

判定纤维素酶系基因组融合重组的统计分析法

艾云灿^{1*} 赵学慧¹ 余家林²

(华中农业大学食品科学系¹和基础课部², 武汉 430070)

摘要 根据纤维素酶系中组分间协同作用现象及多元回归分析原理, 提出同一融合组合 *Aspergillus niger* × *Trichoderma reesei* 子代群体中, 由于遗传物质重组传递的独立随机性, n 个菌株各组分酶活性观测值相当于这一特定融合试验的 n 组观察值, 可作多元线性回归模型分析, 判断融合重组发生。40 个融合子子代菌株分析结果表明, 此法可准确有效的指导纤维素酶高产菌株选育。

关键词 融合育种, 纤维素酶系, 协同作用, 统计分析, *Aspergillus niger*, *Trichoderma reesei*

开发利用纤维素再生资源的诱人前景, 促使纤维素酶系高产菌株选育日益活跃^[1,2]。大量研究表明, 纤维素酶系是由内切葡聚糖酶 (E. C. 3.2.1.4, EG), 外切葡聚糖酶 (E. C. 3.2.1.91, 主要是 Cellobionhydrolase, CBH), β -葡萄糖苷酶 (E. C. 3.2.1.21, BGL) 三个主要组分 (及其各自的多亚组分) 组成的诱导型复合酶系。前两者 (EG + CBH) 主要溶解纤维素, 后者 (BGL) 对糖化起关键性限速因子作用, 溶解与糖化协同作用才能完成纤维素的降解^[3,4]。现已选育到的一些较优良菌株如 *Aspergillus niger* AMS11, *Trichoderma reesei* QM9414 都具有某特定组分酶活性较低的缺陷, 如前者 EG 及 CBH 较低而 BGL 较高; 后者则反之, 因而总酶活性不高。细胞融合方法改造这些较好菌株, 是获得高产纤维素酶系菌株的重要育种途径^[1,2,5]。在广泛变异的融合子代群体中, 快速准确地分析多组分酶系的酶活性表现, 进而推测其遗传重组规律, 指导育种实践, 是值得探索的工作, 迄今有关的工作较少。本文报道运用计算机多元统计分析程序, 分析 *A. niger* AMS11 (CBH、EG 均低, BGL 高) × *T. reesei* QM 9414 (CBH、EG 均高, BGL 低) 属间融合^[5]子代群体中纤维素酶系活性表现, 指导高产菌株选育的实例。

1 数学原理与生化遗传学依据

1.1 数学方法

根据多元统计分析原理^[6,7], 设因变量 Y 随着 K 个自变量 X_1, X_2, \dots, X_k 变化并具有如下线性关系:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + \epsilon \quad (1)$$

其中 ϵ 为随机误差, 则根据 n 组观察值, 得到 n 个方程式:

$$Y_1 = b_0 + b_1X_{11} + b_2X_{21} + \dots + b_kX_{k1} + \epsilon_1$$

$$Y_2 = b_0 + b_1X_{12} + b_2X_{22} + \dots + b_kX_{k2} + \epsilon_2$$

国家自然科学基金资助课题。

* 现在通讯地址: 山东大学微生物学博士后流动站, 济南 250100。

本文于 1993 年 1 月 11 日收到。

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki} + \varepsilon_i \quad (2)$$

式中 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 是相互独立的且服从相同的分布 $N(0, \sigma^2)$ (σ 未知)。 $b_0, b_1, b_2, \dots, b_k$ 为待估的参数, 求出其最小二乘估计值, 可得回归方程:

$$Y = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_k \cdot X_k \quad (3)$$

对所得回归方程 (3) 进行方差分析、显著性检验及通径系数比较, 并据分析结果得出结论:

1.1.1 在显著性水平 α 下, 比较 $FT > F\alpha(k, n-k-1)$ 则回归方程显著, 即 Y 对 X_1, X_2, \dots, X_k 有线性回归关系。

1.1.2 一般回归方程显著, 不一定每个自变量 X 对 Y 的影响都显著。比较 $Fi > F\alpha(1, n-k-1)$, 则自变量 X_i 对 Y 的作用显著。

