技术与方法

中华蜜蜂形态遗传分析方法的研究*

朱翔杰** 周冰峰*** 王 媛 吴显达 汉学庆

(福建农林大学蜂学学院 福州 350002)

摘 要 本文以3个中华蜜蜂 *Apis cerana cerana* 种群为模型,测定常用的30个形态变量数据,通过比较逐步判别分析、主成分分析、聚类分析等多种多变量统计方法的分析结果,找出适合划分中华蜜蜂种群、揭示种群间遗传关系的数据分析方法。结果表明:逐步判别分析对于划分中华蜜蜂种群效果最好,并可进行准确率检验;与聚类分析联用有助于反映种群间遗传关系。主成分分析结果的准确度不高,有样点被分入错误种群,且遗传关系也存在偏差。聚类分析的结果与采用的形态变量和距离-分类方法组合的选择有关,不适合划分中华蜜蜂种群。本研究为中华蜜蜂种质资源调查和种群形态遗传分析等科学问题的研究方法提供参考。

关键词 中华蜜蜂,形态学,判别分析,主成分分析,聚类分析

Study on morphological genetics analytical method of *Apis cerana cerana* population

ZHU Xiang-Jie** ZHOU Bing-Feng*** WANG Yuan WU Xian-Da HAN Xue-Qing

(Fujian Agriculture and Forestry University, Fuzhou 350002, China)

Abstract Mutivariate statistical analyses of 30 morphometric variables collected from three *Apis cerana cerana* populations were used to model the genetic relationships between populations. Stepwise discriminant analysis proved a good method for classifying populations and had the advantage of allowing the percentage of correctly classified specimens to be tested using Cross-Validation. Cluster analysis could assist discriminant analysis reveal the relationship between populations. In contrast, factor analysis assigned some samples to the wrong population and gave much more variable estimates of the genetic relationships between populations. The results obtained from cluster analysis were also relatively imprecise, varying with the variables analyzed, genetic distance and clustering method.

Key words *Apis cerana cerana*, morphology, discriminate analysis, factor analysis, cluster analysis

蜜蜂形态遗传标记在种下分类和资源调查工作中起重要作用。Ruttner(1988)用形态遗传标记将世界饲养的两大蜂种西方蜜蜂 *Apis mellifera* 和东方蜜蜂 *Apis cerana* 进行了亚种分类,得到学术界的重视。杨冠煌等(1986)首次采用比较形态学方法对我国东方蜜蜂进行了初步归类,为我国东方蜜蜂资源调查研究做出了重大贡献。

2000年以来,随着实验技术的进步、计算机数据处理等的发展,形态遗传标记在地方蜂种资源调查和保护(Radloff *et al.*, 2005a, 2005b; Kandemir *et al.*, 2005)、蜜蜂种群遗传学(朱翔杰等,2009a;2009b;2009c)等领域得到广泛应用,同时在应用中发现现有的数据统计方法已不能满足蜜蜂种群层面的分析需要。主成分分析(principal components analysis)和聚类分析(cluster analysis)是目前用于亚种分类、地方蜂种形态分析常用的方法。采用不同的统计分析方法,结果往往存在很大差异,这给研究者们带来很大的困惑。在对

* 资助项目:现代农业峰产业技术体系建设专项资金(NYCYTX-43)、公益性行业(农业)科研专项(nyhyzx07-041)、福建省自然科学基金(2010J05043)。

** E-mail: xiangjie_zhu@126.com

*** 通讯作者,E-mail: bfymshj@pub3.fz.fj.cn

收稿日期:2010-10-05,接受日期:2010-11-12

中国东方蜜蜂的形态学研究中,用主成分分析得到 4 个类群,用聚类分析得到 7 个类群(谭垦等,2005)。对菲律宾东方蜜蜂的研究中,主成分分析把 Palawan 与其他地方区分开成为独立的一支,而聚类分析将 Palawan 划归为北方的 Luzon(Tilde *et al.*,2002)。出现这些问题的原因主要是两种方法本身固有的不足和方法运用过于机械。这给种群形态遗传学研究的数据分析方法提出更高要求。

