

转基因甘蓝型油菜亲和性等级识别方法构建

廖慧敏^{1,*}, 吴超¹, 王国槐², 朱建裕³, 马树宝¹, 汪发松¹

(1. 中南大学资源与安全工程学院, 长沙 410083; 2. 湖南农业大学植物科技学院, 长沙 410128;

3. 中南大学资源加工与生物工程学院 长沙 410083)

摘要: 为预测转基因甘蓝型油菜商业释放的基因流环境生态风险, 基于模糊数学中模糊聚类分析与模糊模型识别, 并结合生物统计学相关知识, 构建了转基因甘蓝型油菜亲和性等级识别方法。该方法包含转基因甘蓝型油菜种间组合亲和性的模糊聚类分析、模糊聚类结果 F 检验、种间组合风险等级划分和风险等级模式识别 4 个基本步骤。并在等级模式识别步骤中针对样本数据收集情况的不同, 构建了不同的等级识别方法, 与传统聚类的硬性划分相比更具灵活性。以收集的常规油菜杂交实验数据聚类形成 6 个风险等级模式, 将一甘蓝型油菜种间杂交实验进行亲和性等级识别为例对该方法进行了演示, 论证了该方法具可操作性。

关键词: 模糊聚类分析; 甘蓝型油菜; 亲和性; 等级识别; 方法

文章编号: 1000-0933(2009)10-5570-10 中图分类号: Q141, Q321, Q948, X826 文献标识码: A

Predicting transgene flow by modeling compatibility with *Brassia napus*

LIAO Hui-Min^{1,*}, WU Chao¹, WANG Guo-Huai², ZHU Jian-Yu³, MA Shu-Bao¹, WANG Fa-Song¹

1 School of Resource and Safety Engineering, Central South University, Changsha 410083, China

2 School of Bioscience and Technology, Hunan Agriculture University, Changsha 410128, China

3 School of Resource Processing & Bioengineering, Central South University, Changsha 410083, China

Acta Ecologica Sinica, 2009, 29(10): 5570 ~ 5579.

Abstract: To predict the environmental risk of the gene flow of transgenic *B. napus* after commercialization, a protocol for assessing its compatibility risks was proposed by the fuzzy clustering analysis and fuzzy model identification along with the biometric theory. The protocol included four steps: the fuzzy clustering analysis of interspecific crossing compatibility of transgenic *B. napus*, the F test of the outcome of the fuzzy clustering analysis, ranking of the risk of the interspecific crossing compatibility, and identifying the risk ranks. At the final stage of the protocol, various methods for identifying risk ranks were proposed to accommodate different data collection situations, therefore this approach was more flexible than the rigid ranking by the traditional clustering analysis. Subsequently, the protocol was demonstrated in a case study of the experimental data of common *B. napus*, and six ranks of risks were established, and the risk rank of the interspecific compatibility was identified successfully in a crossing experiment of *B. napus*. These results proved that the protocol proposed was operational.

Key Words: fuzzy clustering analysis; *B. napus*; compatibility; identification of risk ranks; protocol

甘蓝型油菜是全球一种重要的经济作物, 是中国、加拿大、澳大利亚及欧洲各国油菜的主要栽培类型^[1,2]。自 20 世纪 90 年代以来, 转基因油菜在全球种植趋势发展迅猛, 转基因甘蓝型油菜与转基因芸苔属蔬菜已在好几个国家得到商业释放批准, 其中转基因甘蓝型油菜占全球 2300 万 hm^2 油菜面积的 19%^[3]。转基

基金项目: 国家自然科学基金资助项目(30700008)

收稿日期: 2008-07-07; 修订日期: 2008-12-31

* 通讯作者 Corresponding author. E-mail: liaohuimin201@yahoo.com.cn

因油菜种植展现了巨大的经济前景,2005 年全球种植抗除草剂油菜所带来的农田经济效益增收达 195 百万美元^[4]。但甘蓝型油菜是典型的异花授粉作物,有着复杂的种间与属间亲缘关系,易与近缘种的作物和野生杂草发生异交,国内外通过有性杂交、原生质体融合和子房胚胎挽救等途径已获得了许多其与近缘植物的种间和属间杂种^[5~13]。因此转基因甘蓝型油菜商业化释放产生的比较突出的问题是转基因漂移的环境生态风险。如加拿大是全球最早种植转基因油菜的国家,在 1998~1999 年连续两年报道了转基因油菜基因漂移大田出现自播油菜成为“超级杂草”的事件^[14,15]。其风险评估方法的建立具有重要意义,并对其他转基因作物商业化释放风险评估具有借鉴作用。

模糊聚类分析是模糊数学的一个重要理论分支,其目的在于将特征空间中一组没有类别标记的矢量按某种相似性准则划分到若干子集中,使每个子集代表整个样本集的某些特征和性质。近年来模糊聚类理论已在生物学动植物分类^[16,17]、工程应用^[18,19]等诸多领域得到广泛应用。目前将模糊聚类用于灾害风险评估也有少量报道,如在火灾危害分类^[20]、风暴潮灾情等级识别^[21]中的应用等。本文拟以模糊聚类分析和模糊模式识别为基础,结合生物统计学相关知识构建转基因甘蓝型油菜亲和性风险等级识别方法,以期为转基因油菜环境风险评估方法体系的形成提供新思路。

1 转基因甘蓝型油菜风险等级识别方法构建

1.1 数据来源

利用常规芸苔属植物间的杂交实验资料,研究转基因油菜商业释放对环境生态系统的影响是转基因油菜基因漂移风险评估的一个重要内容^[3]。构建转基因甘蓝型油菜亲和性等级识别方法的数据主要来源于两种途径,其一可从网站(<http://biosafety.landcareresearch.co.nz>)数据库有偿获取,该网站将收集的芸苔属、萝卜属和白芥属的 8 类作物的亲和性实验数据汇总成数据库,其作用主要用于转基因油菜环境生态风险评估和育种资料。实验的杂交方式包含人工授粉、天然杂交和胚胎繁殖技术,资料涵盖 326 项研究和 216 个种间组合,收集的文献年代跨越从 1920 年至今,其数据比较全面。其二收集各国以往常规油菜杂交实验数据进行亲和性等级评估研究,由于这些实验数据获得的气候条件与转基因油菜商业化种植时的环境条件更为相似,因此该来源的数据具实际意义。

