

菌落形态的计算机识别法用于菌种的分离筛选*

阳 葵 王积分¹ 王福东 段世铎 张 臻²

(天津大学化学系 天津 300072)

(天津大学计算机科学与工程系 天津 300072)¹

(天津大学化工学院 天津 300072)²

摘要: 菌落形态是鉴别和分类菌种的重要特征之一。以分形和多重分形理论为基础,以计算机图像识别技术为手段,考察霉菌(绿僵菌)菌落形态的定量描述,分别测定各菌落样本的分形特征(覆盖维)和多重分形特征(多重分形谱)。研究表明,多重分形特征与菌种性能的相关性更大。以多重分形特征 α -right, α -width 和 f -start 为依据设计的分类器,可以用于优良菌种的自动识别,速度快,与人工分离筛选的实验数据相吻合。

关键词: 菌落,形态,菌种筛选,图像识别,多重分形

中图分类号: Q934 **文献标识码:** B **文章编号:** 0253-2654(2000)-01-051-05

SCREENING FOR STRAINS BY COLONY AUTOMATIC RECOGNITION METHOD*

YANG Kui, WANG Ji-Fen¹, WANG Fu-Dong, DUAN Shi-Duo, ZHANG Liu²

(Department of Chemistry, Department of Compute Science and Engineering¹, Institute of Chemical Engineering², Tianjin University, Tianjin 300072)

Abstract: The morphology of microorganism colony is proven to be an important feature to classify and screen for strains. Based on Fractal and Multifractal Theory, the quantitative description of the morphology images of mold (*Metarhizium* sp.) colonies was investigated by means of Computer Image Analysis Technique in this paper. The fractal features (box dimension) and multifractal features (multifractal spectrum) were measured for each sample. The multifractal characteristics extracted were proven to be more related with the biotransformation activities of the strains. Moreover, we showed that the classifier designed depending on multifractal spectrum characteristics α -right, α -width and f -start, can be used for automated screening for fine strains. The result is found to be in good agreement with the experimental data of artificial screening with the much faster screening speed.

Key words: Colony, Morphology, Screening for strains, Computer image analysis, Multifractal

* 国家自然科学基金(No.29606008)

Project of Chinese National Natural Science Fund for Young Researchers (No.29606008)

天津市21世纪青年科学基金资助项目(No.953709211)

Project of the 21st Century Natural Science Fund of Tianjin for Young Researchers (No.953709211)

收稿日期:1998-09-10 修回日期:1998-12-31

利用微生物菌种(如绿僵菌 *Metarhizium* sp.)产生的 11 α -羟化酶,催化甾体底物 16 α , 17 α -环氧黄体酮(16 α , 17 α -epoxy-4-pregnene-3,20-dione)羟化转化为产物 11 α -羟基-16 α , 17 α -环氧黄体酮,是当前甾体激素类药物人工合成中最重要和常用的微生物转化反应之一。选用性能优良的菌种是转化过程卓有成效进行的基础^[1],但菌种的优良特性常常会在传代中因变异而丧失,一般该生产菌株经一定时间(约 1~2 月)使用后须分离纯化。而传统的人工筛选法,周期长、效率低,选择范围也很有限,需要设计一种更为有效的筛选方法。

菌种筛选工作的经验表明,菌株的高产性能与其生物学特性有一定关系。菌落形态是重要的生物学特性之一。在一幅微生物显微图象中,可以看到诸如:透明与暗淡、粗糙与细腻、致密与稀疏、凹陷与隆起、丰满与萎缩等表现的或隐含的形态特征。就菌落而言,形态特征包括:(1)大小,(2)形状,(3)隆起状态,(4)表面状态(光滑、皱褶、粗糙等),(5)边缘情况,(6)表面光泽、透明程度、颜色等方面^[2]。

形态研究方法有两种:(1)直观观测,包括肉眼观察与显微镜观测。(2)计算机识别,用计算机视觉替代人的视觉,量化菌落形态特征,并与其性能相关联,有可能成为缩短分离筛选周期、提高效率与精度的有效

途径之一。

1 材料与方法

1.1 菌种及性能评价条件

1.1.1 菌种:霉菌菌株(绿僵菌 *Metarhizium* sp.),由东北制药总厂提供。

1.1.2 斜面与摇瓶发酵培养基:蔗糖 6g,酵母膏 0.25g,定容 100mL, NaNO₃、KCl、MgSO₄等无机盐适量。发酵培养基调 pH6.5,斜面培养基 pH5.6。

