组合中, 如果最大抗延阻力是唯一的选择性状或者是第一个选择性状, Silverstar 1, 3, 5 和 7 在两个性状上有近似相同的遗传增益。如果延展性是唯一的选择性状或者是第一个选择性状, Silverstar 3 和 7 是理

想的杂交亲本。在与 Machete 的杂交组合中, 如果最大抗延阻力是唯一的选择性状或者是第一个选择性状, Silverstar 3, 4, 7 和 8 都能较大提高 Machete 的最大抗延阻力。如果要同时改良延展性, 应首先选择最大抗延阻力; 如果先选择延展性, 然后选择最大抗延阻力, 则应该选择 Silverstar 4 或 8 与 Machete 的杂交。在与 Diamondbird 的杂交组合中, 如果最大抗延阻力是选择性状, Silverstar 1, 2, 3 和 4 只是轻微地提高最大抗延阻力, 如果延展性是选择性状, 只有 Silverstar 3 和 4 轻微地提高最大抗延阻力和延展性。从模拟试验的结果来看, 即便是表型相近的姊妹系与相同亲本的杂交后代也有很大的差异。

表 4 不同育种目标下适宜于待改良亲本的 Silverstar 姊妹系

Table 4 The best Silverstar sister lines for the four selected parents, under different breeding objectives

待改良亲本 Parent to be improved	育种目标 Breeding objective	选择方案 Selection scheme			
		R0.04	R0.2E0.2	E0.2R0.2	E0.04
Westonia	高抗延阻力 High Rmax (BU)	3, 7	3, 7	3, 7	1, 3
	高延展性 High extensibility (cm)	1	1, 5	1, 3, 5	1, 3, 5, 7
Krichauff	高抗延阻力 High Rmax (BU)	1, 3, 5, 7	1, 3, 5, 7	3, 7	3, 7
	高延展性 High extensibility (cm)	1, 3, 5, 7	1, 3, 5, 7	1, 5	1, 5
Machete	高抗延阻力 High Rmax (BU)	3, 4, 7, 8	3, 4, 7, 8	4, 8	无 None
	高延展性 High extensibility (cm)	1, 2, 5, 6	1, 2, 5, 6	1, 2, 3	1, 2, 3, 4
Diamondbird	高抗延阻力 High Rmax (BU)	1, 2, 3, 4	1, 3, 4	3, 4	3, 4
	高延展性 High extensibility (cm)	无 None	无 None	1, 2, 5, 6	1, 2, 5, 6

R 表示最大抗延阻力, E 表示延展性, 性状后面的数字为选择比例 R for Rmax, E for extensibility, trait followed by selected proportion

2.3 制定设计育种方案

Bernardo 和 Peleman 等提出了设计育种的概念^[1, 28], 万建民报道了在中国开展设计育种的必要性并提供了设计育种的实例^[29], 设计育种可分 3 步进行:

(1) 定位所有相关农艺性状的 QTL; (2) 评价这些位点的等位性变异; (3) 建立设计育种方案、开展设计育种。设计育种将使育种方法由目前的以表型选择为主转变为以基因型为主, 这里我们以水稻品质育种为例说明 QuLine 在制定设计育种方案中的应用。

水稻直链淀粉含量 (amylose content, AC) 和垩白度 (Area of chalky endosperm, ACE) 是两个重要的品质指标^[30], 利用一套水稻染色体片段置换系 (chromosome segment substitution lines, CSSL)^[8, 30] 定位了某一环境条件下控制 AC 和 ACE 的 QTL, 结果见表 5。作图群体的两个亲本为粳稻品种 Asominori (背景亲本) 和籼稻品种 IR24 (供体亲本), 群体包含 65 个置换系 (分别用 CSSL1-CSSL65 表示), 每

