

RAPD 分析中遗传距离计算方法的比较

吕雪梅 杨关福 张细权

(华南农业大学动物科学系, 广州 510642)

摘要 为了更有效地应用 RAPD 分析这一手段研究种群遗传变异, 以广州市力康鸡场 7 个蛋鸡品系为材料, 根据 RAPD 测定的结果, 采用 Nei, Rogerst 和 Hedrick 的遗传距离公式计算上述 7 个品系彼此间的亲缘关系, 同时提出新的遗传距离计算公式。结论认为 Rogers 的公式较适用于 RAPD 数据的分析, 文中提出的公式跟其他距离公式存在显著相关, 但有待进一步验证。

关键词 RAPD 分析; 遗传距离

中图分类号 S 828. 2

力康白壳蛋鸡 a、d、e、h、n、 l_1 、 l_2 7 个品系分别来源于 80 年代从国外引进的 3 个品种: 科伯特 Corbett(a、d、e)、哈巴德 Hutbard(h)、海兰 Hy-Line(n、 l_1 、 l_2), 由于各品系在引进以前选育历史的原始资料不完整, 因此希望利用遗传标记多态性分析, 更好地了解系间的亲缘关系。

RAPD 是 90 年代建立的遗传分析技术, 进行 RAPD 分析时可用的引物数量大, 因此, 在理论上利用一系列引物可以使检测区域几乎覆盖整个基因组。另外, 由于 RAPD 技术还具有操作简单快捷、无需预先知道引物和模板 DNA 序列等优点, 它已被广泛应用于多方面的研究中, 尤其在动植物群体结构分析中应用较多(Cargill et al, 1995; Lu et al, 1996)。作者曾用 RAPD 标记分析家鸡几个品种遗传多态性的结果, 能准确反映它们的起源及选育历史(吕雪梅等, 1996)。但由于绝大多数 RAPD 标记是显性遗传, 目前用 RAPD 技术进行亲缘关系分析时没有一致的统计方法, 本研究用 4 种遗传距离计算方法分析上述 7 个品系群体间差异, 并对计算结果进行了对比和讨论。

1 材料与方法

1.1 试验动物

试验鸡群包括力康公司白壳蛋鸡科伯特 a 系、d 系、e 系, 哈巴德 h 系, 以及由海兰 W-36 分化而成的 n 系、 l_1 系和 l_2 系。其中, l_2 系是 l_1 系的公鸡与海兰 W-36 父母代母系的母鸡交配建立的品系。

1.2 血液样品采集及模板 DNA 的制备

分别从 a、d、e、h、n、 l_1 、 l_2 7 个品系每个品系随机挑出 10 只母鸡、10 只公鸡采取血样 1 mL/只, EDTA 抗凝, 用于基因组 DNA 的提取。从鸡的血样中提取 DNA 根据孟安明(1993)的方法, 采用常规酚/氯仿抽提法。抽提后的 DNA 溶液根据其在紫外分光光度计或琼脂糖电泳所检测到的浓度取适当体积稀释至浓度约为 25 ng/ μ L。

1.3 RAPD—PCR 扩增

用 6 个 10 碱基随机引物和 1 对双引物组合(编号为 OPB8、OPB13、OPG4、OPH4、OPH19、

OPI4 和 OPH5+OPH7) (OPERON 公司) 分别对 7 个品系的个体 DNA 样品 (20 个/系, 10 公、10 母) 进行 PCR 扩增。PCR 扩增反应总体积 20 μ L, 其中包括模板 DNA 80 ng (或双蒸水作阴性对照)、10 \times PCR 缓冲液 2.0 μ L、dNTP 2.0 mmol/L、MgCl₂ 2.5 mmol/L、引物 5 pmol、Taq DNA 聚合酶 1.2 U。反应条件: 94 $^{\circ}$ C 变性 45 s、35 $^{\circ}$ C 退火 1 min、72 $^{\circ}$ C 延伸 2 min, 38 个循环, 72 $^{\circ}$ C 后延伸 10 min。扩增产物于含溴化乙锭的 2% 琼脂糖凝胶中电泳 2~3 h, 电压 ≤ 2 V/cm。电泳结束后在紫外光下观察结果并拍照。

