

雌雄蚕蛹近红外反射光谱的差异 及其模式识别

潘 沈 元

陶 鸣

(江苏省徐州师范学院生物系 徐州 221009) (江苏省镇江蚕种场 镇江 212018)

孙 爱 群

金 同 铭

(江苏省徐州市茧丝绸研究所 徐州 221009) (北京市蔬菜研究中心 北京 100081)

摘要 采用 6250 型近红外光谱分析仪, 从波长 680 到 1 235 nm 分别对日系蚕蛹 217 颗、中系蚕蛹 174 颗、杂交种蚕蛹 330 颗做了扫描测试, 结果表明: 雌雄蚕蛹的近红外光谱有极显著的差异。从日系 217 颗蚕蛹中随机选取雌雄各 50 颗作为学习样本, 采用逐步判别方法从其二阶导数光谱中抽取 5 个特征波长作为观测向量, 利用二次型判别函数对其余 621 颗蚕蛹的性别进行识别, 其准确率达 98.7%。该方法明显优于以蚕蛹重量或大小识别蚕蛹雌雄的方法。

关键词 蚕蛹, 雌雄鉴别, 近红外光谱, 模式识别

家蚕 (*Bombyx mori*) 蛹的雌雄自动识别研究是蚕业工作者长期想解决的课题之一, 因为它会给家蚕良种繁育和育种带来极大方便, 具有重要的经济意义。家蚕遗传学家利用染色体易位, 育成了一些限性品种^[1], 这对区分蚕或茧的雌雄是一个突破性进展, 但由于染色体易位的影响, 限性品种在某些经济性状方面, 还赶不上非限性品种, 所以生产上目前主要以推广非限性品种为主。刘敬权等^[2]利用紫外线激发荧光识别蚕茧雌雄, 但这种方法同样受到所谓“判性品种”的限制, 并且茧壳荧光的遗传规律尚未搞清楚。

近年来, 近红外光谱分析技术已广泛应用于工业、农业、医药等各个方面^[3, 4], 这是一种非破坏性的仪器分析法, 具有安全可靠, 无致变作用的特点, 对家蚕育种和良种繁育中的雌雄鉴别比较合适。本研究首次将近红外光谱分析技术和模式识别理论相结合, 将其应用于蚕茧和蛹的模式识别^[5], 试图为蚕茧和蛹的雌雄识别自动化, 从理论上和方法上打下基础。这里仅对雌雄蚕蛹近红外光谱的差别及其识别方法作一详细报道。

1 材料和方法

1.1 实验材料及测试

本试验所用的蚕蛹, 夏蚕期为杂交种, 由徐州市茧丝绸研究所提供; 中秋蚕期为中

系品种和日系品种, 由江苏省镇江蚕种场选种站提供。测试工作在北京市蔬菜研究中心生化分析室(农业部开放实验室)进行。测试仪器为美国太平洋公司生产的 6250 型近红外光谱分析仪和 D-744 型光导纤维探头, 光栅式扫描, 波长范围 680 ~ 1235 nm, 由硅检测器测量反射信息。被测蚕蛹发育进度从复眼着色到触角着色之间。

首先削茧取蛹, 对每颗蚕蛹做近红外扫描, 由计算机自动将反射信息转换成 $\log(1/R)$ 值贮存, 其次称蛹重, 鉴别蚕蛹雌雄。本试验于 1994 年夏蚕期测试杂交种蚕蛹 330 颗, 其中雌蛹 150 颗, 雄蛹 180 颗; 于 1994 年中秋期测试中系蚕蛹 174 颗, 其中雌蛹 86 颗, 雄蛹 88 颗, 测试日系蚕蛹 217 颗, 其中雌蛹 100 颗, 雄蛹 117 颗。

1.2 雌雄蚕蛹的识别方法

采用 NSAS 光谱数学程序, 求原始光谱 ($\log 1/R$) 的二阶导数光谱。由学习样本二阶导数光谱数据, 采用逐步判别方法, 选取与蚕蛹性别有关的特征波长^[6]。比较雌雄蚕蛹光谱特征的差别, 并对其做统计检验。由学习样本的人选特征建立值 Bayes 二次型判别函数:

$$u_i(X) = (X - \bar{X}^{(i)})' S_i^{-1} (X - \bar{X}^{(i)}) + \ln |S_i| - 2 \ln q_i \quad i=1, 2$$

