

性状加权编码在放线菌数值分类上的应用

马俊才 赵玉峰 张克莉

(中国科学院微生物研究所, 北京)

摘要 本文提出了一种新的微生物形态性状的加权编码方式。使用这种加权编码方式及平均链锁法对放线菌游动放线菌属 (*Actinoplanes*) 和小瓶菌属 (*Ampullariella*) 共 116 株进行了聚类分析。结果表明: 通过性状加权处理可以使游动放线菌属和小瓶菌属得以区分。

关键词 数值分类; 性状加权; 放线菌分类

数值分类学中的一个重要原则是建立分类单元时, 每一性状都是等权的。但在微生物学传统分类中, 往往有些性状在其分类单元中具有较大的影响力。因此在微生物学数值分类中, 对个别性状的加权在所难免。在分类学分析中, 性状加权问题已呈现出新的重要性。这已为正统的种系发生学者所承认^[1]。种系发生系统学者 Henning (1966) 也采用了加权编码。但是为了防止对性状的恣意加权, 人们多年来一直企图寻找一种不局限于个别种类, 而又有较明确的自然意义的数学性的性状加权编码规则。本文对此进行了尝试, 并取得了一定效果。

材料和方法

1. 实验菌株: 游动放线菌属 (*Actinoplanes*) 共 91 株。其中标准菌株 27 株, 未知菌 64 株。小瓶菌属 (*Ampullariella*) 共 27 株, 其中标准菌 7 株, 未知菌 20 株。其它菌属 11 株。标准菌株详见表 1。

2. 性状指标: 主要包括形态、培养特征、生理生化等 75 项指标 (表 2)。全部数据均由阮继生、陆晓涛提供。

3. 聚类方法

(1) 建立原始数据矩阵: 把阳性结果定为 1; 阴性结果定为 0; 颜色等多态结果定为 1— n , 并将所有菌株的性状作成二维表, 作为计算相似性用原始数据矩阵。

(2) 计算菌株间的相似性系数: 设菌株 K 与菌株 L 间, a 为两菌中编码为 1 的性状数目 (即正配对数), b 和 c 为两菌中不相符的性状数目 (即错配对数), d 为两菌中编码为 0 的性状数目 (即负配对数), 则:

$$S_{SM}(K, L) = \frac{a + d}{a + b + c + d}$$

$$S_I(K, L) = \frac{a}{a + b + c}$$

(3) 聚合策略: 采用平均链锁法, 即群间相似性水平定义为群间的各 OTU 间的平均相似性^[2]。

(4) 性状消减: 在两态性状中, 去掉那些 90% 以上的菌株都呈一种状态的性状, 共去掉 14 项性状 (表 2 表 * 号者), 以突出其关键特征的作用。

(5) 性状加权: 在形态方面的多态性状中, 以每项性状的最大状态数为其权数的原则加权。如孢囊孢子形状一项, 状态数为 3, 则加权数为 3。加权后使整个性状数达到 95 项。

(6) 软硬件系统: 该程序用编译 BASIC 编写, 在长城 0520CH 机上运行, 树状谱等结果用点阵式打印机输出。

结果和讨论

先用常规的等权编码原则对性状进行编

本工作承蒙阮继生、陆晓涛老师悉心指导, 特此鸣谢。
加权编码处理后的结果被选入《第七届国际放线菌生物学会议》论文集, 1988, 5, 东京。

表 1 标准菌株表

1. Kcc A-0061 <i>Ampullariella lobata</i>	24. ATCC 15331 <i>Spirillospora albida</i>
2. Kcc A-0059 <i>Amp. campanulata</i>	25. AS4180 <i>Streptosporangium rubroaurantiacum</i>
3. JCM 03235 <i>Amp. regularis</i>	26. AS4. 1083 <i>Stresp. vulgare</i>
4. Y79--15 <i>Amp. kunmingensis</i>	27. Kcc A-0005 <i>Stresp. roseum</i>
5. AS4. 1012 <i>Actinoplanes kinshanensis</i>	28. AS4. 1090 <i>Stresp. amethystogenes</i>
6. AS4. 1154 <i>Amp. cylindrica</i>	29. Kcc A-0252 <i>Act. teichomyceticus</i>
7. AS4. 1011 <i>Amp. pekinensis</i>	30. AS4. 1054 <i>Act. pyriformis</i>
8. AS4. 1155 <i>Amp. pilifera</i>	31. AS4. 1158 <i>Act. missouriensis</i>
9. AS4. 1148 <i>Act. violaceus</i>	32. Kcc A-0251 <i>Act. sarveparensis</i>
10. AS4. 1044 <i>Act. pallidoaurantiacus</i>	33. Kcc A-0122 <i>Act. utahensis</i>
11. AS4. 1048 <i>Act. rutilosporangius</i>	34. Kcc A-0250 <i>Act. liguriae</i>
12. AS4. 1046 <i>Act. roseosporangius</i>	35. Kcc A-0246 <i>Act. azureus</i>
13. AS4. 1149 <i>Act. yunnanensis</i>	36. Kcc A-0196 <i>Act. brasiliensis</i>
14. AS4. 1065 <i>Act. italicus</i>	37. Kcc A-0195 <i>Act. caeruleus</i>
15. AS4. 1053 <i>Act. purpeobrunneus</i>	38. 12427 <i>Act. philippinensis</i>
16. AS4. 1185 <i>Act. minutisporangius</i>	39. Kcc A-0277 <i>Act. ferrugineus</i>
17. AS4. 1052 <i>Act. aurantiacus</i>	40. Kcc A-0248 <i>Act. garbadinensis</i>
18. AS4. 1153 <i>Amorphosporangium castaneum</i>	41. Kcc A-0249 <i>Act. ianahinogenes</i>
19. Kcc A0311 <i>Dactylosporangium matsuzakiense</i>	42. Kcc A-0165 <i>Act. italicus</i>
20. IF0 14181 <i>Dact. vinaceum</i>	43. Kcc A-0264 <i>Act. nipponensis</i>
21. IF0 14104 <i>Dact. variesporum</i>	44. Kcc A-0194 <i>Act. rectilineatus</i>
22. IF0 14103 <i>Dact. salmonium</i>	45. AS4. 1055 <i>Act. tuftstagellus</i>
23. A/23867 <i>Planobispora longispora</i>	