个置换系包含一到几个供体亲本 IR24 的染色体片段, 共计有 82 个供体片段 (分别用 M1-M82 表示), 覆盖了水稻的 12 条染色体。根据 QTL 定位的结果, 可以预测各种可能基因型的表现。假定育种目标是培育具有高直链淀粉含量和低垩白度的品种, 由于在片段 M35、M57 和 M59 上的 QTL 具有不利的一因多效, 因此, 不可能育成同时具有最高 AC 和最小 ACE 的品种。通过模拟试验提出一个设计基因型, 含有 4 个 IR24 的供体片段 (M19、M35、M57 和 M60), 其它染色体片段与 Asominori 相同, 这一设计基因型的 AC 值为 17.73% (理论最高值 22.3%), ACE 值为 9.3% (理论最低值 0) (表 6)。在 65 个置换系中, 发现其中 3 个置换系 CSSL15、CSSL29 和 CSSL49 包含有目标基因型片段, 这 3 个置换系可以作为该项育种计划的亲本材料 (表 6)。

3 个亲本间的三交 (又称顶交) 组合有可能将我们需要的染色体片段聚合在一起。产生三交组合的方

式有3种(表7),即三交组合1:(CSSL15×CSSL29)×CSSL49;三交组合2:(CSSL15×CSSL49)×CSSL29和三交组合3:(CSSL29×CSSL49)×CSSL15。假定我们采用标记辅助方法选择目标基因型,可供选择的方案有很多,这里只考虑其中的2种,标记选择方案1:产生200个三交F₁个体,每个产生20个DH(doubled

haploid)家系,共产生4000个DH家系,然后从中选择目标基因型;标记选择方案2:产生200个三交F₁个体,通过标记辅助选择只保留含有目标染色体片段的个体,每个中选个体产生20个DH家系,然后从中选择目标基因型,以上过程借助遗传育种模拟工具QuLine实现。通过100次模拟试验,方案1在三交组

表5 利用65个染色体片段置换系对水稻AC和ACE的QTL定位结果

Table 5 QTL mapping results of AC and ACE in the population consisting of 65 CSS lines

控制 AC 的 QTL QTL for AC									
标记 Marker	M6*	M14**	M21*	M35*	M38*	M57**	M59**	M60**	M63**
LOD 值 LOD score	1.07	2.60	1.40	0.92	1.37	7.24	4.66	4.34	1.48
加性效应 Additive effect (%)	0.47	-0.61	-0.35	-0.36	-0.43	1.12	1.03	0.71	0.45
贡献率 Percentage of variance explained (%)	1.89	4.83	2.48	1.62	2.41	15.97	9.28	8.59	2.59
控制 ACE 的 QTL QTL for ACE									
标记 Marker	M19*	M35**	M38*	M39*	M43*	M57**	M59**		
LOD 值 LOD score	0.94	2.16	1.19	1.54	1.23	16.86	10.02		
加性效应 Additive effect (%)	-1.80	-1.63	1.20	-1.31	-0.88	5.93	4.96		
贡献率 Percentage of variance explained (%)	1.10	2.66	1.43	1.70	1.47	35.00	16.56		

*显著水平 0.05, **显著水平 0.01。*Significance level 0.05, **Significance level 0.01

表6 一个设计基因型和3个染色体片段置换系的标记基因型和性状预测值

Table 6 Marker types and predicted genetic values on AC and ACE of a designed genotype and 3 CSS lines

染色体 Chromosome	3	3	5	8	9	预测值 Predicted value	
标记 Marker	M19	M21	M35	M57	M60	AC (%)	ACE (%)
设计基因型 Designed genotype	2	1	2	2	2	17.73	9.27
CSSL15	2	2	1	1	1	14.09	0.55
CSSL29	1	1	2	1	1	14.07	0.88
CSSL49	1	1	1	2	2	18.44	16.13

1和2分别代表来自背景亲本 Asominori 和供体亲本 IR24 的染色体片段

1 and 2 represent the chromosome segment from background parent Asominori and donor parent IR24, respectively

