第22卷 第3期
1998年9月

水生生物学报
ACTA HYDROBIOLOGICA SINICA

Vol.22, No.3
Sept., 1998

野生鲫和五个金鱼品种的判别分析和聚类分析

梁前进 彭奕欣 余秋梅[1)]

(北京师范大学生物系 北京 100875)

摘要 对野生鲫鱼(*Carassius auratus*)和五个金鱼品种(*Carassius auratus* var.)的亲缘关系进行了比较研究。(1)两组判别分析表明,以17个形态变异性状的指标衡量野生鲫鱼和金鱼以及金鱼各品种之间的差异,说明金鱼的种下分化是明显的($F \gg F_{0.01}$)。(2)对6个品种(包括鲫鱼)各20尾鱼(计120个个体)进行17个指标的聚类分析,结果大部分个体(占各品种的50～100%)可依品种归类,但也有一部分个体"误分",说明这些品种作为一个物种的成员其分化程度有限,而且指标选择和个体差异问题均需考虑。(3)用逐步判别分析找出17个指标中最主要的14个指标后,各品种回判正确率达85%以上。(4)剔除次要(差异不显著)指标和一个品种(红头蛋白)不存在的指标进行聚类分析,得出了野生鲫鱼、金鱼的系统关系。多因子方差分析结果表明聚类指标的选择合理。

关键词 野生鲫鱼,金鱼,判别分析,聚类分析

原产于中国的金鱼(*Carassius auratus* var.)是名贵观赏鱼类,也是研究种内演变的好材料[1]。迄今为止,已有许多学者验证了金鱼起源于野生鲫鱼(*Carassius auratus* L.)的推断[2],取得了相当一致的结果。随着计算机技术的发展,多元分析和数值分类方法逐渐应用于育种和进化研究,并在不同的生物类型中获得了成功[3—5]。本研究的目的是,利用两组判别分析来比较野生鲫鱼和金鱼的五个代表品种两两间的差异水平;在对所有个体和性状指标的观察数据进行聚类分析,确定可按品种归类后,用逐步判别方法找出最主要的分类指标,用它们进行聚类分析,得出野生鲫鱼和各金鱼品种的系统关系,同时用多因子方差分析进行分类指标检验。

1 材料和方法

1.1 性状值的测量 实验鱼及处理同前文[6]。为使本研究与前人工作有连续性,性状的选择与测量用陈桢的方法进行。其中的各性状(长度或距离取其与头长之比值作性状值)代号:1:体长;2:体高;3:背鳍鳍条数;4:吻和背鳍距离;5:背鳍长;6:吻和胸鳍距离;7:胸

1) 现在国家统计局城调队工作。

本工作承刘来福、李洪兴、黄远樟、朱勇珍、李晓春、何清、程晓莉、吴贤柱和庞尔丽等老师、同学的指导和帮助,谨此致谢。

1995-04-04收到;1997-02-24修回。

鳍长;8:吻和腹鳍距离;9:腹鳍长;10:吻和臀鳍距离;11:臀鳍长;12:最长尾鳍鳍条长;13:最短尾鳍鳍条长;14:眼眶直径;15:侧线上鳞片数;16:侧线上方鳞片行数;17:侧线下方鳞片行数。由于蛋鱼无背鳍,为比较背鳍相关性状,将性状 3、4 和 5 定为零。

为叙述方便,将野生鲫鱼称为一个品种。各品种代码见表 1。

表1　野生鲫鱼和若干金鱼品种的代码和英文名称

Tab.1　The codes of the wild crucian and some goldfish varieties

	野生鲫鱼	金鲫	草金鱼	红文鱼	红龙睛	红头蛋鱼
	The wild crucian	Golden crucian	Grass goldfish	Red wen goldfish	Red dragoneye goldfish	Red-head oval goldfish
代码	1	2	3	4	5	6

1.2　两组判别分析　求两品种的判别函数

$$y = \sum_{j=1}^{p} b_i x_i$$

(p:性状数,b_i:判别系数,x_i为性状;$i = 1,2,\cdots,17$)[7]再作 F 检验[8]

1.3　对全部观察数据的聚类　运用 SAS 统计分析软件对 6 个品种各 20 尾鱼(计 120 个个体)进行 17 个性状指标的聚类分析,所用距离为欧氏距离。

1.4　逐步判别分析　运用 SAS 统计分析:软件进行 6 个品种(各 20 个个体)间 17 个性状指标的逐步判别分析,找出一组能最充分揭示各品种间差异的指标变量,从而舍弃其他包含信息最少的指标变量。