1.4 数据统计分析

1.4.1 多态性带型频率 计算品系中出现某多态性条带的个体数占该品系实验个体总数的比率。

1.4.2 遗传距离 分别使用了 3 种较常用的公式及作者推导的 1 个计算公式, 计算了 7 个品系间的遗传距离。

(1) Nei 的遗传距离 DN (Nei, 1972; Bonnin et al, 1996; Smith et al, 1996): 群体 X 、 Y 的遗传同一性 (Genetic Identity, I)

$$I = \sum_i x_i y_i / \sqrt{\sum_i (x_i)^2 \sum_i (y_i)^2},$$

群体 X 、 Y 的遗传距离 $DN = -\ln I$, (公式 1)

其中 x_i 、 y_i 表示第 i 条带分别在群体 X 、 Y 中出现的频率。

(2) Rogers 的遗传距离 DR (Rogers, 1972):

$$DR = \frac{1}{r} \sqrt{\frac{1}{2} \sum_i (x_i - y_i)^2},$$

$DH = -\ln I$, (公式 2)

其中 r 为群体 X 、 Y 中所有引物产生的条带总数, x_i 、 y_i 的定义同上。

(3) 遗传距离 DH (Hedrick, 1971; Zhang 等, 1995):

$$I = \frac{1}{N} \sum_{i=1}^n \frac{2v_i^{(1)} v_i^{(2)}}{[v_i^{(1)}]^2 + [v_i^{(2)}]^2},$$

$DH = -\ln I$, (公式 3)

其中 N 表示对比的两个群体中所有引物产生的条带总数, $v_i^{(1)}$ 、 $v_i^{(2)}$ 分别为条带 i 在群体 1、2 中出现的频率。

(4) 本文发展的遗传距离公式 DS :

实质上, 此公式是以共同带率 (Nei et al, 1979; Jeffreys et al, 1987) 的计算方法为基础的, 以某特定条带在群体中出现的频率的比较, 代替了共同带率计算中 (见公式 5) 群体混合 DNA 样品间共有条带数的比较。计算方法如下:

$$\text{相似性指数 (Similarity Index, } SI) \quad SI = 2 \sum_{i=1}^n q(xy) i / (\sum_{i=1}^n q(x) i + \sum_{i=1}^n q(y) i),$$

$DS = 1 - SI$, (公式 4),

其中 n 为在群体 X 、 Y 中出现的条带的总数; $q(xy) i$ 为第 i 条带在群体 X 或 Y 中出现的两个频率数值中的最小值; $q(x) i$ 、 $q(y) i$ 分别为 i 条带在群体 X 、 Y 中出现的频率。

1.4.3 聚类分析 分别以 4 种方法计算出的 a 、 d 、 e 、 h 、 n 、 l_1 、 l_2 7 个品系间遗传距离 DN 、 DR 、 DH 、 DS 作为距离系数, 采用算术平均非加权对群法 (Unweight Pair-group Method with Arithmetical Average, UPGMA) 对上述 7 个品系进行聚类分析, 并绘出聚类树形图。

1.4.4 简单线性相关分析 应用 MTB 软件对 RAPD 分析的 4 种遗传距离 DN 、 DR 、 DH 、 DS 进行相关分析。

2 结果与分析

2.1 多态性带型的频率

各引物的碱基序列, 扩增产物数和多态性产物数见表 1。扩增产物条带数共 40 条, 其中 14 条有多态。表 2 列出了这 14 条带在各品系中出现的频率。

表 1 引物及其产物数目

引 物	序列 5' → 3'	扩增产物数(条)	多态性产物数(条)
OPB8	GTCCACACGG	5	2
OPB13	TTCCCCCGCT	4	2
OPG4	AGCGTGTCTG	8	2
OPH4	GGAAGTCGCC	6	1
OPH19	CTGACCAGCC	5	2
OPI4	CCGCCTAGTC	5	2
OPH5	AGTCGTCCCC	6	3
+			
OPH7	CTGCATCGTG	40	14
总 数			

