

微生物诱变的数学模型 及诱变育种的设计

刘垂珩
(安徽农学院)

提 要

控制微生物数量性状的多基因系统的诱变表达,可以用混合分布来描述.群体水平上微生物诱变的数学模型的典型情况为均匀诱变分布,即两个正态分布的混合分布.

用电子计算机的蒙特卡洛模拟,对现有的三种估计方法作出了评价,发现它们都不能令人满意.提出了交互反馈最小卡方估计,这是适宜于电子计算机执行的理想的方法.参数 p (即突变率)的误差估计,可以通过误差传播公式加以测定.

微生物工程的诱变育种,可以分为三个阶段:预备试验,鉴定试验及验证试验.通常所说的“初筛”,实际上是鉴定试验.在此以前,应当安排由一系列实验组成的预备试验;验证试验则由复筛及车间对比组成.给出了一个实例.

培养条件的优化,可以采用析因回归调控的方法.给出了一个简化的实例.

一、引 言

现代微生物工程的推动力之一,就是目标产物的产生菌通过基因突变(广义上包含着重组)使其目标产物的产率持续地大幅度提高.然而,随着微生物工业发酵水平的提高,获得高产突变株的机会愈来愈小,育种的难度愈来愈大.为了克服盲目性,提高效率,除加强遗传分析,探索新的途径,改进工艺以及精确实验操作之外,针对微生物育种本身所具有的随机性,采用数理统计方法来设计实验,收集、整理、分析和解释数据——揭示规律,用以指导进一步的工作.

二、多基因系统诱变的表达

微生物工程中目标产物的产量属于数量性状,受多基因系统所控制^[1],即有核内基因,通常还有核外基因的共同作用.遗传学的基本原理指出,生物数量性状的表现型值是基因型值及环境值的共同结果.除去外部条件的大环境(macroenvironment)效应外,生物体内的微环境(microenvironment)效应造成了在相同的外部条件下个体发育的参差性.微生物除遵循这些共同的原则外,还有着其独特之处.微生物在摇瓶或发酵罐中

1984年8月17日收到,1985年2月25日收到修改稿

本研究课题曾获农牧渔业部颁发的1982年农牧渔业技术改进一等奖.

是一个有生有灭的群体,所观测到的目标产物的产量是这个多世代群体的表现;并且由于空间及资源有限,这个多世代群体的发育必将影响其赖以生存的外部条件,而外部条件的变化又会影响群体的继续发育。多世代群体发育及其与环境的交互作用,是微生物数量性状的一个突出的特点。另外,摇瓶的振荡及发酵罐的搅拌,它们是微生物工艺学所决定的,但在数学工作者的眼光下,却无疑在起着均匀混合的“平均”作用。综上所述,按照概率论中的中心极限定理,可以得出结论:一般微生物菌株在相同的环境下,许多子代摇瓶(或发酵罐)目标产物发酵单位的分布(称为微生物菌株的分布)近似于正态分布。广泛的实践证实了这一结论^[9]。

微生物自发突变的随机性,早已由 Luria 等^[8]、Newcombe^[9] 及 Lederberg 等^[10] 人所证实。诱发突变的随机性则可从分子生物学的进展得到肯定。譬如,紫外线诱变的主要机制是使 DNA 中两个邻位或斜对位的胸腺嘧啶聚合成二聚体,但究竟哪一对胸腺嘧啶发生聚合却是随机的;又如,5-溴尿嘧啶可以呈酮式代替胸腺嘧啶掺入 DNA 中与腺嘌呤配对,而在复制过程中又可以呈烯醇式与鸟嘌呤配对而发生突变,但究竟 5-溴尿嘧啶掺入 DNA 中哪一个位置却是随机的。

长期以来,微生物学中对突变的研究大多基于对特定的表现型的直接计数。但基因突变对于表现型的变化,既不是充分条件也不是必要条件。特别是生物工程着眼于目标产物的产量,而多基因系统的突变型是无法象对待营养缺陷型那样用选择性培养基来加以分辨的。

