

文章编号: 10000-7423(2005)-06-0396-05

【论著】

## 钉螺 AFLP 分子标记电泳图谱信息 数量化数据的分析

周艺彪, 姜庆五, 赵根明, 韦建国

**【摘要】** 目的 探讨扩增片段长度多态性(AFLP)分子标记电泳图谱信息数量化数据的分析方法。方法 从现场采集的钉螺中筛选出 40 只阴性钉螺, 随机分为两组, 用于基因组 DNA 模板的制备。再用 Glyko BandScan 软件将钉螺扩增片段长度多态性电泳图谱信息数量化, 使用不同的读带标准读带, 得到相应的数据集, 然后对这些数据集进行遗传学统计分析 with 描述性总结。结果 不同的标准所得到的遗传变异结果均有所差别, 但随着读带标准值的增加, 反映钉螺种群遗传多样性指标(如: Shannon's 信息指数)也增加, 当其增加到一定水平时, 又开始下降, 而基因流和基因一致度则相反。不同读带标准所得的遗传变异结果均呈明显的正态分布( $P>0.05$ )。以总灰度或以总灰度百分比划分读带标准, 所得遗传变异结果的平均值均十分接近。两组钉螺平均基因一致度在总灰度百分比数据中为 0.956, 在总灰度数据中为 0.958; 两组间的平均遗传距离在总灰度百分比数据中为 0.045, 在总灰度数据中为 0.043。结论 将电泳图谱信息数量化, 再以不同的读带标准去处理与分析数据的模式, 是一种较为合理且准确的分析方法。

**【关键词】** 钉螺; 扩增片段长度多态性; 电泳图谱; 数据分析

中图分类号: R383.241

文献标识码: A

## Linking Quantitation of Electrophoresis Pattern and Data Analysis in AFLP for *Oncomelania hupensis*

ZHOU Yi-biao, JIANG Qing-wu, ZHAO Gen-ming, WEI Jian-guo

(Department of Epidemiology, School of Public Health, Fudan University, Shanghai 200032, China)

**【Abstract】** Objective To search into a method for analyzing the quantitative data in amplified fragment length polymorphism (AFLP) electrophoresis. Methods *Oncomelania* snails collected from the field were screened. Forty snails found uninfected with schistosomiasis were divided randomly into two groups and used to isolate genomic DNA. AFLP electrophoresis pattern was first transformed into quantitative data by Glyko BandScan software, and the bands were read according to different standards of band-reading to acquire the corresponding data. These data sets were analyzed by genetic statistics to get an inference set, and the analysis of this inference set was performed to reach a summary description. Results The results of genetic variation from different standards of band-reading were different. With the increase of the standard value of band-reading, the indices indicating the genetic polymorphism of *Oncomelania hupensis* population (e.g. Shannon's information index) also increased. When the standard value reached at certain level, the values of these indices began to decrease. Compared with the above indices, the change for gene flow turned out contrary to the genetic identity. The distributions of inference results from different standards of band-reading all showed significant normal distribution. The mean value of genetic variation based on total grey was very close to that on the proportion of total grey. The average genetic identity between the "subpopulations" was 0.956 according to proportion of total grey or 0.958 from the total grey with an average genetic distance between the "subpopulations" of 0.045 and 0.043 respectively. Conclusion It seems to be a reasonable and accurate method by quantifying the AFLP electrophoresis pattern followed by analyzing the data through the use of the different standards of band-reading.

**【Key words】** *Oncomelania hupensis*; AFLP; Electrophoresis pattern; Data analysis

Supported by the S&T programs during the National Tenth Five-year Plan (No. 2004BA718B04)

扩增片段长度多态性(amplified fragment length

polymorphism, AFLP)分子标记技术是由荷兰科学家 Zabeau 等<sup>[1]</sup>提出的, 是在随机扩增多态性 DNA(random amplified polymorphic DNA, RAPD)和限制性片

基金项目: "十五"国家科技攻关项目(No. 2004BA718B04)