表 2 多态性产物在品系中出现的频率

引 物	产物编号	品 系						
		a	d	e	h	n	l_1	l_2
OPB8	1	0.75	0.45	0.65	0.60	0.30	0.30	0.10
	2	0.25	1.00	0.80	1.00	0.95	1.00	1.00
OPB13	3	0.50	0.45	0.75	0.10	0.15	0	0
	4	1.00	1.00	1.00	1.00	0.95	0.95	0.95
OPG4	5	0.90	0.90	0.95	0.25	0.40	0.10	0.05
	6	0.70	0.10	0.05	0.80	0.95	1.00	1.00
OPH4	7	0.30	0.05	0.25	0.65	0.50	0	0
OPH19	8	0.50	0.80	0.05	0.25	0.35	0.05	0
	9	0.75	0.50	1.00	0.80	1.00	0.70	0.40
OPI4	10	0.90	1.00	0.95	1.00	0.85	1.00	1.00
	11	0.20	0.10	0.65	0.10	0.50	0.05	0.35
OPH5	12	0.80	1.00	0.75	0.85	0.90	1.00	1.00
+	13	0.90	1.00	1.00	1.00	1.00	1.00	1.00
OPH7	14	0.50	0.50	0.50	0.50	0.50	0.50	0.50

2.2 4 种遗传距离及其相关分析

在 RAPD 分析的报道中, 用来计算群体间遗传距离的方法各不相同(Plotsky et al, 1995; Link et al, 1995; Zhang et al, 1996; Smith et al, 1996), 本文前面所提到的共同带率多用于通过各群体混合 DNA 样品间带型的比较(Yu et al, 1993)来反映群体间相似性, 而以不同的条带在各群体的个体中出现的频率来计算遗传距离时, 较多使用的是本文所用的 3 种遗传距离计算方法, 分别为 DN 、 DR 和 DH 。此外, 作者还以共同带率的计算方法为基础发展了另一种遗传距离的算法, 以 DS 表示(见公式 4)。表 3、表 4 分别是 7 个品系间的 DN 、 DR 、 DH 和 DS 4 种遗传距离计算结果, 可以看出 4 种遗传距离所反映的群体间变异是有所差异的。

而表 5 的简单相关分析结果表明,除 DN 和 DH 基本没有相关外,其余的两两间均有极显著的相关关系,其中以 DS 和 DR 的相关性最高($r=0.921$)。

表 3 RAPD 分析中的系间遗传距离 DN (左下)和 DR (右上)

品系	a	d	e	h	n	l_1	l_2
a		0.057 6	0.057 3	0.059 7	0.069 0	0.073 6	0.081 4
d	0.075 8		0.059 7	0.092 6	0.070 3	0.076 3	0.081 0
e	0.077 2	0.066 0		0.0727	0.068 3	0.086 5	0.090 4
h	0.081 1	0.074 8	0.076 7		0.031 9	0.041 1	0.052 3
n	0.081 6	0.078 9	0.072 0	0.059 2		0.044 7	0.049 9
l_1	0.081 0	0.070 3	0.082 0	0.048 9	0.053 3		0.023 9
l_2	0.088 8	0.075 1	0.086 5	0.055 4	0.056 5	0.030 8	

表 4 RAPD 分析中的系间遗传距离 DS (左下)和 DH (右上)

