

应用微机进行苜蓿花叶病毒 株系分类的研究

李 毅 魏宁生 张文军

(植保系)

摘 要 应用主成分分析和模糊聚类法在微机上研究了苜蓿花叶病毒毒株的分类问题。结果表明, H-10分离物与苜蓿花叶典型株系完全相同, 属同一株系。Wc分离物自成一株系且与黄化斑块株系最接近。这些结果得到血清方法的证实, 并部分得到病毒外壳蛋白AA分析的证实。

主题词 苜蓿花叶病毒, 病毒株系, 分类学, 计算机应用

以在三生烟 (*Nicotiana tabacum* Xanthi-nc) 等7个鉴别寄主上的症状反应为依据, 本文研究了18个国内外已报道的苜蓿花叶病毒 (AMV) 毒株的相关性与分类问题。

1 原理与方法

1.1 分类性状的选择和量化

在本文中, 分类性状即为用以区分不同病毒株系的各种鉴别寄主对病毒株系的症状反应。此反应为定性描述, 因而须加以量化 (编码)。现规定在任一寄主上无反应为0, 有枯斑反应为1, 有系统反应为2, 同时具有后两种反应为3。原始数据见表1。

1.2 主成分分析

该方法^[1]的基本原理是从原来关系错综复杂、互为相关的许多性状中找出能反映这些性状内在联系的, 起主导作用的, 数目较少和互相独立的综合性状。其算法为:

(1) 原始数据整理: 设有 N 个样品 (株系), 每样品有 P 个性状, Z_{ij} 为样品 j 的第 i 个性状观测值, 则原始资料阵为 $Z = (Z_{ij})_{P \times N}$, 此处须对 Z_{ij} 进行标准化处理以消除量纲差异。

(2) 求性状的相关阵 R : $R = (r_{ij})_{P \times P}$, $r_{ii} = 1$, $i = 1, \dots, P$,

$$r_{iR} = \frac{1}{N-1} \sum_{j=1}^N Z_{ij} Z_{Rj}, \quad i, R = 1, \dots, P, \quad i \neq R.$$

(3) 求相关阵 R 的特征值与特征向量: R 的特征值 λ 由方程 $|\lambda E - R| = 0$ 求得。此处 $\lambda_i \geq 0$, $i = 1, \dots, P$, 对应 P 个非0正交向量 (特征向量)。通常选取 m 个主成

文稿收到日期: 1988-12-02

分,使其对应的 m 个特征值之和达 P 的85%以上,此时表示这 m 个主成分所含信息量占 P 个性状信息量的85%以上。

表1 18个AMV毒株的性状编码

毒 株	号 码	寄 主 症 状 编 码						
		三生烟	蔓陀萝	黄 瓜	辣 椒	豇 豆	蚕 豆	菜 豆
		1	2	3	4	5	6	7
苜蓿黄化花叶 ^[4]	1	3	2	0	0	3	3	3
爱达荷AMV ^[4]	2	3	0	0	2	3	3	2
苜蓿黄斑 ^[4]	3	3	2	0	2	3	3	2
叶脉坏死 ^[4]	4	2	2	0	2	3	3	3
块茎坏死 ^[4]	5	3	2	2	3	1	2	3
土豆杂色 ^[4]	6	3	2	2	3	1	3	3
黄化斑块 ^[4]	7	2	2	0	2	1	3	1
典型株系 ^[4]	8	3	2	2	3	1	3	1
辣椒AMV ^[4]	9	3	3	3	2	3	3	3
烟草AMV ^[4]	10	3	3	3	3	3	3	3
黄化斑点 ^[4]	11	2	2	3	0	3	3	2
H-10 分离物 ^[4]	12	3	2	2	3	1	3	1
Wc-分离物 ^[5]	13	2	2	0	2	1	2	1
15/64分离物 ^[6]	14	3	0	1	0	3	3	3
425 分离物 ^[7]	15	3	3	3	3	3	1	1
VRU 分离物 ^[7]	16	3	0	3	2	3	3	1
AA ₁ 分离物 ^[9]	17	3	3	3	2	3	3	1
YSMV ^[10]	18	3	3	2	2	3	3	2