对于微生物诱变,把接受诱变处理的所有个体(孢子或细胞)看作是一个总体。不加诱变其分布一般应近似于正态。诱变后,这个总体将被分解成为三个部分:一部分(D)在诱变剂的作用下致死,另一部(R)虽经诱变剂处理却最终没有发生突变而保持着原来的基因型存活下来,最后一部分(M)则是存活的突变型。如果用 $P(D)$ 、 $P(R)$ 、 $P(M)$ 分别表示被处理个体在诱变后致死、未突变及突变的概率,则有

$$P(D) + P(R) + P(M) = 1. \quad (2-1)$$

致死率及诱变效应的可重现性,使得我们有证据去推断 $P(D)$ 、 $P(R)$ 及 $P(M)$ 在一般微生物诱变试验中是取常数值;这样的诱变称为均匀诱变。

诱变后存活个体目标产物产量所确定的分布称为诱变分布。显然,诱变分布应当是未突变分布及突变分布的混合分布,两个支分布混合时的比例因子恰是突变率。

不妨假设出发菌株近似地服从正态分布 $N(\mu_r, \sigma_r^2)$ 。由于 $P(R)$ 为常数,故未突变分布应与出发菌株的分布相同,具有分布函数

$$R(x) = \int_{-\infty}^x r(t) dt = \int_{-\infty}^x \frac{1}{\sigma_r \sqrt{2\pi}} \exp \left\{ -\frac{(t-\mu_r)^2}{2\sigma_r^2} \right\} dt \quad (2-2)$$

由于 $P(M)$ 也是常数,再加所观测到的突变型目标产物的产量受着多基因系统控制,并且又是多世代群体的表现,因此,简单而合理的假设是:突变分布近似地服从另一正态分布

$$M(x) = \int_{-\infty}^x m(t) dt = \int_{-\infty}^x \frac{1}{\sigma_m \sqrt{2\pi}} \exp \left\{ -\frac{(t-\mu_m)^2}{2\sigma_m^2} \right\} dt \quad (2-3)$$

从而,均匀诱变分布函数 $F(x)$ 具有两个正态分布的混合分布形式(其中 p 即为突变率)

$$F(x) = \int_{-\infty}^x f(t) dt = (1-p)R(x) + pM(x) \\ = \int_{-\infty}^x \left[\frac{1-p}{\sigma_r \sqrt{2\pi}} \exp \left\{ -\frac{(t-\mu_r)^2}{2\sigma_r^2} \right\} + \frac{p}{\sigma_m \sqrt{2\pi}} \exp \left\{ -\frac{(t-\mu_m)^2}{2\sigma_m^2} \right\} \right] dt. \quad (2-4)$$

大量微生物诱变试验的结果证实了均匀诱变分布的这种特性^[3]。

三、交互反馈估计

群体水平上微生物诱变的数学模型,一般可以归结为两个正态分布的混合分布。后者是一类理论上棘手但应用上重要的问题。文献[3]提出了条件矩估计、条件最小 χ^2 估计及信息条件最小 χ^2 估计,但其优劣却未有评价。我们利用电子计算机的 Monte Carlo 模拟作了这项研究。运用均匀变量构成正态变量的经典方法^[6],在电子计算机上随机地产生 189 个 $N(103.72, 42.09)$ 的样品及 311 个 $N(122.99, 40.94)$ 的样品,从而组成一个容量为 500 的混合样本,对每个样本都作出上述三种估计。把所产生的 50 个样本,按产生的顺序,每十个作为一组,分为五组,分别就估计的参数予以统计,再把 50 个样本合并加以统计。结果列于表 3-1。

表 3-1 三种估计方法的评价

方法	参 数		第一组	第二组	第三组	第四组	第五组	五组合计	期 望
信息最小 χ^2 条件估计	p	均 值	0.6186	0.6237	0.6194	0.6211	0.6231	0.6212	0.6220
		标准误差	0.0025	0.0049	0.0078	0.0034	0.0063	0.0023	
	μ_2	均 值	123.21	122.96	123.15	123.07	123.01	123.06	122.99
		标准误差	0.1306	0.1632	0.1910	0.1565	0.0604	0.0699	
条件矩估计	p	均 值	0.5876	0.5790	0.5718	0.5731	0.5908	0.5805	0.6220
		标准误差	0.0208	0.0180	0.0091	0.0205	0.0172	0.0078	
	μ_1	均 值	104.67	104.71	104.75	104.89	104.31	104.67	103.72
		标准误差	0.5063	0.4185	0.2805	0.4925	0.4079	0.1869	
	μ_2	均 值	123.40	123.67	123.93	123.67	123.63	123.66	122.99
		标准误差	0.2748	0.2801	0.1439	0.2976	0.3587	0.1214	
条件最小 χ^2 估计	p	均 值	0.6079	0.6050	0.6050	0.6062	0.6072	0.6063	0.6220
		标准误差	0.0017	0.0012	0.0011	0.0013	0.0016	0.0006	
	μ_1	均 值	104.22	104.11	103.97	104.17	103.96	104.09	103.72
		标准误差	0.1230	0.0740	0.1406	0.0725	0.0909	0.0475	
	μ_2	均 值	123.13	123.29	123.57	123.30	123.35	123.33	122.99
		标准误差	0.0976	0.1094	0.1522	0.1168	0.1444	0.0520	