1.2 方法的基本步骤

1.2.1 模糊聚类基本步骤

聚类分析是按研究对象在性质上的亲疏关系进行分类的一种多元统计方法,能反映样本间的内在组合关系。在 F 数学产生以前,聚类分析已是数理统计多元分析的一个分支,它是一种硬性划分,即把每个待辨识的对象严格地划分到某个类中,非此即彼。但现实的分类问题往往伴有 F 性,事物类型的划分存在隶属程度的问题,对模糊对象给出切合实际的隶属函数,有效利用模糊集合理论和方法,可对现实中的模糊分类问题进行合理有效处理。其基本步骤如下:

(1) 确定待分类的种间杂交组合,收集符合条件的样本数据,待分类的全体样本构成定论域 $U = \{u_1, u_2, \dots, u_n\}$, 其中每一样本的特征数为 $u_i = (v_1, v_2, \dots, v_m) (i = 1, 2, \dots, m)$ 。

(2) 为消除不同量纲数据的差异,对样本数据进行归一化处理,采用级差变换等方法对样本数据进行变换。

(3) 建立 F 相似关系,设矩阵 $R = (r_{ij})_{m \times n}$ $r_{ij} \in [0, 1]$, 则称 R 为 F 矩阵, r_{ij} 为 F 矩阵的元素。对有限论域 $U = \{u_1, u_2, \dots, u_n\}$, $V = \{v_1, v_2, \dots, v_m\}$, 若元素 $r_{ij} = R(u_i, v_j)$, 则 F 矩阵 $R = (r_{ij})_{m \times n}$ 表示从 U 到 V 的一个 F 关系。 R 表示相似矩阵 $r_{ij} = R(u_i, u_j)$, 相似系数 $r_{ij} \in [0, 1]$, 其大小表示两事物的彼此相似程度。相似系数的确定方法有夹角余弦法、相关系数法等多种,可根据需要合理选择。

(4) 将 R 相似矩阵改造成 F 等价矩阵,采用平方法求出 R 的传递闭包 $t(R)$, 即从模糊矩阵 R 出发,依次求平方: $R \rightarrow R^2 \rightarrow R^4 \dots$, 当第一次出现 $R^k \cdot R^k = R^k$ 时,表明 R^k 已具有传递性, R^k 就是所求的传递闭包 $t(R)$ 。

(5) 模糊聚类 在 R 改造成模糊等价矩阵 R^k 后可在适当的限定值 λ 水平上进行截取,当 λ 由 1 逐步降

至 0 时,由 F 等价关系 R 确定的分类所含元素由少变多,逐步归并,最后成一类,从而获得所需的一个动态聚类图分类

1.2.2 模糊聚类结果的 F 检验

(1) 利用公式(1)^[26]计算各 λ 水平下模糊聚类结果的 F 值:

$$F = \frac{\sum_{j=1}^r n_j \|\bar{u}^{(j)} - \bar{u}\|^2 / (r-1)}{\sum_{j=1}^r \sum_{i=1}^{n_j} \|u_i^{(j)} - \bar{u}^{(j)}\|^2 / (n-r)} \quad (1)$$

其中, $\|\bar{u}^{(j)} - \bar{u}\| = \sqrt{\sum_{k=1}^m (\bar{u}_k^{(j)} - \bar{u}_k)^2}$ 为 $\bar{u}^{(j)}$ 与 $\bar{u}$ 的距离; $\|u_i^{(j)} - \bar{u}^{(j)}\|$ 为第 j 类中样本 $u_i^{(j)}$ 与 $\bar{u}^{(j)}$ 的距离。

(2) 将 F 值与各显著性水平自由度临界值进行比较,列 F 检验表。

1.2.3 各种间组合亲和性等级划分

根据 F 检验结果,对各种间组合亲和性的聚类结果确定一个明确分类。然后根据各组合亲和性指标值的变化,进行亲和性等级划分。

1.2.4 亲和性等级模式识别

由于甘蓝型油菜亲和性指标结角率,亲和指数与受精指数均服从正态分布,故其样本数据均为正态模糊集,其隶属函数可以指定为公式(2):

$$A_i(u) = \frac{1}{m} \sum_{j=1}^m \exp \left[- \left(\frac{u_j - \mu_{ij}}{\sigma_{ij}} \right)^2 \right] \quad (j = 1, 2, \dots, m; i = 1, 2, \dots, n) \quad (2)$$

因此视收集样本亲和性指标标准差是否齐备,将转基因甘蓝型油菜亲和性等级模式识别分两种情形:

(1) 当收集到的常规甘蓝型油菜亲和性指标具样本标准差时,用各样本平均数代替总体 μ_{ij} ,各样本标准差代替总体 σ_{ij} 计算待识别样本对明确分类后各等级类别模式的隶属度,根据最大隶属原则,则待识别样本亲和性等级应属于隶属度最大的样本识别模式。

(2) 当收集到的常规甘蓝型油菜亲和性指标缺失样本标准差时,可采用两种途径进行亲和性等级识别。途径一请经验丰富的专家估算总体标准差,然后根据样本亲和性指标具标准差情形(即上述第一种情形)确定待识别样本亲和性等级。途径二通过模糊优先比相似识别或格贴近度两种方法进行模糊亲和性模式识别。