1.1.3 比较通径系数 $Pi \rightarrow y, Pi \rightarrow j \rightarrow y, (i, j=1, 2, \dots, k)$ 相对大小, 确定各对应自变量的相对重要性, 及对 Y 的增量效应程度。

有关线性方程组求解及统计分析的复杂计算, 我们参照有关方法建立程序在计算机上迅速完成⁽⁸⁾。

1.2 生化遗传学依据

这里指出, 就一个菌株而言 (如 *Trichoderma viride* 菌株, 三个酶组分 EG, CBH, BGL 在纤维素粉溶解和糖化两个方面均具有极显著的线性回归关系。其中 (EG+CBH) 对溶解、BGL 对产生还原糖起决定性作用⁽⁴⁾。生理上看, (EG+CBH) 活性高, 部分降解纤维素所产生纤维二糖积累, 依赖于高活性的 BGL 解除其对酶系活性的反馈抑制及对酶系诱导合成的阻遏抑制^(1,3,5)。我们认为, 高活性的 (EG+CBH) 与 BGL 的协同效应, 应该表现出高的滤纸纤维素降解活性。即滤纸酶活 (FP), 可以反映纤维素酶总活力, 体现 (EG+CBH) 与 BGL 的协同累加效应^(1,2,5)。数学上, 这一过程相当于二个自变量 X_1 (BGL), X_2 (EG+CBH) 对因变量 Y (FP) 的二元线性回归关系:

$$Y = b_0 + b_1 X_1 + b_2 X_2 \quad (4)$$

由此可以认为, 来自于同一个融合组合 *A. niger* AMS 11 (EG, CBH 均低、BGL 高) \times *T. reesei* QM 9414 (EG、CBH 均高、BGL 低) 的融合子代群体中, 由于遗传物质重组与传递具有独立随机性, 因而 n 个菌株的 (EG+CBH)、BGL、FP 组分活性的观测值, 相当于这一特定融合组合试验的 n 组观察值, 可适用于上述一般线性回归模型分析。若回归方程显著, 则表明该融合组合的确发生了遗传重组。那么在广泛变异的子代群体中一定存在真正的杂种优势。并且可以排除因非融合重组因素引起的单个或极少数高酶活变异菌株的假象干扰。

2 实例分析

我们在近 2000 个融合子代中, 选取 40 个表型较优的菌株, 参照文献 [9] 方法测定纤维素酶系组分 BGL、(EG+CBH)、FP 活性 (数据未列出), (单位: μMol 葡萄糖 / $\text{min} \cdot \text{ml}$)。运用本文方法作统计分析结果如表 1。

表 1 融合子代纤维素酶系协同作用回归模型分析

Table 1 Statistic analysis for the synergism of the cellulases from fusants

Situation	Number of strain	Regression function and F -test	F -test of partial regression coeff.	Path coeff*
Total variations of activity ($0.41 < FP < 5.08$)	40	$Y = -3.3879 + 0.2952X_1 + 0.2854X_2$ $F_{0.01}(2, 37) = 5.25$ $F = 153.5670^{**}$	$F_{0.01}(1, 37) = 7.40$ $F(1) = 8.5758^{**}$ $F(2) = 43.3778^{**}$	<pre> graph TD Y((Y)) -- 1.1718 --> X2((X2)) Y -- 0.9652 --> X1((X1)) X2 -- 0.8213 --> X1 R1((0.9971)) --> X1 </pre>
Over higher parent ($FP > 1.82$)	19	$Y = -12.1912 + 1.4452X_1 + 0.0881X_2$ $F_{0.01}(2, 16) = 6.23$ $F = 25.0368^{**}$	$F_{0.01}(1, 16) = 8.53$ $F(1) = 26.4864^{**}$ $F(2) = 1.8389$	<pre> graph TD Y((Y)) -- 0.3256 --> X2((X2)) Y -- 3.8265 --> X1((X1)) X2 -- 2.0441 --> X1 R2((0.1739)) --> X1 </pre>
Between two parents ($1.03 < FP < 1.82$)	9	$Y = -2.7248 + 0.4129X_1 + 0.0499X_2$ $F_{0.01}(2, 6) = 10.90$ $F = 69.5068^{**}$	$F_{0.05}(1, 6) = 5.99$ $F_{0.01}(1, 6) = 13.70$ $F(1) = 50.7969^{**}$ $F(2) = 9.9136^*$	<pre> graph TD Y((Y)) -- 0.2945 --> X2((X2)) Y -- 3.1323 --> X1((X1)) X2 -- 1.8991 --> X1 R3((0.1786)) --> X1 </pre>
Below lower parent ($FP < 1.03$)	12	$Y = -0.5226 + 0.0843X_1 + 0.0987X_2$ $F_{0.01}(2, 9) = 8.02$ $F = 43.9251^{**}$	$F_{0.05}(1, 9) = 5.12$ $F(1) = 2.7971$ $F(2) = 9.4574^*$	<pre> graph TD Y((Y)) -- 0.9092 --> X2((X2)) Y -- 0.7229 --> X1((X1)) X2 -- 0.6296 --> X1 R4((0.7918)) --> X1 </pre>