品系	a	d	e	h	n	l_1	l_2
a		0.063 5	0.069 4	0.063 1	0.054 3	0.135 0	0.149 7
d	0.043 0		0.073 6	0.084 6	0.089 7	0.127 6	0.155 6
e	0.042 7	0.044 2		0.1006	0.078 7	0.130 8	0.163 6
h	0.043 7	0.049 5	0.054 8		0.028 8	0.088 1	0.133 3
n	0.049 1	0.056 3	0.050 2	0.025 6		0.109 1	0.120 4
l_1	0.061 2	0.049 6	0.066 7	0.028 4	0.030 8		0.070 5
l_2	0.070 3	0.055 7	0.071 3	0.040 3	0.038 6	0.013 4	

2.3 聚类分析

聚类分析可以更直观地表现各品系间的关系,以及应用不同的距离系数所产生的不同结果。图 1~4 分别是以 DN 、 DR 、 DH 和 DS 作为距离系数得到的 7 个品系聚类分析树形图。 DR 和 DS 所得到的结果基本一致, DN 的结果与前二者有一定相似之处,即可将 7 个品系聚为两大类, a、d、e 3 个系聚为一类, h、n、 l_1 、 l_2 4 个系聚为另一类, 其中 l_1 和 l_2 最先聚合。这一结果与 7 个品系的来源基本相符: n 系和 l_1 系是海兰蛋鸡的两个品系, 具有较近的亲缘关系, l_2 系是从 l_1 系分化出来的, 两者的相似性程度应是最高的。a、d、e 3 个系同属于一个品种, 亲缘关系比较接近; h 系具有单独的来源。由于不具备 h 系以及其他各系的选育过程详细资料, 所以无法从选育历史进行考证它们间

表 5 4 种距离系数的相关系数

遗传距离	DN	DR	DS
DN			
DR	0.833 **		
DS	0.890 **	0.921 **	
DH	0.263	0.551 **	0.617 **

** $P<0.01$

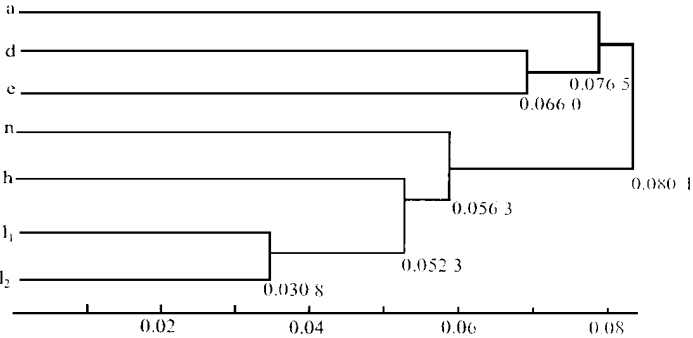


图 1 以 DN 为距离系数的聚类分析树形图

的亲缘关系。但可以肯定的是,它们都具有白来航蛋鸡的血缘,而且具有近似的选育目标,因此 7 个品系虽各有不同的来源,但也存在一定的联系和共同的特点。

另一方面,若将 *DN* 的聚类结果再细分,则会看出它与 *DR*、*DS* 的结果是有差异的。而由 *DH* 得到的聚类分析结果与 *DN*、*DR* 和 *DS* 三者的差异程度更大一些。它将 *l₁*、*l₂* 两系聚为一类,其他 5 个系聚为一类,而 *l₁*、*l₂* 的聚合距离大于 *n* 和 *h*、甚至 *a*、*h*、*n* 3 个系的聚合距离,这一结果与几个系之间的实际情况相差较大。

3 讨论

遗传距离是把两个群体间的遗传差异用单个数值来表示 (Smith, 1977)。假如群体间无差异,则可设其距离为零;但是假如群体在任何基因座位上毫无共同的等位基因,则其距离可设为最大值。一般认为,遗传距离与所比较的群体从同一祖先群体分化以来的时间