逐步判别分析(stepwise discriminate analysis)可能成为划分中华蜜蜂 *Apis cerana cerana* 种群的更有效的数据统计方法。判别分析(discriminate analysis)也是多变量分析方法之一,在研究中一般用于检测预先分类结果的准确性(Kandemir *et al.*,2005)。逐步判别分析是判别分析方法的一种,具有筛选变量的功能。应用 Wilks' Lambda 统计量(组内离差平方和与总离差平方和的比值),首先纳入使该统计量减小最多的变量,这正是区分种群层面所需的标记特征,有助于中华蜜蜂种群的划分。根据蜜蜂形态学研究数据,样本量大,形态标记数量多的特点,选择有利于区分种群的标记,剔除干扰数据成为有效划分种群的一个思路。

本文以 3 个已知种群遗传关系的中华蜜蜂种群为模型,采用多个形态变量数据,通过与真实情况比较,评价逐步判别分析、主成分分析和聚类分析分析结果的准确性,判断逐步判别分析是否能成为更有效中华蜜蜂种群划分、种群间遗传关系的数据分析方法。为蜜蜂形态遗传标记在中华蜜蜂种质资源调查、种群形态遗传分析等科学问题的合理应用提供方法参考。也为形态标记在昆虫领域应用提供一种新的数据分析方法。

1　材料与方法

1.1　样本信息

选取 3 个已知种群,分别来自福州、武夷山和四川北部。福建福州 3 个样本,采自相距较近的连江、九峰和森林公园(FZ1、FZ2、FZ3);武夷山 5 个样本,采自大安、桂林、岚上、高桥和皮坑口(WY1、WY2、WY3、WY4、WY5);四川 5 个样本,采自马尔康(SC1、SC2)和金川(SC3、SC4、SC5)。

1.2　形态测定方法

用带有图形采集器(TK - C921EC,桂林光学仪器厂)的体式显微镜(GL-99TI,桂林光学仪器厂)与电脑形态测量系统(image-pro express)联用技术进行形态测定。

采用的形态标记参考 Ruttner(1988)[66] 并针对中华蜜蜂特征稍做调整:因中华蜜蜂第 5 背板绒毛带没有明显的黑带和浅色带的区分,将第 5 背板绒毛黑带长和绒毛黑带至底缘长合并为第 5 背板绒毛带长;东方蜜蜂和西方蜜蜂在色度(pigmentation)上存在较大差别,无法采用原有色度标准,本研究未采用。共测定了吻长、右前翅长、右前翅宽、胫节长、股节长、基跗节长、基跗节宽、第 3 背板长、第 4 背板长、第 4 腹板长、第 4 腹板蜡镜长、蜡镜斜长、蜡镜间距、第 7 腹板长、第 7 腹板宽、第 4 背板绒毛带长、第 6 背板绒毛长、后翅钩数、11 个翅脉角 A4、B4、D7、K19、G18、J10、L13、J16、N23、O26、E9 以及肘脉指数共 30 个形态遗传标记。

1.3　种群划分方法

逐步判别分析与聚类分析联用。先按样点初步分类,再根据各样点在判别函数 1 和判别函数 2 的二维图中重心的相对位置进行进一步分类;交互验证检测分类准确率;最后用各样点的重心聚类,体现种群间遗传关系。

主成分分析与聚类分析联用。用各样点形态指标平均数进行主成分分析,建立主成分 1 和主成分 2 的因子得分二维图;根据不同样点在图中的相对位置进行分类;最后用各样点的因子得分聚类,分析种群间遗传关系。

聚类分析本文采用多种距离 - 分类方法组合,并结合指标筛选,探索聚类分析运用的合理方法。

1.4　数据处理

逐步判别分析采用 Wilks' Lambda 法,该统计量为组内离差平方和与总离差平方和的比值,比较各形态遗传标记的判别能力。

主成分分析根据每个形态标记的负荷值,评价每个形态标记的贡献大小。负荷值小于 0.3,说明该因子的作用不大。

聚类分析在与判别分析和主成分分析联用时,采用常用系统默认的欧式距离和类平均法。单独使用时,采用系统聚类法,并进行了距离和聚类方法的选择。聚类分析常用的距离有欧式(Euclidean)距离、切比雪夫(Chebychev)距离、马

氏(Mahalanobis)距离和卡方(Chi-square)距离等,常用聚类方法有最短距离法、最长距离法、重心法、中间距离法、类平均法、离差平方和法等(唐启义和冯明光,2007)。