模糊优先比相似识别的基本步骤:①设论域 $X = \{X_1, X_2, \dots, X_n\}$, 其中任意 $X_i, X_j \in X (i, j = 1, 2, \dots, n)$, X_0 为固定待识别样本。以 c_{ij} 表示成对样本 X_i, X_j 与 X_0 的相似优先比,即公式(3):

$$c_{ij} = \frac{D_j}{D_i + D_j} \quad (3)$$

其中,海明距离 $D_i = |X_i - X_0|$, $D_j = |X_j - X_0|$

可得模糊相似优先比关系 R 为映射: $R = (c_{ij})_{n \times n} (i, j = 1, 2, \dots, n)$, 模糊矩阵 $C = [c_{ij}]_{n \times n}$ 为相似优先比关系矩阵。② 利用不同的 λ 水平截割, $\lambda \in [0, 1]$, 得 C 的 λ 截矩阵 $= (c_{ij}^{(\lambda)})$, 当 $c_{ij} \geq \lambda, c_{ij}^{(\lambda)} = 1$; 当 $c_{ij} < \lambda, c_{ij}^{(\lambda)} = 0$ 。③ 令 λ 从 1 到 0 逐渐变化,当 $c_{ij}^{(\lambda)}$ 首次出现某一行除对角线外全部为 1 的情况时,则认为该行元素为第一优先相似,划去该行该列,在相似优先比关系矩阵中除去第 1 优越元素,可得到新的优先关系矩阵。依此类推,可将全部元素排出相似优先比顺序。④ 当样本包含多个特征因素时,首先取每样本第一个因素按上述方法求其相似优先比序号,然后将每个样本的诸因素序号求和,最后将总序号和排序,序号和越小,该样本与固定样本相似程度越接近。

格贴近度模糊模式识别基本步骤:① 列出待识别样本与固定亲和性模式样本因素集;② 计算待识别样本与相邻优先顺序两固定亲和性模式样本的格贴近度,格贴近度的表达式见公式(4);③ 根据格贴近度大小确定待识别样本亲和性等级。

$$N(A, B) = [\bigvee_{u \in U} (A(u) \wedge B(u))] \wedge [1 - (\bigwedge_{u \in U} (A(u) \vee B(u)))] \quad (4)$$

1.3 方法使用的一点说明

亲和性等级识别模式步骤中当收集的样本数据不具标准差时,可采用格贴近度与模糊优先比相似识别两种方法进行亲和性等级识别,其中前者计算比较简单,但可能会出现多组待识别样本与亲和性等级模式样本格贴近度相等的情形,从而不能达到有效识别风险等级的目的。采用模糊优先比相似识别法进行风级识别不会出现上述情形,但该方法计算量要大于前者,因此这两种亲和性等级模式识别方法在实践应用中须合理选用。

2 转基因甘蓝型油菜亲和性等级识别方法演示

2.1 甘蓝型油菜 11 个种间组合亲和性的模糊聚类

本文选用了甘蓝型油菜自交以及其与 *Brassica. campestris* L. ssp. *Chinensis* (L.), *Brassica. oleracea* L. var. *Capitata.*, *Brassica. nigra* (L) Koch., *Brassic. Carinata* A. Braun., *B. juncea* L. Czern& Coss. 5 个种正反交的 11 个种间杂交组合资料,为使模糊分析结果接近大田自然状态,选用的这些种间杂交组合均在大田中完成人工授粉,其授粉过程中未采用辅助手段(如用激素处理柱头等)。杂交组合的染色体组及试验资料数据见表 1,每一种间杂交组合包含 3 对以上品种杂交结果,其中根据文献^[3] $B \times AC, AC \times B, AC \times BC, AC \times C$ 4 个种间杂交组合虽有种间远缘杂交的田间试验文献,但均未获得大田自然条件下的杂交种子,故该 4 个组合的 3 个亲和性指标均为零。定义 $U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9, u_{10}, u_{11}\}$ 为 11 个种间杂交组合的定论域,其每组合的亲和性特征由结角率(v_1)、亲和指数(v_2)和受精指数(v_3) 3 个指标表示,其表达式见表 1 备注。 $u_i = (v_1, v_2, v_3) (i = 1, 2, \dots, 11)$ 。

表 1 各种间组合亲和性指标数值表

Table 1 The compatibility index of interspecific cross

亲和性项目 Compatibility item	AC × AC U_1	A × AC U_2	AB × AC U_3	B × AC U_4	BC × AC U_5	C × AC U_6
结角率(%) The rate of silique set	81.52	70.24	55.98	0	32.52	28.82
亲和指数 Compatibility index	12.65	0.89	3.04	0	0.05	0
受精指数 Fertilization index	13.03	4.63	3.42	0	1.46	0.02
亲和性项目 Compatibility item	AC × A U_7	AC × AB U_8	AC × B U_9	AC × BC U_{10}	AC × C U_{11}	
结角率(%) The rate of silique set	71.15	44.65	0	0	0	
亲和指数 Compatibility index	13.34	0.16	0	0	0	
受精指数 Fertilization index	15.12	1.24	0	0	0	

结角率 = 角果数/授粉花朵数 × 100%; 亲和指数 = 饱满种子数/授粉花朵数; 受精指数 = 总种子数/授粉花朵数; 数据来源于文献^[3, 22-25]

为减少表 1 中 3 指标不同量纲的数据差异,对其数据进行归一化处理,数据变换采用公式(5)进行级差变换^[21, 26],变换后资料数据转换到[0, 1]区间之内,得到归一化后亲和性指标列于如下 U 矩阵:

$$u_{ij} = \frac{u_{ij}^0 - \min(u_j^0)}{\max(u_j^0) - \min(u_j^0)} (i = 1, 2, \dots, n; j = 1, 2, \dots, m) \quad (5)$$