对于表 3-1 所列三种估计的“五组合计”项参数 p, 可以用 SNK 法^[7]对估计值作多重比较, 见表 3-2。其它四个参数, 也有类似结果。

模拟的结果说明:

(1) 信息条件最小 χ^2 估计未发现偏, 显著优于其它两种方法, 但此法必须依赖于

表 3-2 参数 p 三种估计的多重比较

方 法	矩估计	χ^2 估计	信息估计	比较范围	显著范围	极显著范围
均 值 \bar{p}	0.5805	0.6063	0.6212	R	90.05%	90.01%
\bar{p} 的差值		0.0258**	0.0149*	2	0.0131	0.0173
			0.0407**	3	0.0157	0.0196

* 表示 0.05 水平显著.

** 表示 0.01 水平显著.

精确的前馈信息.

(2) 条件矩估计有显著的偏离,且稳定性劣于其它两种方法,不宜采用.

(3) 条件最小 χ^2 估计是有偏的,但优于条件矩估计,其稳定性则优于其它两种方法.

由于上述三种方法都不够理想,故而有必要建立一种无偏、稳定而又不必依赖前馈信息的方法.交互反馈最小 χ^2 估计就是这样一种方法,其具体步骤为:

(1) 按照条件最小 χ^2 估计,在

$$\min \chi^2(t|v_1, v_2, v_3, v_4) \quad (3-1)$$

时得到一组估计值 $p_1, \bar{x}_{(1,1)}, s_{(1,1)}^2, \bar{x}_{(2,1)}, s_{(2,1)}^2$, 其中 v_1, v_2, v_3, v_4 是诱变样本的前四阶矩.

(2) 把 $\bar{x}_{(1,1)}$ 及 $s_{(1,1)}^2$ (或者 $\bar{x}_{(2,1)}$ 及 $s_{(2,1)}^2$) 作为前馈信息,按照信息条件最小 χ^2 估计,在

$$\min \chi^2(p|\bar{x}_{(1,1)}, s_{(1,1)}^2, v_1, v_2) \quad (3-2)$$

(或者 $\min \chi^2(p|\bar{x}_{(2,1)}, s_{(2,1)}^2, v_1, v_2)$) 时,得到一组估计值 $p_2, \bar{x}_{(1,2)}, s_{(1,2)}^2, \bar{x}_{(2,2)}, s_{(2,2)}^2$ (或者 $p_2, \bar{x}_{(1,2)}, s_{(1,2)}^2, \bar{x}_{(2,1)}, s_{(2,1)}^2$).

(3) 再把 $\bar{x}_{(2,2)}$ 及 $s_{(2,2)}^2$ (或者 $\bar{x}_{(1,2)}$ 及 $s_{(1,2)}^2$) 作为前馈信息,按照信息条件最小 χ^2 估计在

$$\min \chi^2(p|\bar{x}_{(2,2)}, s_{(2,2)}^2, v_1, v_2) \quad (3-3)$$

(或者 $\min \chi^2(p|\bar{x}_{(1,2)}, s_{(1,2)}^2, v_1, v_2)$) 时得到一组估计值 $p_3, \bar{x}_{(1,3)}, s_{(1,3)}^2, \bar{x}_{(2,3)}, s_{(2,3)}^2$.

(4) 重复(2)(3)两个步骤,直到相邻两次迭代所得 χ^2 值的差值在绝对值上小于预先给定的精度时,终止迭代.

进一步的 Monte Carlo 模拟研究得到以下一些主要结果:

(1) 按前述方法在电子计算机上产生了十个混合样本,参数 p 的条件最小 χ^2 估计值的均值为 0.6053,对应的标准误差为 0.0014,由于 p 的期望为 0.6220,仍然是有偏的;但相应的交互反馈最小 χ^2 估计值的均值为 0.6209,标准误差为 0.0015,未发现有偏,稳定性一致.其它四个参数也有类似结果.