用选出的性状指标对 6 个品种进行聚类分析:运用 SAS 统计分析软件,以逐步判别选出的最主要指标对 6 个品种进行聚类,得出各品种间的系统关系。所用距离为欧氏距离。

1.5　方差分析　经过方差分析确认用于聚类分析的各指标的重要性(可靠性)。

2　结果

共测试了 6 个品种各 20 个个体的 17 个性状。

2.1　两组判别分析　两两品种比较,参比的两品种分别称为 A、B。运算结果见表 2。F 检验自由度 $df_1 = p = 17$, $df_2 = N_1 + N_2 - p - 1 = 22$(公式中,$N_1$、$N_2$分别为 A、B 两组样品的个体数)。共作了 $C_6^2 = 15$ 次两组判别分析,F 检验结果均为极显著(当 $df_1 = 17, df_2 = 22$ 时,$2.03 < F_{0.05} < 2.22, 2.75 < F_{0.01} < 3.12$)。

2.2　对全部观察数据的聚类　先将 120 个个体看作 120 个类别。第 1 步先将第 93 个个体和第 98 个个体聚为一类,于是总类别数变成 119 类;第 2 步将第 117 和 120 个个体聚为一类,于是总类别数变成 118 类……当聚类进行到 第89 步,总类别数成为 31 类时,已有 65% 的品种 1 被聚成一类(A);当聚类进行到第 100、104、107、111 和 116 步时,分别有 60% 的品种 3、60% 的品种 2、50% 的品种 4、75% 的品种 5 和 100% 的品种 6 个体被聚成一类(分别称 B、C、D、E 和 F 类)。另外,除包含一个个体数占该品种总个体数 50% 以上的品种的类别外,尚有 5 个品种 1 的个体、2 个品种 3 的个体、4 个品种 4 的个体和 5 个品种 5 的个体在第 114 步归作单独一类(M)。各类别间的距离和它们的相互关系见表 3。表 4 表明了按主要品种(相应类别包含其 50% 以上个体数的品种)归类的有效率。

表2 各品种的两组判别分析结果

Tab.2 The results of two-group discriminant-function analysis between every two varieties.

A	B	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}	b_{16}	b_{17}	D^2	F值及显著性
1	2	14.345	−16.580	0.482	0.908	−10.867	0.488	3.731	−2.491	−9.880	7.011	−14.188	−15.104	−16.403	17.125	0.057	−0.020	0.242	316.99	107.95**
	3	4.713	−2.157	0.127	0.701	−4.262	−6.941	0.571	−0.551	−8.350	2.201	3.245	−6.725	0.688	10.664	0.001	−0.303	0.125	152.71	52.00**
	4	6.797	1.428	−0.295	1.271	0.820	−10.518	−0.946	−5.927	3.579	1.936	−4.658	−6.258	−2.032	35.326	−0.463	−0.726	0.253	502.74	171.21**
	5	7.105	−5.987	−0.093	−0.970	−4.180	5.318	0.507	6.238	−17.552	3.013	10.816	−2.022	−10.772	−11.733	0.818	0.277	−0.286	544.55	185.45**
	6	1.516	−37.574	8.250	5.258	−35.980	30.459	59.611	−10.449	−2.720	9.385	−61.728	11.040	−11.710	−107.365	1.798	−1.792	1.441	6242.57	2125.95**
2	3	2.238	12.949	−0.046	6.740	−13.523	0.071	0.156	−0.177	−4.425	−4.741	7.219	−14.297	−4.314	−3.496	0.041	−0.159	−0.347	119.36	40.65**
	4	8.866	−12.122	0.314	−9.989	2.708	0.002	−3.747	1.298	2.179	8.163	−11.562	1.132	−2.229	−7.155	−0.674	−0.587	0.177	319.13	108.68**
	5	12.624	1.218	−2.035	24.992	22.017	0.742	−20.768	7.842	−11.561	−0.250	−27.854	−0.819	−3.274	−35.854	0.969	1.503	−0.062	1107.50	377.17**
	6	15.096	110.405	6.723	208.516	−61.665	−1.253	−102.396	2.492	8.837	−35.468	0.214	−7.518	1.459	62.388	−4.363	1.605	−3.644	16426.88	5594.29**
3	4	−1.330	−7.794	0.462	−4.574	−1.584	−1.319	−6.774	0.141	4.057	13.258	−18.417	0.823	−0.100	12.492	0.303	−0.547	0.154	243.26	82.84**
	5	−1.589	1.288	0.195	−0.476	−2.594	3.181	−6.777	1.022	9.066	5.430	−15.852	2.318	−0.381	−26.299	0.029	0.188	−0.898	254.98	86.84**
	6	1.104	55.347	10.745	24.848	−16.054	−63.587	−3.089	−0.269	2.096	−13.060	−4.818	0.073	8.537	37.120	−2.914	−0.305	−0.830	7981.31	2718.09**
4	5	−0.998	−5.756	0.397	−2.271	−0.288	4.971	−1.104	2.621	4.176	2.497	−6.402	2.772	1.262	−25.353	0.636	0.714	−0.412	217.20	73.97**
	6	−3.428	−41.347	2.978	18.721	11.335	26.616	−18.251	6.371	14.536	3.824	−13.693	7.640	−1.268	−45.357	1.763	0.571	−0.087	3795.00	1292.41**
5	6	−13.251	−12.638	2.060	20.682	17.684	13.790	28.755	−39.972	4.924	19.663	−18.456	4.604	0.400	9.968	−0.269	−3.684	1.668	2847.93	969.88**