表7 不同三交组合及选择方案的比较

Table 7 Comparison of the three top crosses and the two marker selection schemes

标记选择方案 Marker selection scheme	选择前 TCF ₁ 个体数 Individuals in TCF ₁ before selection	选择后 TCF ₁ 个体数 Individuals in TCF ₁ after selection	选择前 DH 家系数 DH lines before selection	选择后 DH 家系数 DH lines after selection (S.E.)	DNA 样品总数 DNA samples to be tested	选中每个家系测定的 DNA 样品平均数 DNA samples per selected line
三交组合 1 Topcross 1: (CSSL15 × CSSL29) × CSSL49						
方案 1 Scheme 1	200	200	4000	27.1 (6.6)	4 000	148
方案 2 Scheme 2	200	13	262	16.7 (6.2)	462	28
三交组合 2 Topcross 2: (CSSL15 × CSSL49) × CSSL29						
方案 1 Scheme 1	200	200	4000	12.9 (4.9)	4 000	310
方案 2 Scheme 2	200	6	124	7.9 (4.5)	324	41
三交组合 3 Topcross 3: (CSSL29 × CSSL49) × CSSL15						
方案 1 Scheme 1	200	200	4000	7.5 (3.1)	4 000	536
方案 2 Scheme 2	200	25	491	7.7 (3.1)	691	89

合 1 中选出 27 个具有设计基因型的 DH 家系,在三交组合 2 中选出 13 个 DH 家系,在三交组合 3 中选出 8 个 DH 家系。因此,采用三交组合 1 选中设计基因型的几率最大。如果考虑到 DNA 测试的费用,则两种方案差异很大。方案 1 需要测试 4 000 个 DNA 样品,而方案 2 需要测定的 DNA 样品数则大大减少,而且不同组合间存在差别,组合 2 需要测试的样品数最少(表 7),仅 324 个。因此,在所考察的杂交和选择方案中,最佳的育种策略为:配制三交组合 1,后代用方案 2 进行选择。

3 讨论

CIMMYT 的小麦育种方法在 1984 年前以系谱方法为主,1985 至 1994 年间以修饰系谱方法为主,之后多采用选择混合法^[24]。在对修饰系谱法和选择混合法 2 种育种方法进行模拟之前,CIMMYT 的育种家已经知道选择混合法可以节省费用(图 3-C 和 3-D),但是不太清楚选择混合法的使用是否会影响产量性状的遗传进度。模拟结果表明选择混合法对产量的遗传改良不但没有下降,而且还获得略高于修饰系谱法的遗传进度(图 3-A);模拟结果还表明,在同样大小的终选群体中,选择混合法保留下来较多的杂交组合(图 3-B),这是育种家事先没有意识到的,但是这一结果随后被多年的田间记载资料所证实。因此育种模拟不但回答了育种家所关心的问题,而且还提供了育种家没有意识到的信息。

模拟研究需要与遗传和育种研究结合,模拟结果也有待育种实践的检验,同时模拟结果和田间数据间的偏差也有助于不断改进育种模拟中采用的遗传模型。就目前来讲,对重要育种性状的认识仍然很不完整,尚不能全面准确提供模拟方法所需要的数据和信息。文中对两种选择方法的比较可视为较宏观的问题,在遗传模型不确定的情况下,可在大量的可能遗传模型下进行模拟比较,如果不同模型或对大多数模型都得到相对一致的结果,那么采用什么样的遗传模型不会对我们的结论有大的影响。对另外 2 个实例来讲,所面对的则是较具体的问题,这时需要准确详实的遗传信息。可以预见,随着生物信息数据的增加,模拟试验可定义的遗传模型将更加符合生物体本身的规律^[16,31]。另一方面,遗传数据可以来自生物的不同水平,如群体水平、个体水平、孟德尔基因水平和 DNA 分子水平等,植物生长模型的发展有可能将遗传模型置于多个模拟的生长环境中来预测目标性状的表现