(2) 首次反馈信息取 $\bar{x}_{(1,1)}$ 及 $s_{(1,1)}^2$,与取 $\bar{x}_{(2,1)}$ 及 $s_{(2,1)}^2$ 所得结果一致,见于表 3-3.

这是对同一样本作估计的结果,微小的出入是由于把 χ^2 差值的精度预定为小于 0.01 所致.

(3) 交互反馈最小 χ^2 估计可以显著地提高符合度(fitness),实例见于表 3-4.

综上所述,交互反馈最小 χ^2 估计是一种适宜于在电子计算机执行的较为理想的方法.

表 3-3 首次反馈信息的影响

首次反馈信息	交互反馈最小 χ^2 估计值					
	p	μ_1	σ_1^2	μ_2	σ_2^2	χ^2
$\bar{x}_{(1,1)}, s_{(1,1)}^2$	0.6262	104.19	40.97	122.79	47.22	1.7445
$\bar{x}_{(2,1)}, s_{(2,1)}^2$	0.6255	104.20	41.05	122.80	46.89	1.7439

表 3-4 交互反馈对符合度的影响

	p	μ_1	σ_1^2	μ_2	σ_2^2	χ^2	df	P
反馈前	0.6175	103.80	35.10	123.29	38.76	5.380	3	0.154
反馈后	0.6267	104.05	39.88	122.88	46.95	1.291	3	0.732

参数 $\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$ 的估计值的误差, 可以从相应的支分布得出, 这里需要研究的是参数 p 的估计值的误差.

令 \bar{x}_1 及 \bar{x}_2 分别表示来自混合样本的 μ_1 及 μ_2 的估计值, \bar{x}_3 表示混合样本的均值; s_1^2 及 s_2^2 分别表示来自混合样本的 σ_1^2 及 σ_2^2 的估计值, s_3^2 表示混合样本的方差. 由于

$$\mu_3 = (1-p)\mu_1 + p\mu_2 \quad (3-4)$$

从而即有

$$\hat{p} = \frac{\bar{x}_3 - \bar{x}_1}{\bar{x}_2 - \bar{x}_1} \quad (3-5)$$

显然, 只要

$$|\mu_2 - \mu_1| > \max(\sigma_1, \sigma_2) \quad (3-6)$$

(对于微生物即为诱变效应显著), 就有

$$\left| \frac{\partial \hat{p}}{\partial x_i} \right| \gg \left| \sum_{n=1}^{\infty} \frac{\partial^{n+1} \hat{p}}{\partial x_i^{n+1}} \sigma_n(\bar{x}_i) \right| \quad i=1, 2, 3. \quad (3-7)$$

故 p 的方差可用误差传播公式来加以近似估计

$$\begin{aligned} \hat{\sigma}^2(p) &= \sum_{i=1}^3 \left(\frac{\partial \hat{p}}{\partial x_i} \right)^2 \hat{\sigma}^2(\bar{x}_i) \\ &= \frac{(\bar{x}_3 - \bar{x}_2)^2}{(\bar{x}_2 - \bar{x}_1)^4} s_{\bar{x}_1}^2 + \frac{(\bar{x}_3 - \bar{x}_1)^2}{(\bar{x}_2 - \bar{x}_1)^2} s_{\bar{x}_2}^2 + \frac{1}{(\bar{x}_2 - \bar{x}_1)^2} s_{\bar{x}_3}^2. \end{aligned} \quad (3-8)$$

对于表 3-1 所列数据“五组合计”项有 $\bar{x}_3 = 115.71$, 恰与其期望 μ_3 吻合, $s_{\bar{x}_3} = 0.0402$. 按式 (3-8) 可得“五组合计”项 p 的均值的标准误差分别为 0.0034 (信息条件最小 χ^2 估计)、0.0059 (条件矩估计) 及 0.0028 (条件最小 χ^2 估计), 而模拟所得的对应标准误差分别为 0.0023、0.0078 及 0.0006. 从而可见, 按式 (3-8) 去估计 p 的误差是可行的.

对于微生物诱变, 即使是非均匀的, 其诱变分布仍可用混合分布来描述; 而对一般的混合分布, 只要其两个支分布都有前两阶矩存在, 就可以用交互反馈最小 χ^2 法来作参数估计.

四、诱变育种的设计分析

微生物诱变育种试验, 一般应当分为三个阶段: 预备试验, 鉴定试验及验证试验. 下面即以龟裂链霉菌 (*Streptomyces rimosus*) 诱变育种的实例来加以说明.