* Note Y: FP X₁: BGL X₂: EG+CBH

可见与亲本 AMS11 ($FP=1.03$)、QM9414 ($FP=1.82$) 相比较而言, 子代 FP 表现出三种情形: 超过高亲、低于低亲、居双亲间。进一步分析可看出: 全部 40 个子代菌株的酶活变异范围虽然较广泛 ($FP=0.41-5.08$), 但群体表现出极显著的线性回归关系。即当其它条件一定时, 每增加 1 个标准单位 (EG+CBH), 将直接增加 1.1718 个标准单位 FP, 且通过 BGL 的相关影响又可间接增加 0.8213 个标准单位 FP; 而每增加一个标准单位 BGL, 将直接增加 0.9652 个标准单位 FP, 且通过 (EG+CBH) 的相关影响, 又可间接增加 0.9971 个标准单位 FP。这里 (EG+CBH)、BGL 贡献水平均极显著。结果说明, 子代群体的确来自于融合组合 *A. niger* × *T. reesei* 的遗传重组 (即发生了融合重组), 而不是因非融合重组因素引起的个别或极少数高酶活变异的干扰菌株。可以预计群体中一定存在真正的杂种优势菌株。

在 19 个超高亲 ($FP > 1.82$) 的融合子代中, FP 与 BGL、(EG+CBH) 也存在极显著的线性回归关系。每增加 1 个标准单位 BGL, 将直接增加 3.8265 个标准单位 FP; 而每增加 1 个标准单位 (EG+CBH), 又将通过 BGL 的相关影响而间接增加 2.0441 个标准单位 FP。显然这里 BGL 贡献最重要。进一步考察这 19 个菌株 (EG+CBH)、BGL 组分活性发现, 这些菌株 (EG+CBH) ≥ 12.64 , 均接近 QM9414 的优势组分 (EG+CBH=13.91), 因此当子代 (EG+CBH) 组分活性较高时, BGL 活性能直接影响 FP 的增加量, 体现出 BGL 的关键因子作用。当子代集中 AMS11 (BGL) 及 QM9414 (EG+CBH) 杂种优势时, FP 大幅度提高。这正是我们的育种目标。结果说明 *A. niger* × *T. reesei* 融合组合进行融合育种是可行的。即使在此时 n 个菌株中未能得到最高酶活菌株, 也值得继

续重复融合实验,且预计获得满意结果是必然的。12个融合子代酶活性低于低亲 AMS11 (FP=1.03),但 FP 与 BGL, (EG+CBH) 间仍然存在极显著的线性回归关系,其中 (EG+CBH) 的影响高于 BGL 的影响 ($P_{2 \rightarrow Y} = 0.9092 > P_{1 \rightarrow 2 \rightarrow Y} = 0.7918 > P_{1 \rightarrow Y} = 0.7229 > P_{2 \rightarrow 1 \rightarrow Y} = 0.6296$)。事实上,这些菌株的 BGL 低于 AMS11,而 (EG+CBH) 则比 QM9414 的低得多,因此推断,这些菌株可能是双亲融合后负向重组的结果。