注:毒株右上角号码均为参考文献号。

1.3 模糊聚类

该方法^[1-3]的原理是按各样品(株系)间的相似性建立一模糊相容关系,将此关系经矩阵合成运算改造为一模糊等价关系矩阵,尔后用 λ 水平截集方法将样品分类。运算过程为:计算 $R = (r_{ij})_{N \times N}$,

$$\text{其中 } r_{ij} = \frac{(x_i, x_j)}{\|x_i\| \|x_j\|}, \quad (x_i, x_j) = \sum_{R=1}^m x_i^R x_j^R,$$

$x_i = (x_i^1, x_i^2, \dots, x_i^m)^T$, $x_1, \dots, x_N$ 为 N 个待分类样品, m 为性状数。

计算 $Q = R = R \circ R \circ R \dots \circ R$,则 Q 为模糊等价关系矩阵。

上述所有运算均以BASIC语言在PC-8801微机上进行。

2 结果与分析

2.1 主成分分析结果

根据计算, 得主成分结果如表 2 所示, 而相关阵 R 为

三生烟	蔓陀罗	黄 瓜	辣 椒	豇 豆	蚕 豆	菜 豆
1	-0.030 60	0.367 23	0.267 26	0.188 98	0.027 86	0.151 52
	1	0.347 70	0.343 55	-0.040 48	-0.238 79	0.000 00
		1	0.265 84	0.156 70	-0.129 39	-0.100 51
			1	-0.471 40	-0.312 77	-0.251 97
				1	-0.147 44	0.267 26
					1	0.236 43
						1

* 表示与其对角对称的部分。

表2 对表1的主成分分析结果

主成分 序号	特征值	贡献率 (%)	累计贡献 率 (%)	特 征 向 量						
				x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	2.080 75	29.72	29.72	0.137 44	0.393 27	0.346 28	0.578 17	-0.326 55	-0.406 10	-0.317 33
2	1.582 56	22.61	52.33	0.564 21	0.185 85	0.511 71	-0.015 06	0.505 33	0.112 04	0.342 72
3	1.006 64	14.38	66.71	-0.565 76	0.583 65	0.092 38	-0.318 65	0.345 13	-0.326 17	0.060 78
4	0.884 18	12.63	79.34	-0.050 02	0.408 59	-0.259 88	0.251 35	-0.361 38	0.249 07	0.712 19
5	0.740 97	10.59	89.93	-0.238 30	0.247 33	0.341 78	-0.016 03	-0.104 44	0.791 16	-0.357 94
6	0.428 04	6.11	96.04	-0.281 19	-0.369 73	0.648 00	-0.219 04	-0.415 18	-0.150 76	0.348 19
7	0.276 87	3.96	100.00	-0.452 03	-0.324 54	0.073 84	0.672 72	0.452 79	0.073 74	0.147 94

由矩阵 R 及表 2 可见, R 的非对角元均非 1, P 个特征值均不为 0, 故各性状间无完全相关性。第一主成分主要由辣椒、蚕豆和蔓陀罗控制; 第二主成分主要由三生烟、黄瓜和豇豆控制; 第三主成分由蔓陀罗和三生烟控制; 第四、五主成分分别由菜豆和蚕豆所控制。若考虑性状的载荷 (重要性), 则各性状的重要性程度 (载荷值大小) 依次为: 辣椒 (0.83) > 三生烟 (0.71) > 蚕豆 (0.68) > 菜豆 (0.67) > 黄瓜 (0.64) > 豇豆 (0.63) > 蔓陀罗 (0.59)。由此可知各寄主在株系划分中的作用是不相同的。

在进行模糊聚类分析时, 为了不重复使用具有相关性的寄主症状反应以排除干扰 (将相关性状当作不相关性状计算距离等于将同一信息使用 2 次, 这在数学上是错误的), 故须选上述主成分分析得到的各病毒株系的综合性状进行模糊聚类。从表 2 可知前 5 个特征值对应 5 个主成分 (信息总量为 89.93%), 由此计算聚类的 5 个综合性状值, 方法为: 设 x_{ij} ($i = 1, \dots, 18; j = 1, \dots, 5$) 为样品 (株系) i 的第 j 个性状

值, 则 $x_{ij} = \sqrt{\lambda_j} \sum_{R=1}^7 \mu_{Rj} \cdot Y_{iR}$, 其中 λ_j 为第 j 个特征值, Y_{iR} 为原始数据 (表 1),