其中, u_{ij} 为归一化值; u_{ij}^0 为第 i 个样本的第 j 个指标观测值; $\min(u_j^0)$ 为第 j 个指标的最小观测值; $\max(u_j^0)$ 为第 j 个指标的最大观测值

$$U = \begin{bmatrix} 1.00 & 0.86 & 0.69 & 0.00 & 0.40 & 0.35 & 0.87 & 0.55 & 0.00 & 0.00 & 0.00 \\ 0.95 & 0.07 & 0.23 & 0.00 & 0.00 & 0.00 & 1.00 & 0.01 & 0.00 & 0.00 & 0.00 \\ 0.86 & 0.31 & 0.23 & 0.00 & 0.10 & 0.00 & 1.00 & 0.08 & 0.00 & 0.00 & 0.00 \end{bmatrix}$$

对归一化论域 U 选用绝对值减数法公式(6)计算模糊相似关系,取 $c = 0.1$,得 F 相似矩阵 R :

$$r_{ij} = \begin{cases} 1 & \text{当 } i=j \\ 1 - c \sum_{k=1}^m |x_{ik} - x_{jk}| & \text{当 } i \neq j \end{cases} \quad (6)$$

$$R = \begin{bmatrix} 1.000 & & & & & & & & & & \\ 0.843 & 1.000 & & & & & & & & & \\ 0.834 & 0.959 & 1.000 & & & & & & & & \\ 0.719 & 0.876 & 0.885 & 1.000 & & & & & & & \\ 0.769 & 0.926 & 0.935 & 0.950 & 1.000 & & & & & & \\ 0.754 & 0.911 & 0.920 & 0.965 & 0.985 & 1.000 & & & & & \\ 0.968 & 0.837 & 0.828 & 0.713 & 0.763 & 0.748 & 1.000 & & & & \\ 0.783 & 0.940 & 0.949 & 0.936 & 0.982 & 0.971 & 0.777 & 1.000 & & & \\ 0.719 & 0.876 & 0.885 & 1.000 & 0.950 & 0.965 & 0.713 & 0.936 & 1.000 & & \\ 0.719 & 0.876 & 0.885 & 1.000 & 0.950 & 0.965 & 0.713 & 0.936 & 1.000 & 1.000 & \\ 0.719 & 0.876 & 0.885 & 1.000 & 0.950 & 0.965 & 0.713 & 0.936 & 1.000 & 1.000 & 1.000 \end{bmatrix}$$

对称

经计算, $R^8 = R^4 \cdot R^4$, 故 R^4 为模糊等价矩阵, 利用模糊矩阵的截矩阵公式(7), 按不同的 λ 值水平, 由此可产生不同的分类, 从而得出如图 1 的聚类图。

$$r_{ij}^{(\lambda)} = \begin{cases} 1 & r_{ij} \geq \lambda \\ 0 & r_{ij} < \lambda \end{cases} \quad (7)$$

$$R^4 = \begin{bmatrix} 1.000 & & & & & & & & & & \\ 0.843 & 1.000 & & & & & & & & & \\ 0.843 & 0.959 & 1.000 & & & & & & & & \\ 0.843 & 0.949 & 0.949 & 1.000 & & & & & & & \\ 0.843 & 0.949 & 0.949 & 0.965 & 1.000 & & & & & & \\ 0.843 & 0.949 & 0.949 & 0.965 & 0.985 & 1.000 & & & & & \\ 0.968 & 0.843 & 0.843 & 0.843 & 0.843 & 0.843 & 1.000 & & & & \\ 0.843 & 0.949 & 0.949 & 0.965 & 0.982 & 0.982 & 0.843 & 1.000 & & & \\ 0.843 & 0.949 & 0.949 & 1.000 & 0.965 & 0.965 & 0.843 & 0.965 & 1.000 & & \\ 0.843 & 0.949 & 0.949 & 1.000 & 0.965 & 0.965 & 0.843 & 0.965 & 1.000 & 1.000 & \\ 0.843 & 0.949 & 0.949 & 1.000 & 0.965 & 0.965 & 0.843 & 0.965 & 1.000 & 1.000 & 1.000 \end{bmatrix}$$

对称

图 1 对甘蓝型油菜自交及与芸苔属种间杂交的 11 个组合亲和性进行动态聚类, 其聚类情况如下: 当相关系数 λ (图 1 左侧数据) 为 1 时, 分为 8 类: $\{u_1\}$, $\{u_2\}$, $\{u_3\}$, $\{u_4, u_9, u_{10}, u_{11}\}$, $\{u_5\}$, $\{u_6\}$, $\{u_7\}$, $\{u_8\}$ 。其中 $\{u_4, u_9, u_{10}, u_{11}\}$ 聚为第一类, 这与试验资料结果相符, 根据文献资料, 这 4 个组合虽有进行杂交试验的报道, 但文献中记载未获得这些组合的杂交种子^[3], 因此该类相关系数的聚类应是已知的。当 λ 为 0.985 时, 11 个组合分为 7 类: $\{u_1\}$, $\{u_2\}$, $\{u_3\}$, $\{u_4, u_9, u_{10}, u_{11}\}$, $\{u_7\}$, $\{u_8\}$, $\{u_5, u_6\}$ 。其中 $BC \times AC$, $C \times AC$ 两组合聚为第二类, 表明两组合相关性最好。当 λ 为 0.982 时, 组合分为 6 类: $\{u_1\}$, $\{u_2\}$, $\{u_3\}$, $\{u_4, u_9, u_{10}, u_{11}\}$, $\{u_5, u_6, u_8\}$, $\{u_7\}$ 。 $AC \times AB$ 与第二类中的两组合共聚为第三类, 这说明该三者相关性较好。当 λ 为 0.968 时, $AC \times AC$, $AC \times A$ 两组合聚为第四类, 种间组合聚为 5 类: $\{u_1, u_7\}$, $\{u_2\}$, $\{u_3\}$, $\{u_4, u_9, u_{10}, u_{11}\}$, $\{u_5, u_6,$