表 2 性状特征表

1. 孢囊	26. 分解尿素
2. 孢囊形状	27. 分解次黄嘌呤
3. 孢囊孢子形状	*28. 分解黄嘌呤
4. 孢囊孢子鞭毛	29. 分解腺嘌呤
5. 孢囊孢子排列	30. 分解鸟嘌呤
6. 孢囊孢子个数	31. 分解酪氨酸
7. 孢囊层	32. 分解酪蛋白
8. 气丝颜色 (ASP)	33. 分解 Tween 80
9. 气丝颜色(燕麦)	34. 分解明胶
10. 基丝颜色 (ASP)	35. 分解七叶素
11. 基丝颜色(燕麦)	36. 分解马尿酸
12. 可溶性色素 (ASP)	*37. 分解苯丙氨酸
13. 可溶性色素(燕麦)	38. 抗甲基紫
14. 基丝有横隔断裂	39. 抗砷罗红 B
*15. 菌丝形成菌核或假孢囊	40. 溶菌酶敏感
16. 孢子丝形态	41. 青霉素敏感
17. 孢子表面形态	42. 链霉素敏感
*18. 孢子带鞭毛	43. 在 10°C 生长
*19. 着生在短孢子梗上的孢子数	*44. 在 28°C 生长
20. 细胞壁含 meso-DAP	*45. 在 45°C 生长
*21. 细胞壁含 LL-DAP	46. 5% 氯化钠 (28°C) 生长
22. 全细胞糖型	47. pH8 生长
23. 硝酸盐还原	48. pH5 生长
*24. 淀粉水解	49. 利用苹果酸钙作碳源
*25. 产生硫化氢	50. 利用柠檬酸钙作碳源

- *51. 利用苯甲酸钠作碳源
- 52. 利用乳酸钠作碳源
- 53. 利用琥珀酸钠作碳源
- *54. 利用草酸钠作碳源
- 55. 利用糖酸钠作碳源
- *56. 利用卫矛醇产酸
- 57. 利用肌醇产酸
- *58. 利用 i-赤藓糖醇产酸
- 59. 利用 d(-)甘露醇产酸
- 60. 利用阿东糖醇产酸
- 61. 利用 d(+)半乳糖产酸
- 62. 利用 d(+)麦芽糖产酸
- 63. 利用 d(+)甘露糖产酸

- 64. 利用 L(+)阿拉伯产酸
- 65. 利用葡萄糖产酸
- 66. 利用 d(+)乳酸产酸
- 67. 利用 d(+)蜜二糖产酸
- 68. 利用甲基-D 葡萄糖苷产酸
- 69. 利用 d(+)棉子糖产酸
- 70. 利用 d(-)山梨醇产酸
- 71. 利用 d(+)木糖产酸
- 72. 利用蔗糖产酸
- 73. 利用海藻糖产酸
- 74. 利用果糖产酸
- 75. 利用鼠李糖产酸