式中, $X = (x_1, x_2, \dots, x_p)'$, $x_j, j=1, 2, \dots, p$ 为与蚕蛹雌雄性别有关的 p 个人选特征值, $\bar{X}^{(i)}, S_i, q_i$ 分别为雌雄群体的样本均值向量, 协方差阵和先验概率。以此判别函数, 对任意一个未知样本 X 做识别。判别法则为: 若 $u_1(X) \leq u_2(X)$ 则判 X 为雌, 否则判 X 为雄。

判别效果的检验方法: (1) 刀切法: 从 n 个学习样本中逐一剔除一个样本, 以 $n-1$ 个学习样本建立判别函数, 对剔除样本做判别。求对 n 个剔除样本的错判率。(2) 用学习样本建立判别函数对检验样本预测。以上判别分析均在 Quick Basic 环境下编程, 计算程序设计采用双精度变量。

2 结果与分析

2.1 雌雄蚕蛹近红外扫描光谱的差异及统计特征

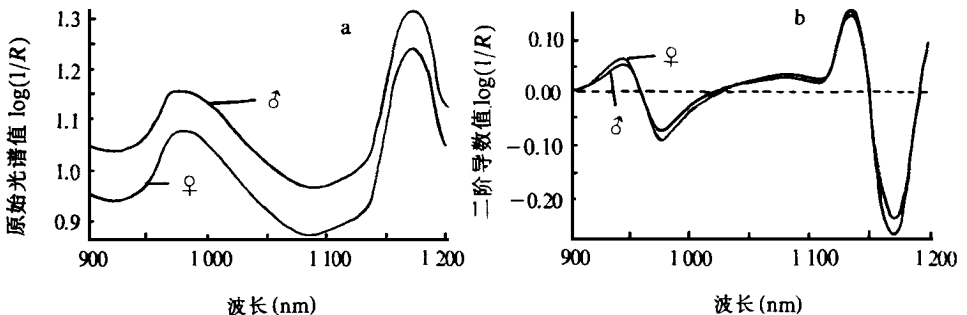


图1 雌雄蚕蛹各 50 个样本的平均扫描光谱

a. 示原始光谱 $\log(1/R)$; b. 示二阶导数光谱

从日系蚕蛹的 217 个样本中随机选取雌雄各 50 个, 分别求其原始光谱和二阶导数光谱的平均数, 由于雌雄蚕蛹的光谱差异主要反映在 900 ~ 1 200 nm 的波段上, 图 1 仅绘制了这个波段上的光谱图。

比较平均原始光谱和平均二阶导数光谱, 似乎原始光谱更能反映雌雄间的差异。其实不然, 因为蚕蛹原始光谱的差别, 受蚕蛹重量(大小)的影响较大, 蚕蛹越大, 反射的能量也就越多, 但是蚕蛹重量在个体间的变异较大, 从后面的分析可以看到, 仅依蚕蛹重量区分蚕蛹雌雄, 错判率是很高的。一、二阶导数光谱反映的是原始光谱的变化率, 它们的差别与蛹体的理化性质有关, 虽然在图示的精度上, 反映的光谱差异较小, 但是, 二阶导数光谱变异较小, 比较稳定, 更能反映出雌雄间的差异。

由于在扫描区间上的光谱值达 700 个, 为简化分析, 采用逐步判别方法, 从众多的波长测试数据中抽取与蚕蛹性别关系密切的特征波长 5 个, 现将 5 个波长处的原始光谱值和二阶导数光谱值分别列于表 1 和表 2。

表 1 原始光谱在 5 个特征波长处的统计特征

波 长 (nm)		921	933	1 041	1 099	1 182
雌 蛹	平均数	0.94076	0.94401	0.95996	0.89324	1.24481
	标准差	0.04277	0.04244	0.04001	0.03992	0.03797
雄 蛹	平均数	1.03666	1.04017	1.04793	0.98358	1.33610
	标准差	0.04091	0.04055	0.30846	0.03830	0.03850
F- 值		102.29**	115.30**	92.36**	13.78**	8.71**

注: F 检验为附加信息法, **表示 1% 的显著水平, $F_{0.01}(1, 94) = 6.91$, 下同

表 2 二阶导数光谱在 5 个特征波长处的统计特征

波 长 (nm)		921	933	1 041	1 099	1 182
雌 蛹	平均数	0.02435	0.04861	0.01229	0.02407	-0.15727
	标准差	0.00115	0.00238	0.00097	0.00100	0.00363
雄 蛹	平均数	0.02251	0.04281	0.01163	0.02176	-0.16301
	标准差	0.00113	0.00215	0.00069	0.00099	0.00344
F- 值		195.70**	319.29**	58.71**	43.41**	29.29**