作者单位: 复旦大学公共卫生学院流行病学教研室, 上海 200032

段长度多态性(Restriction Fragment Length Polymorphism, RFLP)技术的基础上发展起来的新一代 DNA 指纹分析技术, 其集 RAPD 和 RFLP 技术优点, 即在对某种生物遗传背景不甚了解的情况下, 利用该技术可对该物种的基因组各部分可能存在的 DNA 序列变异进行扫描分析, 克服了 RAPD 标记的不稳定性和 RFLP 技术受探针来源限制的制约, 并具有多态性强、分析方法快速、实验结果稳定可靠等优点, 可用于动植物的遗传多样性、基因定位、遗传图谱的构建、种及种下阶元的分类鉴定、系统发生和疾病诊断等多个研究领域<sup>[2-4]</sup>。但 AFLP 分子标记技术在一定程度上存在着读带困难的问题, 即对同一张电泳图、不同的人或不同的读带标准, 其分析的结果可能不一致。本研究先将 AFLP 电泳图谱信息数量化, 并按不同的读带标准对数量化的数据进行转换, 然后分别进行遗传学统计分析和描述性总结。

### 材料与方法

#### 1 钉螺

从湖南省岳阳市君山区采集的肋壳湖北钉螺约 120 只, 置于实验室内饲养约 1 周, 先用逸蚴法剔除阳性钉螺, 再随机抽取阴性钉螺约 50 只, 用 0.3% NaCl 清洗, 再用压碎法, 镜检钉螺软体组织内是否含有尾蚴或子胞蚴或母胞蚴, 阳性者再次剔除, 最后选取两次检查均为阴性的钉螺 40 只, 用于基因组 DNA 模板的制备。

#### 2 基因组 DNA 模板的制备

将上述挑选的 40 只阴性钉螺按文献<sup>[5]</sup>方法进行基因组 DNA 模板的制备。

#### 3 AFLP 分析

选用 5 对重复性好、多态性强的引物进行选择性地扩增<sup>[5]</sup>, Pst I 和 Mse I 荧光标记引物序列详见表 1。将上述制备好的钉螺基因组 DNA 模板进行酶切、连接、预扩增、选择性 PCR 扩增和 PCR 产物凝胶电泳

按参考文献<sup>[5]</sup>进行。

### 4 电泳图谱数量化与数据处理

4.1 电泳图谱信息数量化 用 Glyko BandScan 4.30 软件分别将 5 对引物所扩增的 5 张电泳图谱的条带信息数量化, 建立 2 个原始数据库: 每条带的总灰度和每条带总灰度占其所在泳道的所有条带总灰度之和的百分比(总灰度百分比)。

4.2 数据处理与分析的模式 参考 Bloom 等<sup>[6]</sup>提供的一种新的图谱信息数量化与数据分析的模式, 在结合电泳图谱中条带的清晰度并考虑 2 个数据库中数据的分布和统计结果的前提下, 分别按总灰度>5、>10、>15、>20、>25、>30、>40 和>45(单位:1000 pMols)和总灰度百分比>0.1%、>0.2%、>0.3%、>0.4%、>0.5%、>0.6%、>0.7%、>0.8%和>0.9%的读带标准将数据转化为“1”和“0”的形式, 例如: 总灰度>5 000 pMols 的条带记为“1”, 否则判为“0”, 其他类推。分别得到按总灰度百分比读带标准的 9 个只含“1”和“0”的数据集和按总灰度读带标准的 9 个数据集。另外也将两原始数据库中的一个数据库转化为“1”和“0”的形式(即缺失数据记为“0”,其他记为“1”,这 2 个原始数据按此形式进行转换, 所得到的数据库是相同的, 故只将其中的一个进行转换), 组建含“全部”带的数据库, 总共产生 19 个只含“1”和“0”的数据库, 然后分别对这 19 个数据库进行单独分析, 所得到的结果组成 2 个推论或结果数据库, 其中一个来自于以总灰度作为原始数据的数据库, 另一个来自以总灰度百分比作为原始数据的数据库。最后对推论数据库进行描述性统计分析。

4.3 数据统计分析 参考 RAPD 标记数据分析方法<sup>[7,9]</sup>, 将 AFLP 分子标记作为等位基因进行遗传多态性分析。同时为了比较不同读带标准对钉螺遗传变异研究结果的影响及验证读带标准划分的合理性, 人为随机地将 40 个钉螺标本分为 2 个组来分析钉螺群体的遗传多样性。用 Popgen32 统计软件计算多态位点频率 (P=多态性扩增片段数/扩增片段总数 100%)、Nei's 基因多样性指