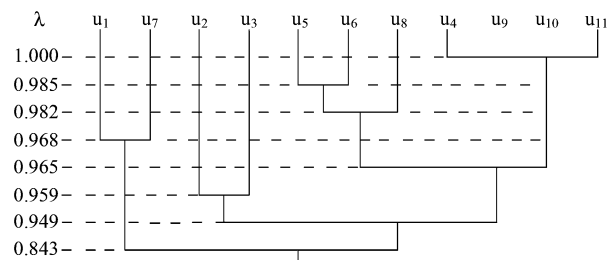


图 1 甘蓝型油菜自交及与近缘种杂交组合亲和性聚类图

Fig. 1 The Fuzzy Clustering of *B. napus* self-cross and its interspecific cross with allied genera

u_8 }。当 λ 为 0.965 时,第一类与第三类聚为较大的第五类,种间组合聚为 4 类: $\{u_1, u_7\}, \{u_2\}, \{u_3\}, \{u_4, u_5, u_6, u_8, u_9, u_{10}, u_{11}\}$ 。当 λ 为 0.959 时, $A \times AC, AB \times AC$ 聚为第六类,种间组合聚为 3 类: $\{u_1, u_7\}, \{u_2, u_3\}, \{u_4, u_5, u_6, u_8, u_9, u_{10}, u_{11}\}$ 。当 λ 为 0.949 时,第六类与第五类聚为更大的第 7 类,当 λ 为 0.843 时,所有指标归一为一个结合类。至此完成甘蓝型油菜自交及与芸苔属种间正反交 11 个组合亲和性的模糊聚类。

2.2 模糊聚类结果的 F 检验

为得到各种间组合亲和性的一个明确分类,对不同 λ 值的各分类进行 F 检验, F 检验统计量计算见公式 (1),其检验结果见表 2。

表 2 不同 λ 值各分类的 F 检验
Table 2 F-test of every class under different λ level

λ	分类数(r) Classification number	F	自由度($r-1, n-r$) Degrees of freedom	$F_{0.005}$
0.985	7	47.13 **	6,4	21.97
0.982	6	173.80 **	5,5	14.94
0.968	5	13.39 **	4,6	12.03
0.965	4	24.44 **	3,7	10.88
0.959	3	38.55 **	2,8	11.04
0.949	2	27.192 **	1,9	13.81

$P < 0.01$

2.3 各组合亲和性等级划分

从表 2 的 F 检验可看出在不同 λ 水平下,各组聚类都达到极显著水平,因此上述各种间亲和性差异显著性分类结果有多个而不唯一。根据 F 值大小来确定最佳分类类别,即 11 个组合聚类的统计学显著差异从大到小的排列顺序为:6 类 > 7 类 > 3 类 > 2 类 > 4 类 > 5 类。当 11 个组合划分为 6 类时统计意义最显著,故 F 值最大即 $\lambda = 0.982$ 时,其聚类类别 $\{u_1\}, \{u_2\}, \{u_3\}, \{u_4, u_9, u_{10}, u_{11}\}, \{u_5, u_6, u_8\}, \{u_7\}$ 为最佳。聚类类别的亲和性指标越高,其亲缘种间发生基因流的风险概率就越高,反之则概率越低,因此根据 6 类别种间组合亲和性指标值从低到高的顺序,将其发生转基因漂移的风险等级分别对应定义如表 3。

表 3 各亲和性类别风险等级
Table 3 The risk rank of every class

聚类类别 Class of clustering	$\{u_4, u_9, u_{10}, u_{11}\}$	$\{u_5, u_6, u_8\}$	$\{u_3\}$	$\{u_2\}$	$\{u_1\}$	$\{u_7\}$
风险等级 Risk rank	安全 Safety	低度风险 Low risk	中度风险 Middle risk	高度风险 High risk	更高风险 Higher risk	极高风险 Highest risk

2.4 亲和性等级的模糊识别

经过模糊聚类 and 聚类结果检验,将上述 11 个组合的风险等级划分为如表 3 中的 6 类,现用一常规油菜种间杂交试验作为待识别风险等级样本 U_0 ,该杂交试验所选用的亲本均为湖南省具代表性的油菜与白菜品种,母本湘矮 A 与 3 个白菜品种的杂交亲和性研究未见相关文献报道,但其他不育系品种如四川 A、714A 等与湖南省本地白菜品种杂交表现了较高的亲和性^[27],可作为本待识别样本参照。同时选用 $u_9, u_6, u_3, u_2, u_1, u_7$ 6 个样本分别作为安全、低度风险、中度风险、较高风险、高度风险和极高风险 6 个风险等级识别模式,对待识别样本 U_0 进行亲和性等级识别。

2.4.1 待识别样本 U_0 试验

该试验于 2007 ~ 2008 年在湖南农业大学理学院蔬菜教学基地进行,母本材料为湘矮 A,父本材料为石板滩白菜、油白菜与长沙迟白菜 3 个白菜品种,供试品种由湖南农业大学油料所提供。父本与母本分别种植于不同小区,小区面积 $6m^2$,随机区组排列,重复 3 次,采用常规管理,得各杂交组合。开花时,各组合杂交 100

朵花,成熟时考察杂交组合的结角数,每果粒数,计算结角率,亲和指数和受精指数。其试验数据见表 4。

表 4 U_0 亲和性指标值
Table 4 Compatibility index of U_0

组合名称(AC × A) Rross combination	结角率 The rate of silique set	亲和指数 Compatibility index	受精指数 Fertilization index
湘矮 A × 石板滩白菜“xiangai” A × “shibantanbaicai”(Chinese cabbage)	76	11.29	13.72
湘矮 A × 油白菜“xiangai” A × “youbaicai”(Chinese cabbage)	66	9.92	11.15
湘矮 A × 长沙迟小白菜 “xiangai” A × “changshachixiaobaicai”(Chinese cabbage)	42	6.36	8.34