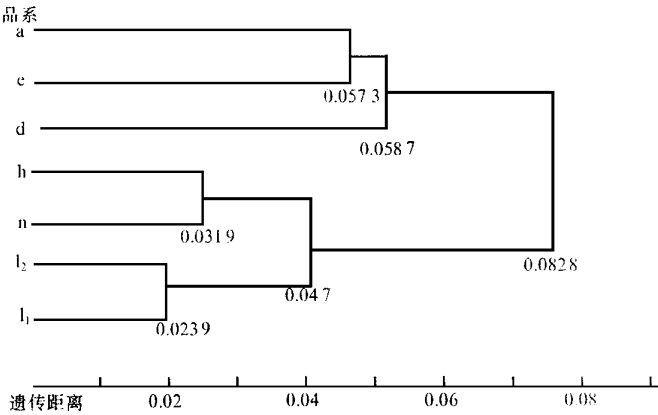


图 2 以 *DR* 为距离系数的聚类分析树形图

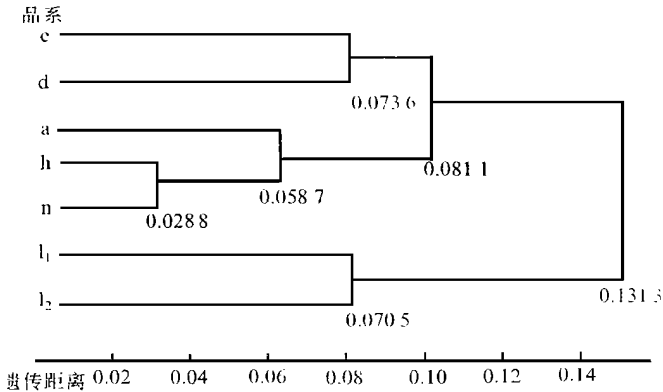


图 3 以 *DH* 为距离系数的聚类分析树形图

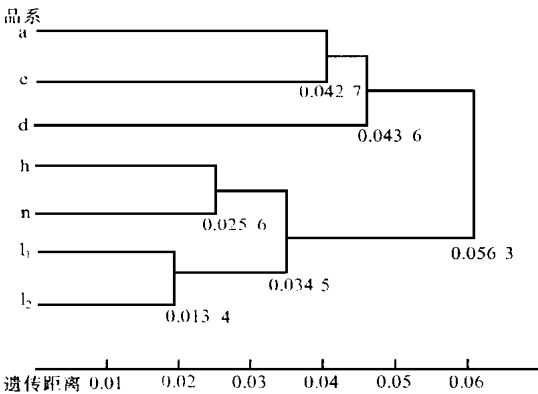


图 4 以 *DS* 为距离系数的聚类分析树形图

有关,因而需要一个说明引起群体分化的突变或漂移等过程的遗传模型。不同的距离有各自不同的原理、假设或模型支持,适用于不同的目的,距离可简单地看作是一种数据缩减机制,或是比较成对远缘群体的一种方法,或是构建群体进化史的基础(Weir, 1994)。

遗传距离估算中应用最广泛的一种方法是由 Nei(1972)提出的。以遗传学观点来看,遗传距离最适宜的测度应是单位长度 DNA 的核苷酸差异或密码子的差数,但很难直接测定。鉴此,Nei 提出了一个以基因频率数据估算每个位点的平均密码子差数的统计方法(见公式 1, DN)(根井正利, 1983)。理论上,只要足够的有效数据,此法就可适用于任何一对分类单位。在蛋白质和微卫星等共显性遗传标记的分析中,多用上公式来计算遗传距离,得到的结果基本能准确地解释遗传差异。但对于显性标记如 RAPD 标记,是否适用还需作进一步探讨。本研究结果显示若以条带出现频率来代替等位基因频率,用 Nei 的距离公式得到的结果与其他计算结果略有差异。