型^[32,33],同时也为基因型到表型建模提供新方法。

4 结论

育种模拟去掉了经典数量遗传研究中的种种假定,实现了对复杂育种过程在真实遗传模型下的模拟,为育种家提供了一个评价不同育种方法选择效果和利用各类基因信息(如基因在染色体上的位置、QTL 的遗传效应、标记和基因间的连锁关系等)的平台。育种模拟工具 QuLine 可以在育种家的田间试验之前,利用各种遗传信息对杂交组合的表现、后代选择效果以及整个育种过程进行模拟,提出最佳的亲本选配、杂交和后代选择策略,从而提高育种效率;QuLine 可以利用大量已知基因信息和 QTL 作图结果,模拟不同世代标记辅助选择的效果,设计符合不同育种目标的基因型并提供获得目标基因型的最佳育种途径,从而在大量的生物信息数据和育种家的需求之间搭起一座桥梁,把大量的遗传研究结果变成能为育种家利用并为育种服务的知识。(模拟工具 QuLine 对国内研究人员免费提供,有需要的读者请与作者联系)

致谢: QuLine 研制过程中得到澳大利亚谷物研究和开发行业协会(Grains Research and Development Corporation, Australia)(2000-2004)、国际遗传资源挑战项目(Generation Challenge Program)(2005-2007)和国际生物加强挑战项目(Harvest Challenge Program)(2006-2007)的资助,特此致谢。

References

- [1] Bernardo R. *Breeding for Quantitative Traits in Plants*. Woodbury, Minnesota: Stemma Press, 2002.
- [2] Fehr W R. *Principles of Cultivar Development. Vol. 1. Theory and Technique*. New York: Macmillian Publishing Company, 1987.
- [3] Zeng Z B. Precision mapping of quantitative trait loci. *Genetics*, 1994, 136: 1457-1468.
- [4] Barton N H, Keightley P D. Understanding quantitative genetic variation. *Nature Review Genetics*, 2002, 3: 11-21.
- [5] Li Z K, Yu S B, Lafitte H R, Huang L, Courtois B, Hittalmani S, Vijayakumar C H M, Liu G F, Wang G C, Shashidhar H E, Zhuang J Y, Zheng K L, Singh V P, Sidhu J S, Srivantaneeyakul S, Khush G S. QTL \times environment interactions in rice. I. Heading date and plant height. *Theoretical and Applied Genetics*, 2003, 108: 141-153.
- [6] Kroymann J, Mitchell-Olds T. Epistasis and balanced polymorphism influencing complex trait variation. *Nature*, 2005, 435: 95-98.
- [7] Tanksley S D, Nelson J C. Advanced backcross QTL analysis: a