顺便指出,我们在 19 株超过高亲的融合子代中,独得了三株稳定的重组子,其纤维素酶系协同作用效果显著高于混合制曲或二次制曲的。详细的生化遗传鉴定也证实了它们的重组性(待发表)。

3 讨 论

3.1 方法的适用前提

数学上,由 n 组观察值估计回归参数,认为随机误差 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 是相互独立的且服从相同的正态分布 $N(0, \sigma^2)^{[6,7]}$ 。育种实践上,在相同的实验条件下,同一融合组合 *A. niger* × *T. reesei* 各个融合子代中,遗传物质重组与传递是独立的随机事件^[10],满足数学要求。另一方面,(EG+CBH) 与 BGL 在纤维素降解过程中的确存在协同作用的线性回归关系^[3,4]。笔者认为可以运用多元线性回归模型分析融合子代群体中 n 个菌株实验观察值来揭示特定的融合组合的遗传规律的全貌。

3.2 运用本法的优点

对于以多组分的纤维素酶系为目标的丝状真菌属间远缘融合育种而言,并非任一融合组合都能融合成功^[1,5]。首先应明确的问题是:是否真正融合了,融合子中是否一定存在杂种优势。通过分析融合子代群体中 n 个菌株酶活性观察值表现(通常具有广泛的变异性),在结果显示出极显著的协同作用线性回归关系情况下,足以指示发生了真正融合重组,且据偏回归系数显著性检验及通径系数比较,可以判定特定目标组分在增量效应上的相对贡献程度。从而排除由非融合重组因素引起的个别或极少数高酶活变异菌株的干扰。甚至即使本次融合实验未能获得高酶活性的杂种优势菌株,但只要继续融合实验,当样本数量(融合子数)足够大时,一定可获得满意的结果。这种指示,可减少盲目性,提高育种效率。

参 考 文 献

- [1] 王祖农. 山大微生物, 1986, 1: 1—14.
- [2] 赵学慧, 尹麦生. 国家自然科学基金资助项目研究成果汇编(二), 国家自然科学基金委员会综合局编, 长春: 吉林大学出版社, 1988, p. 60.
- [3] Wood T M *et al.* Biochemistry and Genetics of Cellulose Degradation. FEMS Symp. 43. London: Academic, 1988, pp. 31—52.
- [4] 高培基等. 生物工程学报, 1988, 4(4): 321—326.
- [5] 艾云灿. 中国科协首届青年学术年会湖北卫星会议论文集(武汉, 1992. 12), 湖北省科学技术委员会编, 北京: 中国科协出版社, 1993. 5(印刷中)亦可见: 论文摘要集(农科分册), 1992, p. 56.
- [6] 方开泰, 全辉. 实用回归分析, 北京: 科学出版社, 1988, pp. 64—103.
- [7] 莫惠栋. 农业试验统计, 上海: 上海科学技术出版社, 1984, pp. 449—545.
- [8] 黄因敏. ANALYST 统计分析软件包, 北京医科大学分析计算中心, 1986, p. 227.
- [9] Ghose T K. Pure and Appl Chem, 1987, 59(2): 257—268, 987.

[10] 盛祖嘉. 微生物遗传学 (第二版), 北京: 科学出版社, 1987, pp. 185—252.

The Statistic Method for Determining the Fusion-recombination of Cellulases Genome

Ai Yuncan¹ Zhao Xuehui¹ Yu Jialin²

(The Department of Food Sciences¹ and Basic Course², Huazhong Agricultural University, Wuhan 430070)

Abstract A hypothesis was proposed on the basis of cellulases synergism and multiple variable statistics that the experimental value of N strains out of total fusants from the same cell fusion pool, *Aspergillus niger* × *Trichoderma reesei*, could be a sample for the definite fusion cross due to the independence and random of DNA recombination. It could be reduced if fusion recombination happened by analyzing the sample with the provided computer programe. Forty strians out of about 2000 fusants of *A. niger* × *T. reesei* were analyzed under this procedure as an example and showed the method was effective on guiding the cellulases breeding.

Key words Cell-fusion breeding, cellulases synergism, statistic analysis, *Aspergillus niger*, *Trichoderma reesei*