表3　各类别间的距离

Tab.3　The distances between every two clusters

聚类步骤	结合类别	距离	新类别
101	A和C	0.168581	(AC)
109	(AC)和B	0.242396	(ABC)
113	(ABC)和D	0.302508	(ABCD)
117	M和E	0.635711	(EM)
118	(ABCD)和(EM)	1.018244	(ABCDEM)
119	(ABCDEM)和F	1.523249	(ABCDEFM)

表4　按主要品种归类的有效率

Tab.4　The deficiency of the classification according to the main varieties.

类　别		A	B	C	D	E	F	M
总个体数		15	19	19	16	15	20	16
主要品种	品种	1	2	3	4	5	6	—
	个体数	13	12	12	10	15	20	—
次要品种	品种	3	1 4	1 2	2 3	—	—	1 3 4 5
	个体数	2	1 6	1 6	2 4	—	—	5 2 4 5
主要品种包含率		86.7%	63.2%	63.2%	62.5%	100%	100%	—

2.3　逐步判别分析　共涉及 6 个品种各 20 个个体的观察数据。对 17 个性状指标进行逐步选择法分析,指标的引入和保留以 0.1500 作为显著性水平默认值。每个品种观察数据占总量的比例是 20/120=0.16667。选择共进行了 15 步,剔除了 3 个性状(性状 6、7 和 8),

表5　逐步判别剔除的性状指标

Tab.5　The characters removed by Stepwise selection

性状指标	F 值	Prob>F	容忍值(Tolerance)
6	1.218	0.3065	0.0445
7	1.543	0.1833	0.0445
8	0.899	0.4850	0.0415

表6　逐步判别选出的性状指标

Tab.6　The important characters entered by Stepwise selection

选择步骤	选出性状指标	F值	Prob>F
1	3	1319.826	0.0001
2	1	46.954	0.0001
3	9	84.027	0.0001
4	14	50.390	0.0001
5	5	28.466	0.0001
6	4	19.206	0.0001
7	13	11.158	0.0001
8	15	5.389	0.0002
9	11	5.880	0.0001
10	10	4.448	0.0010
11	2	5.566	0.0001
12	16	3.656	0.0044
13	17	3.507	0.0058
14	12	1.998	0.0853

其余 14 个性状指标能充分揭示各品种间差异，被选作聚类分析指标(表 5、6)。

用二次判别函数在各品种间进行选出性状指标的交互有效性 (Cross-validation) 分析。所用通用平方距离函数为：

$$D_j^2(x) = (x - \bar{x}(x)_j)' \, cov^{-1}(x)_j (x - \bar{x}(x)_j) + ln|cov(x)_j|$$

观察值 x 个体归属于品种 j 的后验概率为：

$$Pr(j|x) = e^{-0.5D_j^2(x)} / \sum_k e^{-0.5D_k^2(x)} .$$

回判结果见表 7。

表7 对主要聚类指标的回判结果(用二次判别函数)

Tab.7 Cross-validation Summary using Quadratic Discriminant Function

来源品种 \ 归入品种 / 观察值及归类概率(%)	1	2	3	4	5	6	合 计
1	18	0	1	1	0	0	20
	90.00	0.00	5.00	5.00	0.00	0.00	100.00
2	0	20	0	0	0	0	20
	0.00	100.00	0.00	0.00	0.00	0.00	100.00
3	2	0	17	1	0	0	20
	10.00	0.00	85.00	5.00	0.00	0.00	100.00
4	0	0	0	20	0	0	20
	0.00	0.00	0.00	100.00	0.00	0.00	100.00
5	0	0	0	3	17	0	20
	0.00	0.00	0.00	15.00	85.00	0.00	100.00
6	0	0	0	0	0	20	20
	0.00	0.00	0.00	0.00	0.00	100.00	100.00
合计	20	20	18	25	17	20	120
百分率	16.67	16.67	15.00	20.83	14.17	16.67	100.00