微生物诱变育种中,通常所说的“初筛”,实际上是鉴定试验,在此之前应当先做一系列的预备试验。预备试验可由若干小而灵活的实验构成,每次实验应当只有一两个明确的目的。一般,预备试验的目标应包括:(1)出发菌株的遗传变异能力得到估计,证实其适宜作为出发菌株;(2)诱变剂的效应得到估计,证实其确有产生高产突变株的可能;(3)所选用的环境条件确实能使高产突变株得到表达。

通过遗传背景分析得知,从原始菌株 *S. rimosus* T1001 到所选用的出发菌株 *S. rimosus* 151 (华北制药厂提供) 已经历 44 次诱变或自然分离。其中用得最多的诱变因素是烷化剂,共 13 次;氯化锂只在中期用过 4 次,并且都是作为后处理因素与其它因素复合使用的。鉴于氯化锂前处理增强射线诱变作用已有报道^[5],故而选用了氯化锂前处理与紫外线复合作为诱变因素。

第一次预备试验估计了氯化锂前处理后紫外线不同照射时间 *S. rimosus* 151 的致死率,结果列于表 4-1。由于前处理后已有半数左右的孢子死亡,故而选择了总致死率为 96.9% 的 60 秒钟为照射时间。

表 4-1 致死率的估计

氯化锂前处理后紫外线照射时间(s)	0	30	60	90
致死率 (%)	54.5	90.3	96.9	~100

第二次预备试验的目的是在检验所选定的诱变因素的有效性的同时估计出发菌株的遗传变异能力,其结果列于表 4-2。

表 4-2 诱变效应及变异能力

参 数	项 目		
	不 诱 变	后 处 理	前 处 理
平均效价 \bar{x}	100	101.2	113.2
标准差 s	6.18	11.55	10.42
样本容量 n	18	14	16
稳度 ^[2] G_s	16.18	8.76	10.86

从表 4-2 的第一列数据,可以对出发菌株的遗传变异能力作出估计。在预备试验的条件下,出发菌株的稳度达 16 左右,证实 *S. rimosus* 151 是一株高产稳产的菌株,适宜作为诱变育种的出发菌株。从表 4-2 的前三行数据,可以用 SNK 法对三个均值作多重比较而对诱变效应作出估计,从而可以有足够的把握去推断:“前处理”的平均效价高于其它两者,所选定的诱变因素是有效的。

按文献[4]所述,选定突变率 $p=0.5$, $P_0=0.023$ (即预备试验中发酵摇瓶效价超过 112.4 的频率), $P_1=0.08$ (即当诱变后平均效价能被接受为显著差异时摇瓶效价超过 112.4 的概率), $\alpha=0.01$ (即未发生高产突变而被接受的风险), $\beta=0.05$ (即已发生高产突变而遭拒绝的风险),即可算出:诱变后至少应挑取 213 个单孢子菌落。

实际上共做了 323 个单孢子菌落的发酵摇瓶,并同时做了 41 个出发菌株的对照摇瓶。用交互反馈最小 χ^2 估计可得 $\hat{\mu}_r=103.72$, $\hat{\sigma}_r^2=42.09$, $\hat{\mu}_m=122.99$, $\hat{\sigma}_m^2=40.94$,

$\hat{p}=0.6220$ 。从而可见, 诱变的突变率略高于 0.60, 突变均值较出发菌株的均值有极显著的提高, 突变方差与未突变方差均较小, 突变型与出发菌株的基因型很少有可能混淆起来。可以推断: 诱变效应良好。由于诱变是成功的, 因而确定了一个较高的“门槛”: 从前述 323 个参试菌株中选出 16 株摇瓶效价超过 128 的菌株作为入选株。这个过程就是鉴定试验——初筛。对于整个育种工作来说, 鉴定试验是关键环节。

鉴于初筛入选株较多, 为了在同一环境下作比较, 就把全部初筛入选株在出发菌株的对照下作半复筛。通过半复筛, 从 16 株初筛入选株中挑出了其平均效价超出半复筛中出发菌株平均效价一成以上的 6 株参加复筛。