将 U_0 亲和性指标按公式(5)进行归一化处理得该种间组合的结角率、亲和指数和受精指数分别为 0.75, 0.73, 0.73。

2.4.2 待识别样本的风险等级识别

在本应用实例中,收集的风险等级识别模式样本亲和性指标缺失标准差,根据以上构建的风险等级识别方法,可采用专家估算总体亲和性指标标准差,通过隶属度法进行待识别样本风险等级识别,也可通过模糊优先比识别法与格贴近度进行待识别样本风险等级识别,现以模糊优先比识别法为例进行说明。

(1) U_0 第一亲和性指标结角率分别与 6 个样本从低风险等级到更高风险等级的两样本进行相似比较, i 为两样本中低风险等级, j 为两样本中高风险等级。首先 U_0 与 U_9, U_6 样本值比较,海明距离为 $D_i = |0 - 0.75| = 0.75, D_j = |0.35 - 0.75| = 0.4$

$$\text{相似优先比为 } C_{12} = \frac{0.4}{0.75 + 0.4} = 0.35$$

同理以此类推建立第一指标结角率的相似优先比矩阵

$$C(1) = \begin{bmatrix} 0.00 & 0.35 & 0.07 & 0.13 & 0.14 & 0.25 \\ 0.65 & 0.00 & 0.13 & 0.22 & 0.23 & 0.38 \\ 0.93 & 0.87 & 0.00 & 0.65 & 0.71 & 0.81 \\ 0.87 & 0.78 & 0.35 & 0.00 & 0.52 & 0.69 \\ 0.86 & 0.77 & 0.29 & 0.48 & 0.00 & 0.68 \\ 0.75 & 0.62 & 0.19 & 0.31 & 0.32 & 0.00 \end{bmatrix}$$

(2) 做 λ 截矩阵,取 λ 有如下关系见公式(8)

$$c_{ij} = \begin{cases} 0 & C_{ij} < \lambda \\ 1 & C_{ij} > \lambda \end{cases} \quad (8)$$

对 C_{ij} 作 λ 截矩阵(限于篇幅,其过程略),可得 6 样本结角率相似程度的优先顺序依次为: $U_3 = 1, U_2 = 2, U_7 = 3, U_1 = 4, U_6 = 5, U_9 = 6$

(3) 依次求出第 2 指标亲和指数和第 3 指标受精指数的相似优先比顺序,将 3 指标相似优先比顺序求和得 6 个风险等级识别模式相似程度序列,其结果见表 5。取表中综合相似优先比顺序最小值 6,认为 U_0 与 U_1 样本亲和性等级相似,为划分为高度风险等级。

2.5 亲和性等级识别模式的生物学意义

基因漂流是转基因植物可能引起的生态问题的主要风险之一,基因漂流问题一直是转基因植物生态风险评估和管理的关键问题和研究热点^[28]。两植物亲缘关系越近其亲和性越高,两者间发生基因流的可能性增大,发生基因流的风险越高。通过对常规甘蓝型油菜与其近缘 5 个种的 11 个正反交杂交组合的亲和性等级识别,认为 AC × C 杂交组合亲和性最高,发生基因流的风险最大;其次为 AC × AC 等,结果详见表 3,在农业生产实践中,可根据该识别模式对栽培作物及其周围的近缘植物进行亲和性高低的风险模式识别,这对转基因油菜和其他转基因芸苔属作物商业释放时与亲和性等级不同的近缘物种进行合理规划种植和采取相应的

农业栽培管理措施具有重要指导意义。同时油菜是一种典型的异花授粉作物,转基因油菜基因流风险的研究是转基因植物基因流风险的一个重要研究内容,其结果可为基因流风险评价提供最直接的依据,因此该评判方法亦对其他转基因作物基因流风险评估具有借鉴意义。

表 5 U_0 与 6 个风险等级识别模式的相似程度序列
Table 5 The analogy degree list between U_0 and six risk grade identify model

U_0	结角率 The rate of silique set	亲和指数 Ccompatibility index	受精指数 Ffertilization index	综合顺序 Comprehensive sequence	从小到大排序 In ascending order
U_9	6	5	5	16	6
U_6	5	5	5	15	5
U_3	1	3	4	8	3
U_2	2	4	3	9	4
U_1	4	1	1	6	1
U_7	3	2	2	7	2

3 结论

(1) 基于模糊聚类构建的转基因甘蓝型油菜亲和性等级识别方法,充分考虑了各亲和性等级差异的过渡,从而使各级分类具一定弹性,与传统聚类的“硬性划分”相比更符合人类的认知习惯。同时该方法在风险等级模式识别时针对数据收集情况的差异而采用了不同的风险等级识别方法,方法操作更为灵活。通过以常规甘蓝型油菜种间杂交亲和性等级识别为例对该方法的演示,论证了该方法具操作实用性。

(2) 近年来国内外对转基因油菜基因流的研究逐渐增多,如转基因抗除草剂油菜对近缘作物与十字花科杂草的基因漂移规律研究^[29,30]等。但转基因作物的环境生物安全评价起步较晚,许多对转基因油菜基因流风险评价方法构建还在摸索中,至今还未形成比较完善的转基因作物环境生物安全评价体系,该亲和性等级判别模式可作为转基因作物环境生物安全评价体系中的一个重要方法。

(3) 风险等级识别方法可利用以往常规甘蓝型油菜杂交文献资料对转基因甘蓝型油菜商业释放的亲和性高低进行分类,有效识别其发生基因流风险的高低,对转基因甘蓝型的商业释放农业生产实践具有指导作用,同时该风险识别方法避免了大田重复试验,降低了试验成本和转基因甘蓝型油菜大田试验时的生态环境风险。