可见用这 14 个主要性状指标进行聚类，对 6 个品种的回判正确率均达 85% 以上。

2.4 用选出的性状指标对各品种进行聚类分析 这里用选出的性状指标对 6 个品种进行聚类分析。考虑到性状 3、4、5 是涉及背鳍的，而品种 6(红头蛋鱼)缺背鳍，所以尽管它们

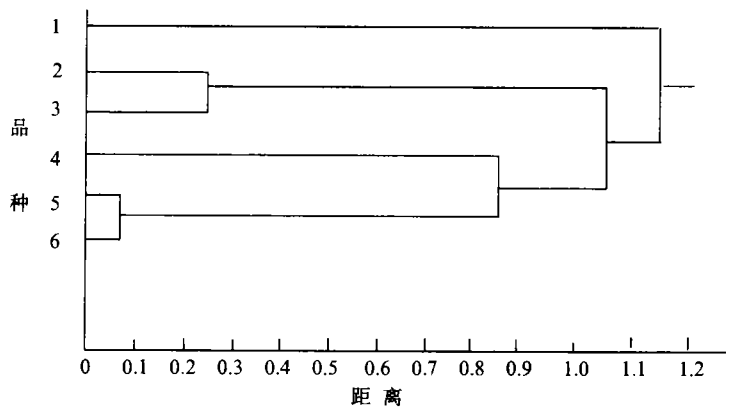

图1 各品种的聚类结果

Fig.1 The result of cluster analysis for the varieties

是其余 5 个品种间分类比较的有效指标,这里还是舍弃了它们,于是实际应用的是其他 11 个性状指标。各品种间的系统关系见图 1。

2.5 方差分析 用 6 个品种共 120 个个体的观察数据对 17 个性状指标进行方差分析,发现只有性状 6 对聚类分析总体无显著影响,其余 16 个性状均有显著影响,它们包括了上述选出的聚类指标的全部(表 8)。处理间和重复间自由度分别为 5 和 119。

表8 方差分析结果

Tab.8 The results of variance analysis

性状	1	2	3	4	5	6	7	8	9
F	59.65	3.43	1319.83	264.40	177.96	1.93	9.36	4.44	43.63
Prob>F	0.0001	0.0064	0.0001	0.0001	0.0001	0.0945	0.001	0.0010	0.0001
性状	10	11	12	13	14	15	16	17	
F	26.45	36.37	29.63	12.92	78.09	13.25	2.70	5.46	
Prob>F	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0243	0.0002	

3 讨论

3.1 本文在研究形态变异时,各长度性状都按其测量值与头长之比值作为观察值。这是因为鲫鱼和金鱼的头长在全身的比例相当稳定[1]。本文起初所选 17 个性状是早在 1959 年陈桢就已发表的用以有效地进行金鱼家化中形态变异分析的性状指标。17 个性状的方差分析(表 8)也表明这些性状(性状 6 除外)几乎都是影响显著的分类指标,但考虑到逐步判别分析剔除了 6、7、8 三性状指标,所以用其余 14 个性状更可靠。另外,由于性状 3、4 和 5 涉及品种 6(红头蛋鱼)不存在的背鳍,尽管逐步判别和方差分析表明它们对于其余品种的分类影响显著,这里最终也没有用它们作聚类指标。

3.2 金鱼由野生鲫鱼演化而来,仅有 1000 余年历史,而身体各部变化之大,成为生物进化、尤其是种下进化的直接证据。本研究的两组判别分析表明,野生鲫鱼和金鱼,以及金鱼各品种间差异极显著(表 2),因此,可认为金鱼已分化出明显不同的品种。从判别系数 b_i 绝对值相对大小上,可估计在区别两个特定品种时,哪些性状更重要些。比如选前 5 个重要性状,可发现在品种 1(野生鲫鱼)和 2(金鲫)的比较中,它们是性状 14、13、2、12 和 1。不同的品种对间,显示其差异的主要性状不同,因为品种形成时,各有特定的形态演变。本文最终进行的聚类分析(图 1)是采用总体体现 6 个品种差异的性状指标的结果。