* 者为性状消减项目

码, 采用 S_{SM} 系数和 S_j 系数按平均链锁方式进行

行聚类, 结果如图 1 所示。从图 1 中可以看出,

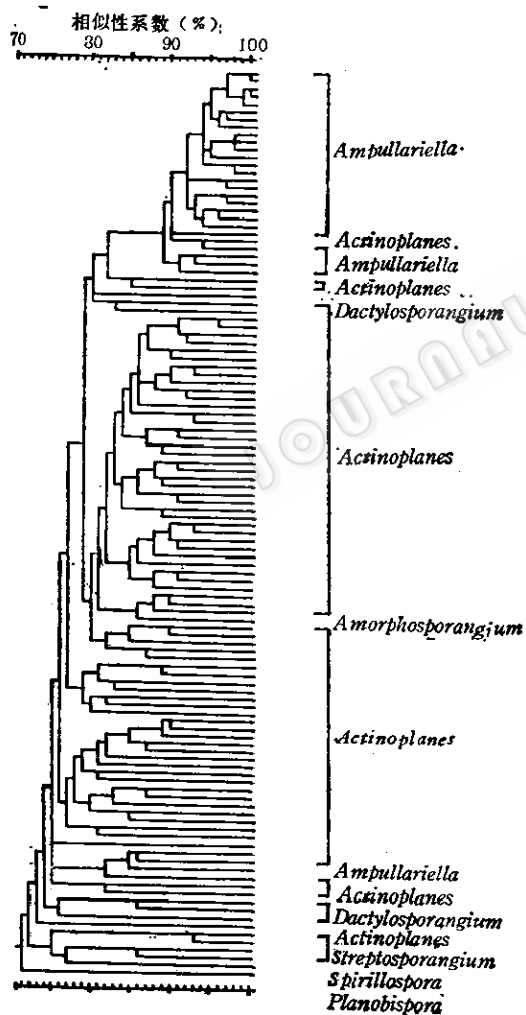


图 1 加权前用 S_{SM} 系数和平均链锁法得到的树状谱 (本树状谱由 MINTS 系统自动生成)

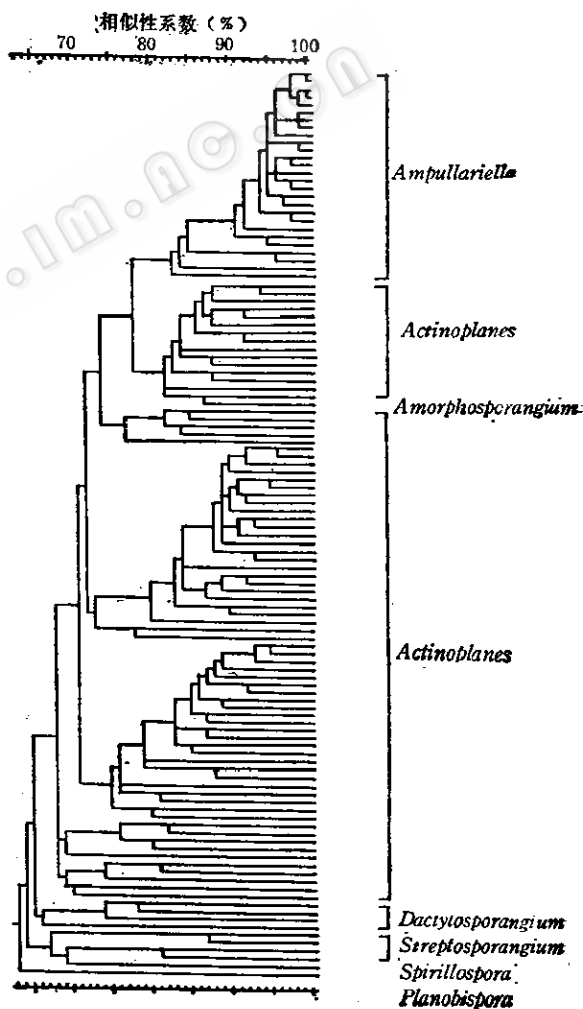


图 2 加权后用 S_{SM} 系数和平均链锁法得到的树状谱 (本树状谱由 MINTS 系统自动生成)

按一般效果较好的 S_{SM} 系数聚类后仍有 3 株游动菌属的菌株被划到小瓶菌属中去, 其中有 2 株是标准种, 一株小瓶菌属的菌株也进入游动菌属, 另有一株指孢囊菌属的菌株也被划入小瓶菌属。这说明虽然被选用性状项目很多, 但其中大量性状十分接近时, 按常规的编码原则会影响性状分辨率, 可能得不到正确结果。

针对以上结果我们对整个性状进行了消减性状和加权编码处理。消减性状为的是消除冗余性状, 而加权编码可以突出关键性特征的作用, 以增强分辨率。用 S_{SM} 系数和平均链锁进行聚类。结果如图 2。在图 2 中, 小瓶菌属呈紧密结合状, 说明其自身相似程度较高, 性状表现也较为统一; 无定形孢囊菌属位于游动放线菌属中, 并且相似性程度较高, 这一结果也符合《伯捷细菌学鉴定手册》新版已把无定形孢囊菌属归于游动放线菌属的分类方法; 指孢囊菌属、

链孢囊菌属、螺孢菌属、游动双孢菌属在树状谱上的成形较好, 说明其彼此间的性状互斥, 属内相似性较好。证明以形态为重要特征划分属是十分必要的。

我们认为通过对性状在一定范围内加权和性状消减, 可以起到突出属内关键特征的作用, 进而得到满意的结果。但这种加权切不可随意进行。只能根据客观的测试结果方可进行。

性状加权编码方式目前在数值分类中的应用尚不够广泛。但随着传统分类系统的进一步改进, 加权编码方式作为一种重要手段在数值分类学中, 特别是在微生物学数值分类中, 必将逐步得以发展和完善。

参 考 文 献

- [1] Sneath and Sokal: 数值分类学——数值分类的原理和应用, p233, 科学出版社, 北京, 1984。
- [2] 马俊才、赵玉峰等: 聚类分析在微生物数值分类上的