μ_{Rj} 为 λ_j 对应特征向量的第 R 个分量。由 $x_{ij}, i = 1, \dots, 18; j = 1, \dots, 5$ 即可进行模糊聚类了。

2.2 模糊聚类结果

由上述综合性状与转换值对 18 个 AMV 毒株进行模糊聚类, 得模糊等价关系矩阵为:

1
.9736 1
.9742 .9566 1
.9742 .9566 .9841 1
.9591 .9566 .9591 .9591 1
.9591 .9566 .9591 .9591 .9955 1
.9591 .9566 .9591 .9591 .9792 .9792 1
.9591 .9566 .9591 .9591 .9780 .9780 .9780 1
 $\theta =$.9591 .9566 .9591 .9591 .9774 .9774 .9774 .9774 1
.9591 .9566 .9591 .9591 .9774 .9774 .9774 .9774 .9947 1
.9655 .9566 .9655 .9655 .9591 .9591 .9591 .9591 .9591 .9591 1
.9591 .9566 .9591 .9591 .9780 .9780 .9780 1.000 .9774 .9774 .9591 1
.9591 .9566 .9591 .9591 .9921 .9921 .9792 .9780 .9774 .9774 .9591 .9780 1
.9548 .9736 .9566 .9566 .9566 .9566 .9566 .9566 .9566 .9566 .9566 .9566 1
.9847 .9548 .9548 .9548 .9548 .9548 .9548 .9548 .9548 .9548 .9548 .9548 .9548 1
.9537 .9537 .9537 .9537 .9537 .9537 .9537 .9537 .9537 .9537 .9537 .9537 .9537 1
.9591 .9566 .9591 .9591 .9774 .9774 .9774 .9774 .9893 .9893 .9591 .9774 .9774 .9566 .9548 .9537 1
.9591 .9566 .9591 .9591 .9774 .9774 .9774 .9774 .9995 .9947 .9591 .9774 .9774 .9566 .9857 .9537 .9893 1

- 为矩阵的对称部分

根据矩阵 θ ，以 λ 水平集方法进行分类。由 θ 可见， $\lambda=0.953$ 时，所有18个毒株属同一类。 $\lambda=1.0$ 时，18个AMV毒株可划分为17类， λ 由大到小，则分类由细变粗。结合实际情况，作者认为取 $\lambda=1.00$ 为宜，此时8（苜蓿花叶典型株系）与12（H-10分离物）完全相同，可认为属同一株系。其它16个毒株各成一株系。此外，由 θ 可见6（土豆杂色）与5（块茎坏死），9（辣椒AMV）与18（YSMV），7（黄化斑块）与13（W₁分离物），17（AA₁分离物）与9及18的关系均非常密切，但仍各为一独立株系。

3 讨论

植物病毒在其寄主上的症状反应由病毒基因(RNA或DNA)所控制^[1],故它是相对稳定的性状,由它可反映出不同病毒或同病毒不同株系间的差异。但要用多种鉴别寄主找出株系间的相互关系——即进行株系间的交叉比较是非常困难的。原因有三,其一是若不借助于数学手段,则不可能在诸多因素组合中对不同株系间在不同寄主上的症状反应进行系统的、客观的交叉比较;其二,各种鉴别寄主在株系划分中的重要性不同;其三,不同寄主对同一病毒株系的症状反应间有一定的相关性。以具有高度计算和综合能力的计算机为工具的主成分分析和模糊聚类分析可以克服上述的困难。这类方法有以下优点:主成分分析可得出每种鉴别寄主在株系分类中的作用,更重要的是它可给出足以代表这些寄主反应的信息以进行模糊聚类分析的不同株系的综合因子,从而排除了相关寄主反应或作用较小的寄主反应的干扰。后者省时省事。不足之处是不能区分系统反应中的各种