3.3 从主要品种归类有效率(表 4)来看,初选的 17 个性状指标的测量结果大体可做自然品种的分类指标,但存在少数个体的"误分"现象。除了个体有多样性差异外,所用性状显然需要精选,逐步判别分析就是要达到这个目的。另从没有任何品种可作主要品种(包含 50% 以上个体数)的 M 类的组成(含品种 1、3、4、5 的少数个体)来看,野生鲫鱼和各品种金鱼间相似性还很大,以至一些个体形态分化不明显(除去测量误差),毕竟它们都属于一个物种(鲫鱼种)。

3.4 除去逐步判别剔除的和品种 6(红头蛋鱼)不存在的性状外,用其余 11 个主要性状指标做出了 6 个品种的系统关系图(图 1)。聚类过程中,首先是品种 5 和 6 归为一类,说明它们演化关系近,这与蛋白成分分析结果一致[6]。第二层次品种 2 和 3 归为一类,也与进化历程相符。这两个品种是从野生鲫鱼最早演化来的(先是金鲫,然后发展为草金鱼),而两

者除单双尾鳍之差外，形态十分相近，有人将其视为同一品系[9]。第三层次是品种 4 与品种 5 和 6 归为一类，与传统看法一致，即龙种鱼和蛋种鱼等品系是由文鱼分化来的，在形态性状上有相近之处。前人研究表明，文鱼由草金鱼演变而来。本文的聚类结果还表明，野生鲫鱼和金鱼之间在形态性状上已相差甚远，所以它最后单作一"类"同其他几个金鱼品种相联系；从形似野生鲫鱼的金鲫、草金鱼到典型的金鱼（文、龙、蛋各种品系）也有典型差别，从而文、龙、蛋（对应于品种 4、5、6）聚为一类（第三步聚类），而金鲫和草金鱼在另一类。

需要说明的是，对生物类型进行系统分析，应结合多方面资料，采用多种方法，本文与前文[6]的结论就是一致性和差异性共存的，不难理解，从个体形态和从生化上对生物进行研究，情况不同，但可相互借鉴。另外，即使是同样的数据，不同的数理分析方法所得结果也会有别，本文只是 SAS 统计分析软件的运用结果。

参 考 文 献

[1] 陈 桢. 金鱼的家化与变异. 北京：科学出版社. 1959
[2] 王春元等. 金鱼染色体组型的研究 I. 遗传学报，1982，**9** (3)：238—242
[3] 伍时照等. 早籼稻优质品种数量性状的遗传距离与聚类分析. 华南农业大学学报，1988，**9** (2)：56—62
[4] 陈 斌. 星天牛属的数值分类研究初探（鞘翅目：天牛科）. 动物分类学报，1989，**14** (1)：96—103
[5] 赵铁桥. 叶尔羌条鳅的数值分类和种下分化. 动物分类学报，1983，(4)：438--446
[6] 梁前进等. 野生鲫鱼和五个金鱼代表品种的肌肉蛋白电泳分析. 动物学研究，1994，15(2)：68—75
[7] 王鉴明. 生物统计学. 北京：农业出版社. 1988
[8] 罗积玉等. 经济统计分析方法及预测. 北京：清华大学出版社. 1987
[9] 王占海等. 金鱼及热带鱼的饲养. 上海：上海科技出版社. 1982

THE DISCRIMINANT AND CLUSTER ANALYSES OF THE WILD CRUCIAN CARP (*CARASSIUS AURATUS*) AND FIVE REPRESENTATIVE VARIETIES OF GOLDFISHES (*C. AURATUS* VAR.)

Liang Qianjin　Peng Yixin and Yu Qiumei

(*Department of Biology, Beijing Normal University, Beijing*, 100875)

Abstract The two-group discriminant-function analysis, stepwise discriminant analysis and numerical taxonomy were used for the comparative study or the intraspecific evolutionary relationship among the wild crucian carp (*Carassius auratus*) and five varieties of goldfish (*C. auratus* var.). Seventeen mophorlogical characters were analysed, and the divergences among the wild crucian carp and the varieties of goldfish are extremely significant, indicating the presence of intraspecific differentiation. The cluster analysis revealed, with the most important characters selected by the stepwise discriminant analysis, that the wild crucian carp and the goldfish are phyletically corre lated. The variance analyses proved that the characters used for cluster analysis are reasonable.

Key words Wild crucian carp (*Carassius auratus*), Goldfish (*Carassius auratus* var.), Discriminant analysis, Cluster analysis.