逐步判别分析和主成分分析采用 SPSS13.0 软件,聚类分析采用 DPS 软件进行。

2 结果与分析

逐步判别分析对于中华蜜蜂种群划分效果最好,并可进行准确率检验;与聚类分析联用可反映种群间遗传关系。主成分分析结果的准确度不高,有样点被分入错误种群,且遗传关系也存在偏差。聚类分析的结果与采用的形态变量和距离 - 分类方法的选择有关,不适合单独用于中华蜜蜂种群划分。

2.1 逐步判别分析 - 聚类分析联用

根据 Wilks' Lambda 统计量,按变量判别能力以从大到小的顺序分别为翅长、第 7 腹板长、胫节长、第 6 背板毛长、蜡镜间距、翅宽、蜡镜长、第 7 腹板宽、翅脉角 N23、A4、股节长、G18、第 4 腹板长、J16、第 4 背板长,各变量的判别效果显著($P < 0.01$)。

根据 13 个样点在判别函数 1、2 二维图中相对位置,进一步分类,明显分出 3 个种群。判别函数 1 和 2 分别解释总体变异的 42.6% 和 31.0%(图 1)。

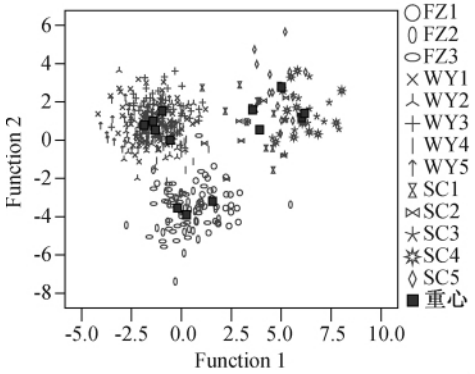


图 1 各样点样本形态标记判别式分析结果

Fig. 1 The discriminatory analysis result of morphometric variables from each sampling polint

将所有个体形态变量数据归入对应的 3 个种群,进行交互验证,得出能准确落入这 3 个种群的个体百分率分别为 96.7%、98.1%、93.8%。

采用各样点重心的判别函数值进行聚类分析。距离 - 分类方法的选择对结果的影响不大,采用 SPSS 系统默认常用的欧式距离 - 类平均法。3 个种群得到很好区分,并分类层次鲜明,准确地揭示了 3 个种群间的遗传关系(图 2)。

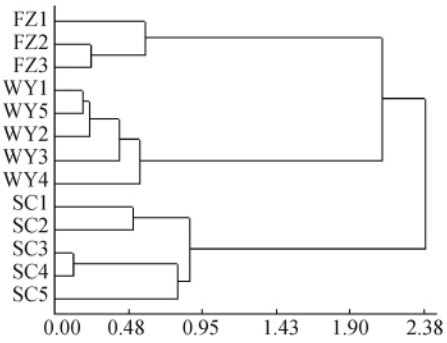


图 2 判断分析样点重心聚类

Fig. 2 The result of cluster analysis with center of gravity of each samplig point

2.2 主成分分析 - 聚类分析联用

根据特征根大于 4,得到 2 个主成分。主成分 1 包括翅长、翅宽、第 4 背板长、第 5 背板绒毛带长、第 6 背板毛长、第 4 腹板长、蜡镜斜长、第 7 腹板长、基跗节长宽、翅脉角 A4、B4、D7、K19、G18,解释了总体变异的 41.076%。主成分 2 包括肘脉指数、后翅钩数、蜡镜间距,解释了总体变异的 15.119%。每个变量负荷值的绝对值大于 0.6。

主成分分析结果准确度不高,有样点被分入错误种群,且遗传关系也存在偏差。各样点的因子得分图中只有 WY 的 5 个样点分布集中,SC 和 FZ 各样点分部较为分散,分类结果人为因素影响较大。特别是 SC1 样点很容易被分到 FZ 中(图 3)。采用各样点主成分 1 和主成分 2 的因子得分作为数据模块进行聚类分析,得到的结果错误:将 SC1 错分入 FZ 中,且 SC 与 FZ 蜜蜂关系较近,WY 蜜蜂离开较远(图 4)。

2.3 聚类分析

用 13 个样点所有形态变量的平均值,使用 SPSS 系统默认常用的欧式距离 - 类平均法聚类,得到的结果是错误的:SC2、SC3、SC4、SC5 聚为一类,FZ1 单独聚为一类,其余样点聚为一类,并且分类层次性不强(图 5)。