症状变化,但这一点在AMV株系分类不是主要的因素。

用上述方法研究AMV株系分类问题,结果得到血清学^[8]和病毒外壳蛋白AA分析^[12]的证实。Hull用血清学方法证明VRU株系与15/64株系关系最近,其次为425株系,而与其它株系较远^[8]。从矩阵 θ 可看出类似的结果。Krall^[12]通过对9个AMV分离物外壳蛋白AA成分的分析获得了组成每种分离物外壳蛋白的18种AA的数目,并应用数学手段得到了不同株系之间差异的数值,由此绘出株系分类图(见图1),结果表明YSMA与AA₁株系及P(Pepper)株系最为接近,故划为一亚组,425株系为一亚组,VRU株系和15/64株系为一亚组,但该亚组与前两个亚组关系较远。它们与作者所得结果的相同之处是:YSMV, AA₁及P三株系关系较近,且425株系与这三株系的关系也较近,而VRU和15/64两株系与前四个株系的关系均较远,VRU与15/64间的关系也较远,因为 $\lambda = 0.953$ 时才可归为一类(见矩阵 θ)。但Krall的结果表明VRU与15/64两株系关系较近^[12]。蔡发兴等认为H-10分离物与苜蓿花叶典型株系关系最近^[4],我们的结果是这两个株系完全相同。因为他们选择的寄主多,但仅在心叶烟(*Nicotiana glutinosa* L.)上有细微差异。本文结果仍需通过进一步分析这些株系基因组RNA的同源性关系及序列分析后才能得到最后的肯定。

近年来在植物病毒株系的鉴定与分类中有不少新进展,尤其是在利用计算机技术分析病毒生化属性方面,诸如外壳蛋白AA数目^[12],主要基因组RNA之间同源性关系及序列分析可直观地反映不同病毒株系的差异,从而研究株系的分类问题。这也是近年来研究相关病毒或同一病毒不同株系分类问题的趋势。

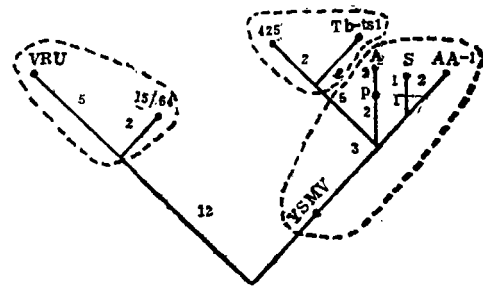


图1 AMV 9个分离物相互关系图

参 考 文 献

- 1 陈国良, 蒋定生, 帅启富等. 微机应用与农业系统模型. 西安: 陕西科技出版社, 1986, 120, 327
- 2 彭祖赠, 孙毓玉. Fuzzy 聚类分析. 模糊数学, 1982 (1): 125
- 3 龙漫远. 模糊聚类在作物遗传差异分析中的应用研究. 四川农业大学学报, 1986, 4 (2): 239
- 4 蔡发兴, 莽克强. 北京苜蓿花叶病毒H-10分离物的研究. I, 病毒的生物学测定、提纯及其理化性质. 微生物学报, 1982, 22: 233
- 5 魏宁生, 李毅. 苜蓿花叶病毒 (AMV) 白三叶草分离物—Wc的研究. 病毒学杂志, 1987, 4: 54
- 6 Hull R. Alfalfa mosaic virus. *Advances in Virus Research*, 1969, 15: 365
- 7 Hagedorn D J, Hanson E W. A strain of Alfalfa mosaic virus severe on *Trifolium pratense* and *Melilotus alba*. *Phytopath*, 1963, 53: 188
- 8 Hull R. Studies on Alfalfa mosaic virus TV, an unusual strain. *Virology*, 1970, 42: 283
- 9 饭塚典男, 饭田格. Bulletin of the Tohoku National Agricultural Experiment Station. 1969, 37: 43
- 10 Zaumeyer W J. Two new strains of Alfalfa mosaic virus systemically infections to bean. *Phytopath*, 1963, 53: 444

11 姜维善. 植物病毒学 (第二版). 北京: 科学出版社, 1985

12 Krall B. Amino acid analysis of Alfalfa mosaic virus coat proteins, an aid for viral strain identification. *Virology*, 1975, 66: 336

Classification of Alfalfa Mosaic Virus Strains Using Microcomputer

Li Yi Wei Ningsheng Zhang Wengjun

(Department of Plant Protection)

Abstract This paper deals with the classification of Alfalfa Mosaic Virus (AMV) strains by using main element analysis and fuzzy clustering method by means of microcomputer. The results show that H-10 isolate, is completely the same as typical AMV strains and Wc isolate is very close to yellow patch strain, but it is an independent strain. All these results are confirmed by serological method and partly by the method of analyzing amino acids (AA) composition of virus coat protein.

Subject words Alfalfa mosaic virus, virus strains, taxonomy, computer application