- method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics*, 1996, 92: 191-203.
- [8] Wan X Y, Wan J M, Jiang L, Wang J K, Zhai H-Q, Weng J F, Wang H L, Lei C H, Wang J L, Zhang X, Cheng Z J, Guo X P. QTL analysis for rice grain length and fine mapping of an identified QTL with stable and major effects. *Theoretical and Applied Genetics*, 2006, 112: 1258-1270.
- [9] Frary An, Nesbitt T C, Frary Am, Grandillo S, Knaap E V D, Cong B, Liu J P, Meller J, Elber R, Alpert K B, Tanksley S D. *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science*, 2000, 289: 85-88.
- [10] Comstock R E, Robinson H F, Harvey P H. A breeding procedure designed to make maximum use of both general and specific combining ability. *Agronomy Journal*, 1952, 41: 360-367.
- [11] Goldman I L. Prediction in plant breeding. *Plant Breeding Reviews*, 2000, 19: 15-40.
- [12] Lynch M, Walsh B. *Genetics and Analysis of Quantitative Genetics*. Sunderland, MA: Sinauer Associates, Inc., 1998.
- [13] Falconer D S, Mackay T F C. *Introduction to Quantitative Genetics*. (4th edition). Essenx, England: Longman, 1996.
- [14] 马育华编著. 植物育种的数量遗传学基础. 南京: 江苏科学技术出版社, 1982.
- Ma Y H. *The Principle of Quantitative Genetics in Plant Breeding*. Nanjing: Jiangsu Science and Technology Press, 1982. (in Chinese)
- [15] Kempthorne O. *An Overview of the Field of Quantitative Genetics*. In: Weir B S, Eisen E J, Goodman M M, Namkoong G (eds). *Proceedings of the 2nd International Conference on Quantitative Genetics*, Sunderland, MA: Sinauer Associates, Inc., 1988: 47-56.
- [16] Cooper M, Podlich D W, Smith O S. Gene-to-phenotype and complex trait genetics. *Australian Journal of Agricultural Research*, 2005, 56: 895-918.
- [17] Frisch M, Bohn M, Melchinger A E. Comparison of selection strategies for marker-assisted backcrossing of a gene. *Crop Science*, 1999, 39: 1295-1301.
- [18] Frisch M, Melchinger A E. Marker-assisted backcrossing for simultaneous introgression of two genes. *Crop Science*, 2001, 41: 1716-1725.
- [19] Podlich D, Cooper M. QU-GENE: A platform for quantitative analysis of genetic models. *Bioinformatics*, 1998, 14: 632-653.
- [20] Podlich D, Cooper M, Basford K E. Computer simulation of a selection strategy to accommodate genotype-environment interactions in a wheat recurrent selection program. *Plant Breeding*, 1999, 118: 17-28.
- [21] Wang J, Ginkel M, Podlich D, Ye G, Trethowan R, Pfeiffer W, DeLacy I H, Cooper M, Rajaram S. Comparison of two breeding strategies by computer simulation. *Crop Science*, 2003, 43:1764-1773.
- [22] Wang J, Ginkel M, Trethowan R, Ye G, DeLacy I H, Podlich D, Cooper M. Simulating the effects of dominance and epistasis on selecting response in the CIMMYT wheat breeding program using QuCim. *Crop Science*, 2004, 44:2006-2018.
- [23] Wang J, Van Ginkel M, Basford K, Cooper M, Delacy I, Pfeiffer W, Podlich D, Trethowan R, Ye G. *QuCim 1.1 (CD-ROM)*. Mexico D.F.: CIMMYT, 2004.
- [24] van Ginkel M, Trethowan R, Ammar K, Wang J, Lillemo M. *Guide to Bread Wheat Breeding at CIMMYT (rev)*. *Wheat special report No. 5*. D.F., Mexico: CIMMYT, 2002.
- [25] Wang J, Eagles H A, Trethowan R, Ginkel M. Using computer simulation of the selection process and known gene information to assist in parental selection in wheat quality breeding. *Australian Journal of Agricultural Research*, 2005, 56: 465-473.
- [26] Eagles H A, Hollamby G J, Gororo N N, Eastwood R F. Estimation and utilization of glutenin gene effects from the analysis of unbalanced data from wheat breeding programs. *Australian Journal of Agricultural Research*, 2002, 53: 367-377.
- [27] Eagles H A, Eastwood R F, Hollamby G J, Martin E M, Cornish G B. Revision of the estimates of glutenin gene effects at the *Glu-B1* locus form southern Australian wheat breeding programs. *Australian Journal of Agricultural Research*, 2004, 55: 1093-1096.
- [28] Peleman J D, Voort J R. Breeding by design. *Trends in Plant Science*, 2003, 8: 330-334.
- [29] 万建民. 作物分子设计育种. 作物学报, 2006, 32: 455-462.
- Wan J M. Perspectives of molecular design breeding in crops. *Acta Agronomica Sinica*, 2006, 32: 455-462. (in Chinese)
- [30] Wan X Y, Wan J M, Weng J F, Jiang L, Bi J C, Wang C M, Zhai H Q. Stability of QTLs for rice grain dimension and endosperm chalkiness characteristics across eight environments. *Theoretical and Applied Genetics*, 2005, 110: 1334-1346.
- [31] Hammer G L, Chapman S, van Oosterom E, Podlich D. Trait physiology and crop modeling as a framework to link phenotypic complexity to underlying genetic systems. *Australian Journal of Agricultural Research*, 2005, 56:947-960.
- [32] Reymond M, Muller B, Leonardi A, Charcosset A, Tardieu F. Combining quantitative trait loci and an ecophysiological model to analyze the genetic variability of responses of maize leaf growth to temperature and water deficit. *Plant Physiology*, 2003, 131: 664-675.
- [33] Yin X, Struik P C, Kropff M J. Role of crop physiology in predicting gene-to-phenotype relationships. *Trends in Plant Science*, 2004, 9: 426-432.

(责任编辑 孙雷心)