解决的思路有二,选用判别能力强的形态变

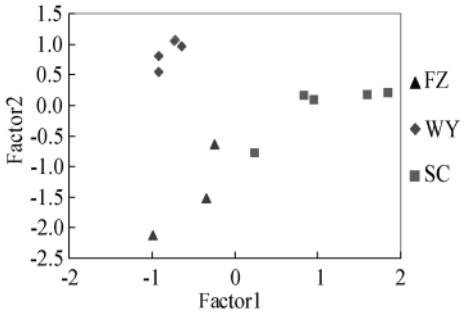


图 3　主成分分析三个样点因子得分散点图

Fig. 3　Scatter of factor scores of each sampling point

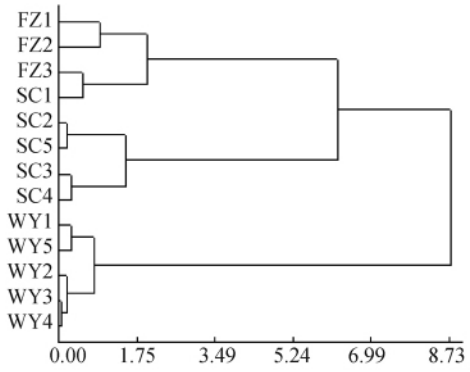


图 4　各样点因子得分聚类

Fig. 4　The result of cluster analysis with factor score of each sampling point

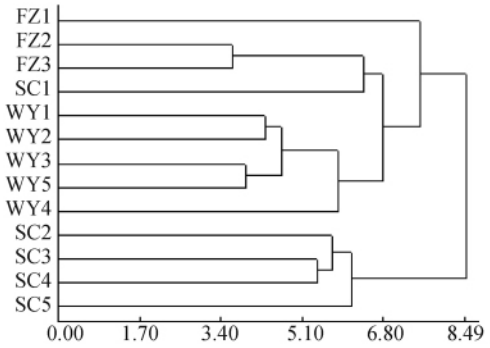


图 5　所有形态变量进行欧式距离 – 类平均法聚类结果

Fig. 5　The result of cluster analysis by Euclidean-Cluster-Average with all of the morphometric variables

量;寻找合理的距离和分类方法组合。对各种距离 – 分类方法组合尝试,选出认为结果较合理的组合方式。这里选择卡方距离 – 离差平方和法,这个聚类结果很好的将 3 个种群加以区分,层次性强,较好的揭示了 3 个种群间的亲缘关系(图 6)。用各样点逐步判别中进入判别函数的形态变量的平均值,使用卡方距离 – 离差平方和法聚类,聚类结果也较好(图 7)。

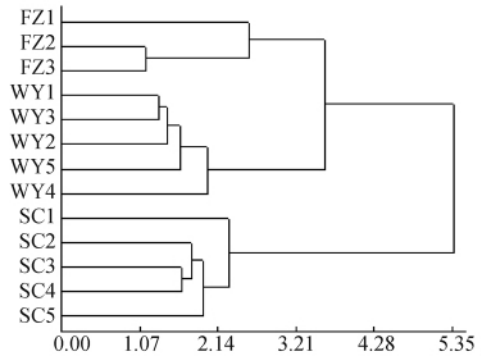


图 6　所有指标卡方距离 – 离差平均和法聚类结果

Fig. 6　The result of cluster analysis by Chi-square-Derivate Square with all of the morphometric varibles

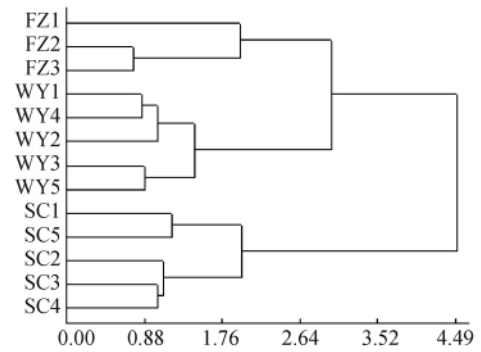


图 7　15 个指标卡方距离 – 离差平方和法聚类结果

Fig. 7　The result of cluster analysis by Chi-square-Derivate Square with the 15 morphometric variables

3　讨论

(1)逐步判别分析适合作为蜜蜂形态遗传分析的统计方法,用于中华蜜蜂种群的划分和遗传关系分析。从统计学的角度,判别分析的目的是对已知分类的数据建立由数值指标构成的分类规则,然后把这样的规则应用到未知分类的样本去分类。作者借助这个方法的原理,假定每个样点为一个分类单位,建立判别函数,不同样点的形态相似性可通过样点重心的判别函数值得以体现,进而进行分类。逐步判别分析中的 Wilks' Lambda 统计量,充分考虑每个变量组内和组间差异的关系,提高种群划分能力。该统计量为组内离差平方和与总离差平方和的比值,使用该方法