对 6 个参加复筛的菌株, 在出发菌株的对照下, 就它们的平均效价作序贯 t 检验, 其原假设为 $\mu_1=\mu_2$, 备择假设为 $1.1\mu_1\leq\mu_2$ (即可望突变型较出发菌株增产一成以上), 第一类统计学错误的风险取定为 $\alpha=0.05$, 第二类统计学错误的风险取定为 $\beta=0.25$ 。本来计划逐批调整累计, 但第一个批次就得出了接受备择假设的推断; 为慎重起见, 还是再做了两个批次, 仍然单独地就可以得出接受备择假设的推断。编号 99 的突变株在三个批次中表现最好, 因而入选进入下一阶段的试验——车间对比。

车间对比, 是投产前的最后的也是最重要的一个步骤, 它和复筛构成了验证试验。在经过试生产证实不会因采用编号 99 菌株而招致重大经济损失以后, 安排了“同步对比”的中间试验性生产检验, 即在同一车间由同样的操作人员在同样的设备上按原来的工艺把选育株与出发株随机地安排罐批次序投入车间生产 (采取“同步对比”的理由是诱变效应显著)。对比的结果见于表 4-3。

表 4-3 “同步对比”结果

菌株编号	生产罐数 n	平均放罐效价 \bar{x}	平均放罐效价的标准误差 s, \sqrt{n}
151	12	100	0.90
99	9	115.68	4.22

从表 3-3 可见, 经过 21 个罐批的车间对比, 即已得出接受备择假设的推断, 可以预期编号 99 菌株投产后较 151 菌株增产一成以上。

编号 99 菌株正式投产后, 实际增产达 20.6%。

这次试验由于事先有比较完善的设计, 每做一个实验都有明确的目的, 通过分析而得到的推断通常要比“经验”更有把握, 故整个试验 (包括车间对比) 只用了不到半年的时间, 育种效率大为提高。

五、培养条件的优化

在微生物工程中, 培养条件所涉及的因素很多, 是一个复杂的多因素优化问题。由于所观测到的表现型值是基因型与环境的共同结果, 培养条件的优化就是使环境特别是环境与基因型的交互作用取得尽可能大的正效应。

微生物工程目标产物的产量一般呈非线性变化, 优化的目的又在于寻求使得产量这

个数量性状达到极大值的培养条件,这样就使得一些常用方法,例如逐步回归,难以奏效。但微生物实验可以较均匀地控制条件而在短期内获得大量信息,却是一个难能可贵的优点。从微生物工程的实际出发,作者提出了析因回归调控的方法来优化培养条件。

“析因”指的是析因设计^[7]。根据目标产物的代谢途径及生产经验,选定若干因素及每一因素的若干水平而设计出一项析因试验,以验证这些因素的主效应及交互效应。

“回归”指的是回归分析。当所选因素在所选水平上通过析因试验证实它们的主效应或交互效应确实对目标产物的产量有显著影响后,采用回归分析的方法建立起回归方程(一般以多元二次多项式为首选模型)。

“调控”指的是在析因及回归的基础上进行调整和控制。有了回归方程,就可以作出预测。由于预测的结果仍然具有随机性,即以一定的概率水平出现在某个范围内,故而对预测的结果不能无条件地接受。如果预测极值点并非实验点所能表述,或预测结果不能接受,则应再设计另一项析因试验,力求把这项试验的中心向量取得与极值点一致,并再作一次回归分析。直到预测结果与实验结果相符,就可以把所得结果作为控制条件,并用在这个控制条件下的生产状况作为反馈信息来进一步调整控制条件。

根据作者的实践,在目前的实验条件下,所选定的因素以一次不超过五个为宜,一般不超过三次循环就可以得到比较满意的结果。在所研究因素相当多时,可以采用交互反馈优化的方法来进行析因回归调控。

根据作者的实践,温度、pH及接种量等条件一经优化就比较稳定,然后应当把重点放在碳、氮、磷等主要元素上,其它元素可再放后一步。当然,这是就一般而论,对于具体的目标产物更重要的是应作具体分析。

考虑到篇幅,下面仅就豆饼粉及饴糖两个因素对生米加链霉菌(*Streptomyces mycarofaciens*)某菌株的代谢产物的效价的影响为例来加以说明。实验数据载于表5-1。

表5-1 3×3析因试验的观测数据

因 素		B ₂ 饴 糖 (%)		
		4	6	8
		效 价		
A :	2	71.3	95.0	69.4
		87.2	77.5	71.0
		78.8	80.8	73.4
		81.3	88.3	77.9
豆 饼 粉	3	76.7	107.4	83.3
		86.7	101.6	87.5
		92.5	97.4	81.7
		92.5	104.9	90.8
C %	4	80.0	70.5	67.0
		85.8	90.5	77.8
		82.5	68.0	74.5
		73.3	78.0	68.7