(4) 该方法还处于理论探索阶段,有一定局限性。风险等级识别模式步骤中当收集的样本数据具标准差时,风险等级模式利用隶属度进行识别,这种识别方式较为简明,但按最大隶属原则进行识别是一种直接识别方法,只考虑隶属向量中的最大分量,其它分量均被忽略,造成部分信息丢失,在实际中可能会出现一些误差。同时芸薹属植物由白菜(*B. campestris* L., AA, $2n = 20$)、甘蓝(*B. oleracea* L., CC, $2n = 18$)、黑芥(*B. nigra* L., BB, $2n = 16$) 3 个基本种和甘蓝型油菜(*B. napus* L., AACC, $2n = 38$)、埃塞俄比亚芥(*B. carinata* Braun L., BBCC, $2n = 34$)、芥菜(*B. juncea* L., AABB, $2n = 36$) 3 个复合种组成^[31],其基本种间染色体组具一定的同源性,大田种植时有着复杂的亲缘关系,而且转基因甘蓝型油菜基因逃逸也与风力、传粉昆虫密度等诸多环境因子有关,因此该方法还有待在实践中进一步完善。

References:

- [1] Wen J, Tu J X, Li Z Y, *et al.* Improving Ovary and Embryo Culture Techniques for Efficient Resynthesis of *Brassica Napus* from Reciprocal Crosses between Yellow-seeded Diploids *B. Rapa* and *B. Oleracea*. *Euphytica*, Springer Science Business Media B. V., 2007. 93 — 102.
- [2] Liu Z S, Guan C Y, L X, *et al.* Study on interspecific cross between *Brassia napus* and *Brassica juncea*, *Chinese Journal of Oil Crop Sciences*, 2001, (2): 82 — 86.
- [3] Richard G. Fitz John, Tristan T. Armstrong, Linda E Newstrom Lloyd, *et al.* Hybridisation within *Brassica* and Allied Genera: evaluation of potential for Transgene Escape. *Euphytica*, 2007 (158): 209 — 230.
- [4] Graham Brookes, Peter Barfoot. GM Crps: The First Ten Years-Global Socio-Economic and Environmental Impacts. ISAAA Brief, No. 36, 2006: 2

—3. <http://biosafety.ihe.be>.

- [5] Zhu Y, Lin L B. Compatibility of distant hybridization between *Brassic napus* L. and *Eruca sativa* Mi. Chinese Agricultural Science Bulletin, 2005, (12): 170—173.
- [6] Zhu W S, Xu J, Hua Y W, *et al.* Genomic analysis of intergeneric and *orychophragmus violaceus* hybrids between *Brassica carinata* and *orychophragmus violaceus* by using in Situ hybridization. Acta Bot Boreal-Occident Sin, 2005, 25(4): 662—667.
- [7] Liu X L, Meng J L. Embryological researches on interspecific crossability of *Brassica Napus* L. and *B. carinata* braun. Acta Agronomica Sinica, 1995, 21(4): 385—392.
- [8] Chen S Z, Yin J M, Tang Z L, *et al.* A preliminary study on interspecific hybridization between *Brassica Napus* and *B. oleracea* var. *acephala*. Journal of Southwest Agricultural University, 2000, 22(3): 208—211.
- [9] Xu A X, Huang J Y, Jin P A, *et al.* Studies on interspecific hybridization between *Brassica Napus* and *Brassica juncea*. Acta Bot Boreal -Occident Sin, 1999, 19(3): 402—407.
- [10] Zh Y T, Li A M, Lu L, *et al.* Disomic alien addition lines generated through microspore culture from progeny of intergeneric hybrids between *Brassica Napus* and *Sinapis alba*. Acta Agronomica Sinica, 2006, 32(11): 1764—1766.
- [11] Yamagishi H, Landgren M, Forsberg J. Production of *Asymmetric Hybrids* between *Arabidopsis Thaliana* and *Brassica Napus* Utilizing an Efficient Protoplast Culture System. Theor Appl Genet, 2002 (104): 959—964.
- [12] Sridevi O, Sarla N. Production of Intergeneric Hybrids between *Sinapis Alba* and *Brassica Carinata*. Genetic Resources and Crop Evolution, 2005, (52): 839—845.
- [13] Li M T, Li Z Y, Zhang C Y, *et al.* Reproduction and Cytogenetic Characterization of Interspecific Hybrids Derived from Crosses between *Brassica Carinata* and *B. Rapa*. Theor Appl Genet, 2005, (110): 1284—1289.
- [14] Hall L, Topinka K. Pollen Flow between Herbicide-resistant *Brassica Napus* is the Cause of Multiple-resistant *B. apus* Volunteers. Weed Science, 2000, (48): 688—694.
- [15] Beckie H J, Hall L M. Impact of Herbicide-resistant Crops as Weeds in Canada Proceedings Brighton Crop Protection Council Conference. Weeds, 2001, 135—142.
- [16] Chen C W, Zhou S B. Isoenzyme pattern analysis for five kinds of enzyme in six species of *polygonatum* from Dabieshan Mountain. Guihaia, 2006, 26(4): 395—399.
- [17] He C X, Wu W W, Yin K S, *et al.* Fuzzy grouping analysis of population dynamics of *Liriomyza huidobrensis* and its parasitic enemies. Southwest China Journal of Agricultural Science, 2006, 19(4): 639—643.
- [18] Kang X B, Xu M, Zhang Q, *et al.* Applying fuzzy clustering analysis to forecast the disaster of tunnel discharge. Journal of Hunan University of Science & Technology (Natural Science Edition), 2007, 22(3): 81—85.
- [19] Fan Z C, Chen J S, Dong H Z, *et al.* Applying fuzzy clustering method to study leakage of dyke. Rock and Soil Mechanics, 2007, 27(7): 1214—1219.
- [20] Ma Y Z. Application of fuzzy clustering analysis to classification of fire harm of China. Journal of Disaster Prevention and Mitigation Engineering, 2006, 26(4): 414—419.
- [21] Sun Z, Zang L, Feng Q M. Fuzzy cluster analysis-based storm surge disaster gradation. Journal of Natural Disasters, 2007, 16(4): 49—55.
- [22] Liu Z S. Genetic and breeding studies on interspecific hybridization of rapeseed II. crossabilities and the F_1 hybrids between *Brassica Napus* and *Brassica juncea*. Chinese Journal of Oil Crop Sciences, 1994, 16(3): 1—6.
- [23] Liu Z S. Genetic and breeding studies on interspecific hybridization of rapeseed III. interspecific cross between *Brassica Napus* and *Brassica plants*. Crop Research, 1994, 8(8): 27—30.
- [24] Zhao H C, Du D Z, Liu Q Y, *et al.* Interspecific distant hybridization between *Brassica juncea* with multiloculus and *Brassica Napus*. Acta Bot Boreal-Occident Sin, 2003, 23(9): 1587—1591.
- [25] Wang G H, Guan C Y, Chen S Y, *et al.* Compatibility of distant hybridization between different sterile lines of rapeseed and vegetable varieties of cruciferae. Journal of Hunan Agricultural University (Natural Sciences), 2000, 26(5): 337—339.
- [26] Li H J. The Basic and Practical Algorithm of Fuzzy Mathematics. Beijing: Science Press, 2005. 210—211.
- [27] Wang G H, Guan C Y, Chen S Y, *et al.* Studies on the utilization of vegetative heterosis of intergeneric hybrids in Brassica. Acta Agronomica Sinica, 2003, 29(1): 54—58.
- [28] Fan L J, Zhou X P, Hu B M, *et al.* The gene flow risks of genetically engineered plants. Chinese Journal of Applied Ecology, 2001, 12(4): 630—632.
- [29] Pu H M, Qi C K, Zhang J F, *et al.* Studies on the gene flow from herbicide-tolerant GM rapeseed to its close relative crops. Acta Ecologica Sinica, 2005, 25(3): 581—588.