时系统会首先纳入使该统计量减小最多的变量。也就是让总体变异尽可能的大,组内变异尽可能的小,两者达到比值最小。这个过程本质上也是对形态变量进行选择,而这样的形态标记正好也是用于种群划分的理想标记。

(2)主成分分析进行分类研究的原理是提取出的每个因素都是原来多个指标的线性组合,彼此相互独立,又能综合反映原来多个指标的大部分信息。作者是借助提取的主成分与各形态指标关系的表达式,建立第一主成分和第二主成分的二维图,根据不同样点在图中的相对位置进行分类分析。主成分分析效果的评价是采用 KMO 统计量,比较变量间的简单相关性和偏相关性的大小,越接近 1,效果越好。在进行蜜蜂形态分析时,影响主成分分析分类结果准确性的原因有:一是提取主成分的效果不够理想,二是忽视个体间差异。在形态标记较多的情况下,往往 KMO 值较低,提取效果不够理想。提出的第一和第二主成分解释所有变异的百分比不高。一般采用各样点的形态指标的平均值进行分析,忽视了个体间的差异,而仅仅追求样点间的差异,缺少对变量的筛选过程。

(3)直接采用聚类分析进行蜜蜂形态数据的处理,不适合研究中华蜜蜂种群分化和亲缘关系鉴定问题。虽然在生物学领域,聚类分析最常用于系统发育研究,生成的树状图代表了可能的进化关系。但该方法是一种探索性统计分析方法,理论背景尚不完备。选用不同计算参数,对于微小差异的样本,也许会得到千差万别的结果。本文采用聚类方法时,通过选用判别能力强的形态变量,寻找合理的距离和分类方法组合等操作,对该方法进行改进才能得到准确结果。聚类分析本身目的性不强。在做之前并不清楚各观察对象属于哪一类,甚至于可以分为几类都不清楚,不一定能表达各个样本间的真实关系(唐启义和冯明光,2007),且无法评估得到结果的准确性。

致谢:四川省蜂业管理站王顺海副站长为本项研究提供四川中华蜜蜂样本,特此感谢!

参考文献(References)

Kandemir I, Kence M, Kence A, 2005. Morphometric and electrophoretic variation in different honeybee (*Apis mellifera* L.) populations. *Turk J. Vet. Anim. Sci.*, 29: 885—890.

Radloff SE, Hepburn HR, Hepburn C, Fuchs S, Otis GW, 2005a. Multivariate morphometric analysis of *Apis cerana* of southern mainland Asia. *Apidologie*, 36(1): 127—139.

Radloff SE, Hepburn HR, Fuchs S, 2005b. The morphometric affinities of *Apis cerana* of the Hindu Kush and Himalayan regions of western Asia. *Apidologie*, 36(1): 25—30.

Ruttner F, 1988. Biogeography and Taxonomy of Honeybee. Berlin: Springer – verlag. 66, 284.

谭垦,张炫,和绍禹,周丹银,2005. 中国东方蜜蜂的形态学及生物地理学研究. 云南农业大学学报,20(3): 410—414.

唐启义,冯明光,2007. DPS 数据处理系统——实验设计、统计分析及数据挖掘. 北京: 科学出版社. 682—690.

Tilde AC, Fuchs S, Koeniger N, 2002. Morphometric diversity of *Apis cerana* Fabr. within the Philippines. *Apidologie*, 31(2): 249—263.

杨冠煌,许少玉,匡邦郁,1986. 东方蜜蜂在我国的分布及其亚种分化. 云南农业大学学报,4(1): 89—92.

朱翔杰,周冰峰,吴显达,徐新建,王青,2009a. 龙栖山中华蜜蜂种群分化形态遗传分析. 福建农林大学学报,38(1): 75—78.

朱翔杰,周冰峰,吴显达,徐新建,王青,2009b. 大门岛中华蜜蜂种群分化形态遗传分析. 中国蜂业,60(1): 5—7.

朱翔杰,周冰峰,吴显达,徐新建,2009c. 第四届中国畜牧科技论坛论文集——武夷山中华蜜蜂种群形态遗传分析,北京: 中国农业出版社. 454—459.