1.1.3 斜面菌种培养条件:以菌悬液接种至斜面培养基,25℃ 恒温培养 5~6d,至长成成熟菌种斜面。

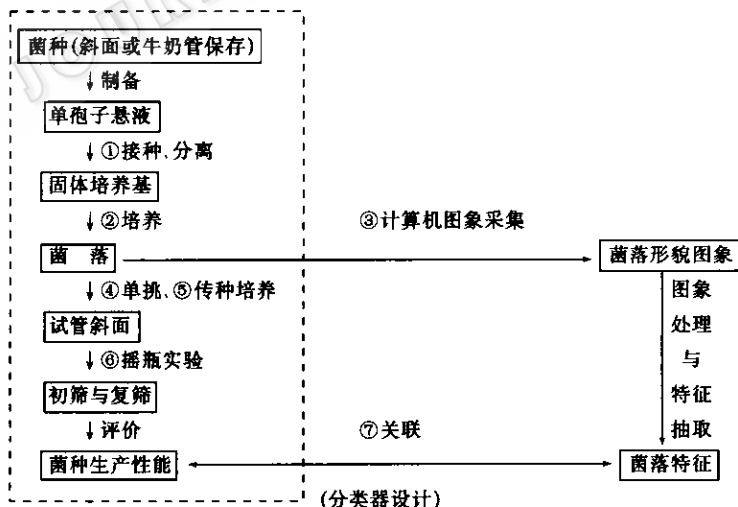
1.1.4 摇瓶评价实验条件:适量斜面菌种接种至摇瓶培养基,25℃ 恒温培养 48h,投入反应底物(投料量 2g / 1000mL 发酵液),30℃ 恒温转化 48h,取样分析转化率。

转化率分析采用硅胶 GF-254 薄板层析法。用三氯甲烷萃取发酵液中的甾体(包括底物与产物),取少量萃取液点板、层析(层析液为苯、丙酮混合液,体积比 8:2),晾干后置于 254nm 紫外分析仪中与标准样板对照分析底物转化率。

1.2 计算机图象识别系统^[3]

1.3 实验流程

实验流程如框图所示(左面虚框为常规人工分离筛选的实验步骤):



2 结果与讨论

2.1 菌落图象获取

菌种单孢子悬液分散接种至固体平板培养基上,至长出肉眼可见的单个菌落后,经生物显微镜放大 50 倍,通过 CCD 摄像机和 CA540 图象采集卡,将菌落图象转化

为 512 × 512 × 8 bit 格式的数字化图象,录入计算机。

2.2 图象前处理

(1) 实验观察发现,菌落不只是平面蔓延生长,而是呈现“山丘”状隆起的立体形态。通过微调焦距分层拍摄的方法获取了菌落底部、隆起部分中部以及顶部

三级层次的序列图象(如图 1 所示)。实验结果表明,①由一级层次(即菌落底部)图象解析得到的菌落形态特征与菌种性能的对应关系不明显(见表 1);②由二级层次(即中间隆起部分)图象解析得到的形态特征与菌种优劣的对应关系较好,较大的形态特征值对应较高的菌种活性(见表 1);③三级层次(即菌落顶部)图象清晰

度较差,易导致所获取信息的失真,不宜采用(见图 1)。

(2) 实验观察还发现,菌落的生长不是匀速的而是近似于指数过程。前期以平面铺展生长为主,后期生长速度加快,以垂直隆起生长为主。实验中获取了不同生长阶段菌落的时间序列图象。实验证明,试管斜面菌种经 88h 培养、牛奶管菌种经 106h 培养后所得的菌落

表1 菌落形态特征(覆盖维)与菌种性能的关系

菌落编号	第一级图象特征 D_1	第二级图象特征 D_2	菌种性能评价(转化率)
1	2.0946	2.0783	82%
2	2.0891	2.0881	82%
3	2.0926	2.0880	80%
4	2.0640	2.0820	74%
5	2.0960	2.0640	68%
6	2.0822	2.0513	65%
7	2.0840	2.0500	64%