Rogers 的遗传距离是另一个较常用的距离公式(见公式 2, DR),可以看作为欧氏系数,也是以基因频率的函数表示群体的遗传变异,具有很强的几何意义。它可看作是数据缩减的一种手段,一个群体可根据某一基因座位上 n 个等位基因的频率,用 n 维空间中的一个点表示(Weir, 1994)。而 Hillis(1984)指出 Rogers 距离太易受分类群体内异质性的影响。Bonnin 等(1996)用 RAPD 标记同时计算了群体间 Nei 和 Rogers 的遗传距离,得到了高度相关的结果。本研究显示两种计算结果也是显著相关,但一些差异可由聚类分析树形图中直观地看出。由于缺乏选育过程的详细资料,因此很难判断哪种遗传距离所反映出的品系间遗传关系更可靠。但从公式的提出原理和计算方法来看,作者认为对于 RAPD 这种显性标记, Rogers 遗传距离更适合。

遗传距离 DH 是由 Hedrick(1971)提出的, Kuhnlein 等(1989)将它用于指纹 DNA 的分析中,此后在 RAPD 分析中也使用较多(Yu et al, 1993; Zhang et al, 1995),均能有效地反映出群体间亲缘关系。但本研究结果显示, DH 与其他计算结果相差较大,也不能很好的反映各品系间关系的实际情况。这可能是由于计算方法中存在着一些不合理造成的,即等于零的一些频率数值所产生的影响,使计算过程中产生较大偏差,因此,作者认为在以 RAPD 多态性条带频率来计算遗传距离时,此公式不合适。除非用于比较的多态性条带数量很大,而很少以频率为零出现于群体中。

为将限制位点数据转化为距离数据, Nei 和 Li(1979)提出了一对分类群 X 和 Y 以一个公共祖先分化后出现的核苷酸置换的数目估计量:直到现在还保留未变的祖先限制位点的比例值为:

$$S = 2N_{xy} / (N_x + N_y), \quad (\text{公式 5})$$

N_{xy} 为两群体共有的一致位点的数目; N_x 和 N_y 分别是 X 和 Y 中限制位点的总数。在指纹 DNA 和 RAPD 分析中,将每一条带作为一个限制位点以此公式计算出的共同带率作为距离数据,比较个体间或群体混合 DNA 样品间差异。但此公式不适于通过多个个体两两对比的平均值来计算群体间遗传距离。因此,作者设想是否可以通过对比各个不同的条带在群体中出现频率来表示群体间差异,并由此设计出公式 4 (DS)。结果表明, DS 与其他几种距离数据有显著相关,尤其与 DR 相关程度最高,而且也能有效反映出群体间实际关系。但它的适用性需要在理论与实践上作进一步验证。