- [30] Pu H M, Qi C K, Zhang J F, *et al.* The studies on gene flow from GM herbicide-tolerant rapeseed to cruciferous weeds. *Acta Ecologica Sinica*, 2005, 25(4):911—917.
- [31] Guan C Y. Study on Rapeseed Ecology and Genetic Breeding. Changsha: Hunan Science & Technology Press, 1989. 292—294.

参考文献:

- [2] 刘忠松,官春云,李恂,等. 甘蓝型油菜与芥菜型油菜种间杂交研究,中国油料作物学报. 2001, (2):82~86.
- [5] 朱媛,林良斌. 甘蓝型油菜和芸芥属间杂交的亲性和研究. 中国农学通报,2005, (12):170~173.
- [6] 朱旺升,徐晶,华玉伟,等. 埃塞俄比亚芥与诸葛菜属间杂种的基因组原位杂交分析. 西北植物学报,2005,25(4):662~667.
- [7] 刘兴梁,孟金陵. 甘蓝型油菜与埃塞俄比亚芥种间可交配性的胚胎学研究. 作物学报,1995,21(4):385~392.
- [8] 陈树忠,殷家明,唐章林,等. 甘蓝型油菜与羽衣甘蓝远缘杂交初步研究. 西南农业大学学报,2000,22(3):208~211.
- [9] 徐爱遐,黄继英,金平安,等. 甘蓝型油菜和芥菜型油菜种间杂交研究. 西北植物学报,1999,19(3):402~407.
- [10] 张永泰,李爱民,陆莉,等. 通过甘蓝型油菜和白芥属间杂种后代的小孢子培养获得二体异附加系. 作物学报,2006, 32(11):1764~1766.
- [16] 陈存武,周守标. 大别山区六种黄精属植物的五种同工酶分析. 广西植物, 2006,26(4):395~399.
- [17] 何成兴,吴文伟,尹可锁,等. 南美斑潜蝇及其寄生蜂种群动态的模糊聚类分析. 西南农业学报, 2006, 19(4):639~643.
- [18] 康小兵,许模,张强,等. 模糊聚类分析预测隧道涌水灾害. 湖南科技大学学报(自然科学版),2007, 22(3):81~85.
- [19] 樊哲超,陈建生,董海洲,等. 模糊聚类方法研究堤坝渗漏. 岩土力学,2007,27(7):1214~1219.
- [20] 马咏真. 模糊聚类分析在中国火灾危害分类中的应用. 防灾减灾工程学报,2006,26(4):414~419.
- [21] 孙峥,庄丽,冯启民. 风暴潮灾情等级识别的模糊聚类分析方法研究. 自然灾害学报,2007,16(4):49~55.
- [22] 刘忠松. 油菜远缘杂交的遗传育种研究 II. 甘蓝型油菜与芥菜型油菜杂交的亲性和及其杂种一代. 中国油料,1994,16(3):1~6.
- [23] 刘忠松. 油菜远缘杂交的遗传育种研究 III. 甘蓝型油菜与芸薹属植物远缘杂交亲和性. 作物研究,1994,8(8):27~30.
- [24] 赵洪朝,杜德志,刘青元,等. 芥菜型多室油菜与甘蓝型油菜的种间远缘杂交. 西北植物学报,2003,23(9):1587~1591.
- [25] 王国槐,官春云,陈社员,等. 油菜雄性不育系与十字花科蔬菜远缘杂交亲和性研究. 湖南农业大学学报(自然科学版),2000,26(5):337~339.
- [26] 李鸿吉. 模糊数学基础及实用算法. 北京:科学出版社,2005. 210~211.
- [27] 王国槐,官春云,陈社员,等. 十字花科芸薹属种间杂种营养优势的利用研究. 作物学报,2003,29(1):54~58.
- [28] 樊龙江,周雪平,胡秉民,等. 转基因植物的基因漂流风险. 应用生态学报,2001,12(4):630~632.
- [29] 浦惠明,戚存扣,张洁夫,等. 转基因抗除草剂油菜对近缘作物的基因漂移. 生态学报,2005,25(3):581~588.
- [30] 浦惠明,戚存扣,张洁夫,等. 转基因抗除草剂油菜对十字花科杂草的基因漂移. 生态学报, 2005,25(4):911~917.
- [31] 官春云. 油菜生态和遗传育种研究. 长沙:湖南科学技术出版社,1989. 292~294.