表中 5 个波长处的原始光谱值与蚕蛹重量的复相关系数为 0.937, 与蚕蛹性别(0 和 1)的复相关系数为 0.925, 而二阶导数光谱值与蛹重的复相关系数为 0.935, 与蚕蛹性别的复相关系数为 0.972, 可见, 二阶导数光谱更能反映雌雄蚕蛹间的差异。

2.2 雌雄蚕蛹的光谱识别

以上面日系品种的 100 个样本的测试数据作为学习样本, 921, 933, 1 041, 1 099,

1 182 nm 5 个波长处的二阶导数值作为观察向量, 建立 Bayes 二次型判别函数, 采用刀切法估算错误率, 并以此判别函数对其余 117 个日系品种的光谱进行识别, 其识别效果见表 3。

在 100 个学习样本的刀切法检验和对 117 个检验样本的预测中, 仅有 2 个错判。根据试验记录知道其中一个为僵死蛹(雌), 另一个是软化的不正常蛹(雄)。为进一步检验对不同品种雌雄蛹的识别效果, 我们仍用前面所建立的二次型判别函数, 对 174 个中系蚕蛹和 330 个杂交种蚕蛹进行雌雄识别, 结果见表 4。

表 3 二次型判别函数对日系蚕蛹的雌雄识别效果

原属类别	判别结果			
	方法 1		方法 2	
	雌	雄	雌	雄
雌	50	0	49	1
雄	0	50	1	66
错判率(%)	0		1.71	

方法 1: 100 个学习样本的刀切法检验;

方法 2: 对日系品种 117 个检验样本的预测

虽然中系品种、日系品种和杂交种的蚕蛹在形态和重量上有一定的差别, 测试时间也不尽相同, 但品种间的雌雄识别仍然取得了很好的效果, 对 621 个检验蚕蛹性别预测的准确率达 98.7%, 这也说明不同品种的二阶导数光谱在反映雌雄蛹的差异方面具有共同特征。

2.3 以蛹重为特征值的雌雄判别效果与光谱识别的比较

人们曾以蚕蛹重量来区分雌雄蚕蛹, 其效果如何? 为与光谱识别法作一比较, 我们仍用 100 个日系蚕蛹为学习样本, 以蚕蛹重量为特征值建立判别函数, 对其余 621 个日系、中系、杂交种蚕蛹做雌雄鉴别, 其结果见表 5。

比较表 3、表 4、表 5 可知, 依蛹重识别蚕蛹雌雄的错判率是光谱识别法的 12 倍。这是由于蚕蛹重量在品种内和品种间有较大变异, 造成雌雄蛹间的差异不明显之缘故。我们还知道, 蚕蛹重量与蚕期的饲养环境条件有很大关系, 即使是同一品种, 在不同环境下饲养, 蛹重也会有很大差异, 所以仅依蚕蛹重量来区分其雌雄将会造成较高的错判率。而用蚕蛹的光谱识别方法, 基本上可以排除由于品种的不同和

表 4 二次型判别函数对不同品种蚕蛹的雌雄识别效果

原属类别	判别结果			
	方法 1		方法 2	
	雌	雄	雌	雄
雌	86	0	149	1
雄	0	88	5	175
错判率(%)	0		1.82	

方法 1: 对中系品种 174 个检验样本的预测;

方法 2: 对杂交种 330 个检验样本的预测

表 5 以蛹重为特征值对蚕蛹的雌雄判别效果

原属类别	判别结果			
	方法 1		方法 2	
	雌	雄	雌	雄
雌	43	7	266	20
雄	2	48	76	259
错判率(%)	9		15.46	

方法 1: 100 个学习样本的刀切法检验;

方法 2: 对日系、中系、杂交种 621 个检验样本的预测

环境因素影响所造成的蚕蛹重量差异的干扰。

3 讨论

影响雌雄蚕蛹近红外光谱差异的原因是值得进一步研究的。这种差别并不是某些品种特有的现象,因为我们选用的中系、日系以及杂交品种都是从生产上随机选取的,它们的二阶导数光谱都表现了共同的特征。如果说这种差别来自于蛹体的大小或重量,从表 5 已清楚地看到,仅由蛹重是难以识别蚕蛹雌雄的。我们从入选的 5 个二阶导数特征值与蛹重的相关系数(0.935),和与蚕蛹雌雄的相关系数(0.972)来看,入选的特征值中除了包含蛹重的信息外,还一定包含其他与雌雄性别有关的信息,我们初步推测这与蛹体内蚕卵的有无,脂肪的含量以及含水率有一定的关系,虽然我们还没有进一步证实,但从近红外技术的其它应用领域来看^[3, 4],我们可以概括地说,蚕蛹近红外光谱的差别是蛹体的大小和体内化学成分的综合影响。