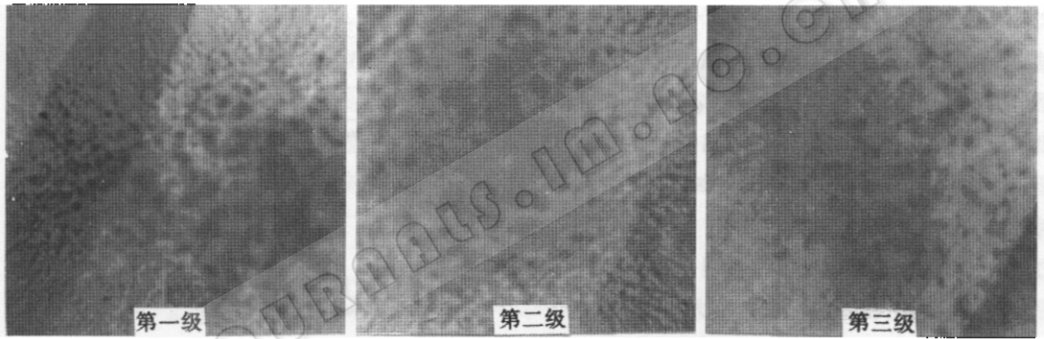


图1 菌落的三级形貌图象

图象最宜于识别且与菌种性能的关联性最优。

(3) 菌落的隆起层次、隆起形状和表面状态等形态差异,是与菌种性能相关的有效形态信息。纹理分割是可实现分割图象为若干有意义的区域。本实验中将每幅菌落图象分割为平面生长部分和隆起部分。着重分析处理隆起部分的形态特征。

2.3 特征抽取、识别与分类

菌落的形成经历了小全息元(如菌种细胞或孢子)通过分裂、数目增加到大全息元,向无穷集团(菌落)逼近的生长过程,符合分形(fractal)结构。其结构特征可用分形维数来定量表征^[4]。

(1) 覆盖维的计算:覆盖维算法应用广泛^[5],且易于实现。对于不规则表面 φ 的分维估算,首先考虑用厚度为 ε 的“毯子”分别覆盖在 φ 的上下表面,使在三维空间中所有到 φ 的距离小于 ε 的点均被包含在其中,表面

φ 的面积可近似由覆盖后所形成的体积除以 2ε 得到,从而可以估算覆盖维值:

$$D = 3 - \log[V(\varepsilon_1)/V(\varepsilon_2)]/\log(\varepsilon_1/\varepsilon_2)$$

式中 $V(\varepsilon_1)$ 、 $V(\varepsilon_2)$ 分别为对应 ε_1 、 ε_2 的体积。由覆盖维的定义和计算方法的基本原理可见,覆盖维值可以较好地反映二维图象表面隆起和起伏程度的信息,用于量化菌落形态是可行的。实验结果(如表 1 和文献 [3])也表明,由菌落形貌图象解析所得的覆盖维特征,在一定程度上与菌种性能的优劣相对应。

但实验中发现,覆盖维的描述并非全然有效。由表 2 的实验结果可见,同一批实验中的两个菌落 a、b,其性能评价结果差异较大,覆盖维值却非常接近,表明覆盖维作为描述菌落形态的唯一特征是不充分的。这是因为覆盖维作为一类单标度分形维数,虽然揭示了集团的自相似构造规律,但不能完全刻划其局域特性。

表2 菌落图象的覆盖维与菌种性能的关系

菌落编号	覆盖维值	初筛实验转化率(%)	复筛实验转化率(%)	活性评价
a	2.247478	53	40	劣
b	2.247807	70	68	良

(2) 多重分形谱的计算: 多重分形(multifractal)是定义在分形结构上的由多个标度指数的分形测度所组成的无限集合, 可以反映分形结构的微观性质^[6], 因此在一定程度上弥补了单标度分形的不足。描述多重分形局域性质有两套基本语言: 一是奇异指数 α 和多重分形奇异谱 $f(\alpha)$, 另一是描述多重分形分布均衡程度的参量 q 和广义分形维数 D_q 。利用两者之间存在的关系式:

$$f[\alpha(q)] = q \cdot \alpha(q) - (q-1) \cdot D_q$$

$$\text{其中 } \alpha(q) = \frac{d}{dq} [(q-1) \cdot D_q]$$

可以求出多标度分形谱 $f(\alpha)$ 。

对表1中菌落a、b的形貌图象抽取所得的多重分形谱(图2)存在较明显的差异, 显示出多重分形在描述具有相同形成机制、不同几何特征和相似单标度维数的分形集团时, 可以提供更深入、细致的信息, 局域特性的刻画更真实、有效。