参 考 文 献

- 吕雪梅, 杨关福, 张细权, 等. 1996. 以 RAPD 标记分析广东四个鸡品种的亲缘关系. *Animal Biotechnology Bulletin*, 5(Suppl): 63 ~ 65
- 孟安明, 齐顺章, 宫桂芬. 1993. 用四个探针产生的家禽 DNA 指纹图. *生物化学与生物物理进展*, 20(2): 139 ~ 142
- 根井正利. 1983. 分子群体遗传学与进化论. 王家玉译. 北京: 科学出版社, 169 ~ 172
- Weir B S. 1994. 遗传学数据分析—群体遗传学离散型数据分析方法. 徐云碧等译. 北京: 中国农业出版社.
- Bonnin I, Huguet T, Gherardi M, et al. 1996. High level of polymorphism and special structure in a selfing plant species, *Medicago truncatula* (leguminosae), shown using RAPD markers. *Am J Botany*, 83(7): 843 ~ 855
- Cargill S L, Anderson G B, Medrano J F. 1995. Development of a species-specific marker using RAPD analysis to distinguish between sheep and goats. *Animal Biotech*, 6(2): 93 ~ 100
- Hedrick P W. 1971. A new approach to measuring genetic similarity. *Evolution*, 25: 276 ~ 280
- Hillis D M. 1984. Measure and modification of Nei's genetic distance. *Syst Zool*, 33: 238 ~ 240
- Jeffreys A J, Morton D B. 1987. DNA fingerprints of dogs and cats. *Animal Genetics*, 18: 1 ~ 15
- Kuhnlein U, Dawe Y, Zadworny D, et al. 1989. DNA fingerprinting: a tool for determining genetic distances between strains of poultry. *Theor Appl Genet*, 77: 669 ~ 672
- Link W, Dixkens C, Singh M, et al. 1995. Genetic diversity in European and Mediterranean faba bean germ plasm revealed by RAPD markers. *Theor Appl Genet*, 90: 27 ~ 32
- Lu R, Rank G H. 1996. Use of RAPD analyses to estimate population genetic parameters in the alfalfa leaf-cutting bee, *megachile rutundata*. *Genome*, 39: 655 ~ 663
- Nei M. 1972. Genetic distance between populations. *American Nature*, 106: 283 ~ 292
- Nei M, Li W H. 1979. Mathematical modelling for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of sciences*, 76: 5268 ~ 5273
- Plotsky Y, Kaiser M G, Lamont S J. 1995. Genetic characterization of highly inbred chicken lines by two DNA methods: DNA fingerprinting and polymerase chain reaction using arbitrary primers. *Animal Genetics*, 26: 163 ~ 170
- Rogers J S. 1972. Measures of genetic similarity and benetic distances. *Studies in Genetics VII*. University of Texas Publications, 7213: 145 ~ 153
- Smith C A B. 1977. A note on genetic distance. *Ann Human Genet*, 21: 254 ~ 276
- Smith J F, Pham T V. 1996. Genetic diversity of the narrow endenmic *Allium aasæae* (Alliaceae). *American Journal of Botany*, 83(6): 717 ~ 726
- Yu K, Pauls K P. 1993. Rapid estimation of genetic relatedness among heterogeneous populations of alfalfa by random amplification of bulked genomic DNA samples. *Theor Appl Genet*, 86: 788 ~ 794
- Zhang X, McDaniel G R, Giambrone L L. 1995. Random amplified polymorphic DNA comparisons among broiler lines selected for incidence of tibial dyschondroplasia. *Poultry Science*, 74: 1253 ~ 1258

COMPARISONS OF CALCULATION METHODS FOR GENETIC DISTANCE IN RAPD ANALYSIS

Lu Xuemei Yang Guanfu Zang Xiquan

(Dept. of Anim. Sci., South China Agric. Univ., Guangzhou, 510642)

Abstract

In order to study population genetic variations with RAPD analysis more available, the genetic distances among 7 layer lines in Guangzhou Leghorn Company Ltd were calculated with Nei's equation, Rogers equation, Hedrick's equation and the equation developed by the authors based on RAPD analysis for 7 lines. It was concluded that Rogers genetic distance was more suitable to RAPD analysis data and that the authors genetic distance was correlated with the other genetic distances significantly, but had to be tested further.

Key words RAPD analysis; genetic distance

(上接第 89 页)

ANALYSIS OF RELATIONSHIPS BETWEEN SIX PIG BREEDS USING RANDOM AMPLIFIED POLYMORPHIC DNA

Liu Dewu Yang Guanfu Li Jiaqi

(Dept. of Anim. Sci., South China Agric. Univ., Guangzhou, 510642)

Abstract

A total of 140 short primers, of arbitrary nucleotide sequence, were used singly in polymerase chain reactions to amplify DNA fingerprints in pools of DNA from six pig breeds, including Duroc, Seghers, Landrace, Large White, Lantang and Dahuabai. Of these primers, 23 amplified no DNA fingerprint, 66 produced polymorphic DNA fingerprints and 51 monomorphic DNA fingerprints. Average number of bands per primer was 4.2, ranging from 1 to 11 bands. Similarity coefficients and cluster analysis, calculated according to the DNA fingerprints, indicated that there were close relationships between the Lantang and Dahuabai, and between the Seghers and Landrace. Duroc expressed a farther relationship to native breeds than to other imported pig breeds.

Key words RAPD (Random amplified polymorphic DNA); pig breeds; similarity coefficients; relationship