表 1 和表 2 中,虽然入选特征值的平均数在雌雄群体间存在极显著的差异,但从每个特征值的标准差来看,仅仅依靠某一特征值是难以区分蚕蛹雌雄的,也就是说雌雄蚕蛹的光谱差异是光谱整体结构的综合差异,而不是某一、二个波长特征所能描述的。蚕蛹光谱识别的任务,就是通过适当的数学手段,使其光谱的结构能够反映出雌雄间的差异,并从众多的光谱数据中,抽取少数与蚕蛹性别有关的特征值,根据这些特征值构造判别函数,综合地反映雌雄蚕蛹光谱结构的差异,从而达到雌雄识别的目的。因此,光谱识别的效果关键在于我们能否利用光谱中所包含的有用信息。许多蚕业工作者也曾利用射线或光谱识别蚕蛹或蚕茧的雌雄,其未成功的原因很可能就是没有对数据进行合理的分析和利用,因为光谱的微小差别有时只凭肉眼难以察觉,通过适当的数学变换,就可以把这种差别反映出来。本研究正是结合了模式识别的原理和近红外光谱分析技术,才达到了对蚕蛹光谱识别的效果。

当然,本研究只是对蚕蛹光谱识别方法和可行性的探讨,离蚕蛹的自动识别还有一段距离,还有许多问题值得进一步研究。譬如学习样本的代表性和特征值选取的合理性问题,可以预计,当这些问题解决之后,蚕蛹识别的准确率还能进一步提高。当我们通过大量的实验确定了少数几个波长之后,对蚕蛹的光谱识别,就不需要做 680 ~ 1 235 nm 的全波段扫描了,这样可以大大减少测试时间。

参 考 文 献

- 1 中国农业科学院蚕业研究所主编. 家蚕遗传育种学. 北京: 科学出版社, 1981, 77 ~ 78, 344 ~ 345
- 2 刘敬权等. 荧光黄色判性品种和后代杂交遗传表现. 蚕业科, 1990, 16(3): 180 ~ 182
- 3 Biston R *et al.* Proceeding of Third International Conference on Near Infrared Spectroscopy. Gembloux (Belgium): Agricultural Research Centre Publishing, 1991, 1 ~ 726
- 4 Murray I *et al.* Making Light Work: Advances in Near Infrared Spectroscopy. Weinheim · New York Basel · Cambridge, 1992, 1 ~ 652
- 5 潘沈元等. 蚕茧近红外反射(NIR)光谱的模式识别 I. 对雌雄茧、死笼茧的非破坏性识别. 生物物理学报,

1995, 11(1): 53 ~ 59

6 孙尚拱等. 实用判别分析. 北京: 科学出版社, 1991, 67 ~ 136

7 沈清等. 模式识别导论. 长沙: 国防科技大学出版社, 1991, 56 ~ 104

DIFFERENCE BETWEEN MALE AND FEMALE PUPAE OF *BOMBYX MORI* IN NEAR INFRARED REFLECTANCE SPECTRA AND THEIR PATTERN RECOGNITION

Pan Shenyuan

(Biology Dept., Xuzhou Teachers
University Xuzhou 221009)

Tao Ming

(Zhejiang Silkworm Eggs Farm Jiangsu
Zhenjiang 212018)

Sun Aiqun

(Xuzhou cocoon and silk research
institute Xuzhou 221009)

Jin Tongming

(Beijing Vegetable Research Center
Beijing 100081)

Abstract In this paper, difference between male and female pupae of *Bombyx mori* in near infrared reflectance spectra (NIRS) and the feasibility and method of spectra recognition to identify the sex of fresh pupae were studied. By Model 6250 NIR instrument, indestructive determination of 721 fresh silkworm pupae selected from Chinese race, Japanese race and their hybrid was made with suitable wavelengths chosen between 680nm and 1235nm. Fifty male and fifty female pupae was randomly selected from 217 pupae of Japanese race as learning samples. The feature at 5 wavelengths was extracted as observation vector from data of their second derivative spectra by a stepwise discriminant analysis. Statistical analysis showed that there were extremely significant differences between male and female pupae in NIRS at the 5 wavelengths. When the Bayes' quadratic discriminant function built by the feature was used to predict the sex of the other 621 test samples, 98.7% of correct rate of the identification recognition was obtained. This method is much better than other identification ones which use weight or size of pupae as their feature. The mechanism is discussed.

Key words silkworm (*Bombyx mori*) pupa, determination of pupa sex, NIRS, pattern recognition.