表5-2给出了对表5-1数据作析因分析而得到的结果。

表 5-2 表 4-1 数据的析因分析

变异来源	自由度	平方和	均方	F
A	2	1633.95	816.98	20.14
B	2	781.35	390.68	9.63
AB	4	389.68	97.42	2.40
误差	27	1095.33	40.57	
总和	35	3900.31		

从表 5-2 可见, 因素 A(豆饼粉)及因素 B(饴糖)在所试验的水平上, 对目标产物的效价均有极显著的效应. 另外, 从表 4-2 中的误差项均方 (MS_E) 还可以得到一项公共标准误差 (r 表示重复度, 本试验中 $r=4$).

$$\sqrt{\frac{MS_E}{r}} = \sqrt{\frac{40.57}{4}} = 3.18. \quad (5-1)$$

设 X_1 表示豆饼粉的百分含量, X_2 表示饴糖的百分含量, Y 表示目标产物的效价, 从表 5-1 所列数据就可以得到回归方程

$$Y = -98.1049 + 84.1647X_1 + 26.7616X_2 - 14.063X_1^2 - 0.20975X_1X_2 - 2.16875X_2^2. \quad (5-2)$$

进而可以从式(5-2)得到预测极值点

$$(X_1=2.95, X_2=6.03) \quad (5-3)$$

预测极值点与试验的中心 ($X_1=3, X_2=6$) 相近证实了试验设计是成功的. 由于预测极大值的近似值

$$E(Y | X_1=3, X_2=6) = 106.54 \quad (5-4)$$

与在 3% 豆饼粉及 6% 饴糖下的实际观测目标产物的平均效价 102.825 间的差异在一个标准误差左右, 预测结果与实验结果相容. 因此, 对所试验的菌株, 仅就豆饼粉及饴糖两个因素而言, 其优化条件可控制为 3% 豆饼粉及 6% 饴糖.

作者对中国科学技术大学杨纪珂教授的指导谨致谢忱.

参 考 文 献

- [1] 刘垂珩, 自然辩证法通讯 1, 46~52 (1984)
- [2] 刘垂珩, 微生物学报 20, 2, 166~172 (1980)
- [3] 刘垂珩, 中国科学, 7, 625~633 (1983)
- [4] 刘垂珩, 科学通报, 19, 1199~1203 (1981)
- [5] 周琪文等, 抗生素, 6, 8~11 (1979)
- [6] 李锡碚, 实验的数学处理, 1983, 科学出版社: 305
- [7] 斯蒂尔等, 杨纪珂等译数理统计的原理和方法, 科学出版社, 138、238 (1979)
- [8] Luria, S. E., Delbrück, M., Genetics, 28, 491 (1943)
- [9] Newcombe, H. B., Nature, 164, 150 (1949)
- [10] Lederberg, J., Lederberg, E. M., Jour. Bacteriol, 63, 399 (1952)

A MATHEMATICAL MODEL FOR MICROBIAL MUTAGENESIS AND THE DESIGN OF MUTAGENIC BREEDING

LIU CHUIYU

(Anhui Agricultural College)

Abstract

The expression of mutagenesis on polygenic system which controls a quantitative character of microbe can be represented by a mixed distribution. A mathematical model for microbial mutagenesis on populational level is a uniform-mutagenic distribution (i. e. a mixed distribution of two normal distributions).

An evaluation on three estimating procedures has been made by an electronic computer with Monte Carlo Simulation. It is found that all of these procedures are not satisfactory. A procedure called Interfeedback Minimum Chi-Square Estimation is proposed which is found to be a satisfactory one and is suitable for executing at electronic computers. An estimate of error for parameter p (i. e. mutation rate) can be measured through the formula of error propagation. A mutagenic breeding in microbial engineering is divided into three stages: a preparative experiment, an appraisal experiment and a checking experiment. A primary screening, as it is generally called, is actually an appraisal experiment. Before the appraisal experiment, a preparative experiment composed of a series of trials should be taken. A checking experiment is composed of a secondary screening and a plant contrast. An illustration is given.

An optimisation technique called Factorisation-Regression-Modulation-Control is recommended to optimise the circumstances of microbial culture. A simplified illustration is given.