(3) 分类器设计: 计算机模式识别是用计算机模拟人的各种识别的技术^[7], 主要有两种方法: 判别法和句法识别。前者相对成熟, 实用性好, 是通过对样本集的“学习”, 寻找有效的分类规律, 来完成未知对象的识别。判别法主要有两类: 有监督和无监督识别算法。在训练样本预先已知归属时使用前者。

实验分3步进行: 第1, 特征抽取。为便于特征选择和分类器设计, 在多重分形谱线上定义了如图3所示的14个特征用来描述谱线形状, 构成用于模式识别的特征群。第2, 特征选择。采用人机交互的方式从特征群中选出最适合训练样本聚类的有效特征, 作为分类器的特征参数。经3批菌落训练样本反复实验, 确定特征 α -right、 α -width和 f -start为用于分类器设计的特征矢量。第3, 分类器设计。采用有监督识别算法, 设计了统计分段线性分类器, 用于菌落形态的自动识别和菌种

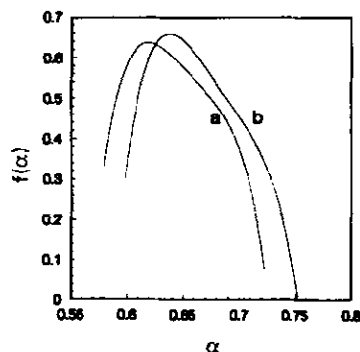


图2 表2中菌落a与b的多重分形谱
 α -奇异指数 $f(\alpha)$ -多重分形的奇异谱

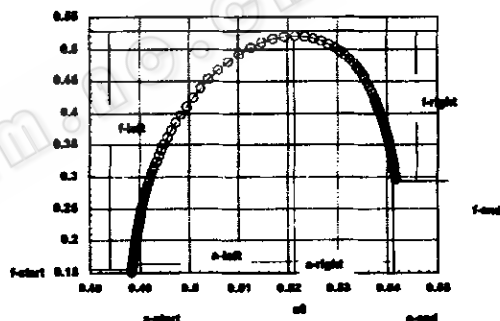


图3 多重分形谱特征的定义

- 1 α_o , 2 α -right, 3 α -left, 4 α -start, 5 α -end 6 f -right,
7 f -left, 8 f -start, 9 f -end, 10 α -width = α -right + α -left,
11 s -left = f -left / α -left, 12 s -right = f -right / α -right,
13 α lr = α -left / α -right, 14 f lr = f -left / f -right

筛选, 与人工挑选菌落筛选法的比较如表3所示。

2.4 讨论

(1) 菌落形态的多重分形表征比单标度覆盖维特征的描述更细致、有效, 与菌种性能的相关性更大。

(2) 以多重分形特征 α -right、 α -width和 f -start为依据设计的分类器, 可以用于优良菌种的自动识别, 速度快, 与人工分离筛选的实验数据相吻合。

(3) 菌种自动筛选的实验步骤为(其中①~⑤为

表3 菌落计算机识别法与人工筛选法的比较

筛选方法	每批筛选数目	工作量和强度	筛选周期	准确度
计算机识别法	80~120	小	10d	误识率5%
人工筛选法	20~40	大	20~25d	—

落图象采集③→特征抽取→进入分类器→指导菌落单
挑④→斜面培养⑤→优良斜面菌种。

参 考 文 献

- [1] 俞俊棠, 唐孝宣. 生物工艺学. 上海: 华东化工学院出版社, 1991, 278.
- [2] 复旦大学、武汉大学生物系微生物学教研室编. 微生物学. 北京: 高等教育出版社, 1990, 56.

- [3] 王积分, 阳葵, 刘宏等. 高等学校化学学报, 1994, 15(5): 750.
- [4] Viesek T. Fractal Growth Phenomena. Singapore: World Scientific, 1989, 53.
- [5] Halsey T C. Phys. Rev. A. 1986, 33(2): 1141
- [6] Benzi R, Paladin G. J. Phys. A, 1984, 17(18): 3521.
- [7] 李介谷. 计算机模式识别技术. 上海: 上海交通大学出版社, 1986, 1.