表 1 AFLP 引物及其序列  
Table 1 AFLP primer and sequence

Pst I 引物 Pst I primer	序列 Sequence	Mse I 引物 Mse I primer	序列 Sequence
P1	5'-GACTGCGTACATGCA GAA-3'	M1	5'-GATGAGTCCTGAGTAA CTG-3'
P2	5'-GACTGCGTACATGCA GAA-3'	M2	5'-GATGAGTCCTGAGTAA CTT-3'
P3	5'-GACTGCGTACATGCA GAC-3'	M3	5'-GATGAGTCCTGAGTAA CAA-3'
P4	5'-GACTGCGTACATGCA GTC-3'	M4	5'-GATGAGTCCTGAGTAA CTG-3'
P5	5'-GACTGCGTACATGCA GTG-3'	M5	5'-GATGAGTCCTGAGTAA CTT-3'

数<sup>[10]</sup>、Shannon's 信息指数<sup>[11]</sup>和 Nei's 无偏估计的遗传距离<sup>[12]</sup>以及基因流与基因分化系数<sup>[13]</sup>等。其他数据的处理与分析用 Microsoft Excel 和 SPSS11.5 软件完成。

## 结 果

### 1 钉螺 AFLP 扩增片段

采用 5 对引物分别对 40 个钉螺标本的基因组 DNA 进行了选择性扩增,共产生 5 张电泳图。P5/M5 引物组合对 40 个钉螺基因组 DNA 进行扩增,片段大小在 35~500 bp 之间。每对引物扩增的片段数在 42~74 之间(图 1)。

机分为 2 组,也无论是总灰度百分比数据,还是总灰度数据,用不同的标准去读带,所得到的遗传变

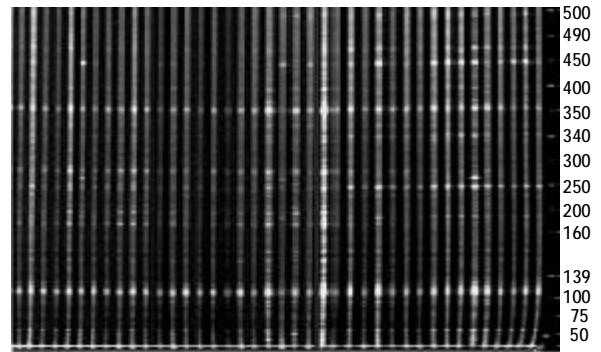


图 1 P5/M5 引物对钉螺基因组 DNA 进行扩增的 PCR 产物电泳图  
Fig.1 Electrophoretic chart of the amplified products of P5/M5 primers for Oncomelania snails

### 2 不同读带标准的遗传变异

无论是将 40 个钉螺作为一组样本,还是将其随

表 2 钉螺不同读带标准的遗传变异结果

Table 2 Genetic variation of different band-reading standards for Oncomelania hupensis

读带标准 Standard of band-reading	位点数 No. loci	多态位点数 No. polymorphic loci	多态位点率(%) Percentage of polymorphic loci(%)	Nei's 基因多样 性指数(H) Nei's gene diversity(H)	Shannon's 信息指数(I) Shannon's information index (I)
全部 All	282	193	68.44	0.228	0.343
总灰度百分比(%) Proportion of total grey (%)					
>0.1	282	243	86.17	0.292	0.438
>0.2	280	249	88.93	0.301	0.452
>0.3	279	254	91.04	0.298	0.449
>0.4	279	259	92.83	0.293	0.443
>0.5	274	254	92.70	0.287	0.436
>0.6	272	253	93.01	0.280	0.426
>0.7	263	246	93.54	0.279	0.426
>0.8	262	245	93.51	0.268	0.410
>0.9	257	240	93.39	0.262	0.401
平均 'Mean'	273(9)	243(19)	89.01(9.89)	0.279(0.022)	0.422(0.032)
正态性检验 ** Normal Test**	0.768	1.024	0.930	0.644	0.773
	0.579	0.245	0.352	0.801	0.589
总灰度(x1000)(pMols) Total grey (x1000)(pMols)					
>5	282	221	78.37	0.270	0.404
>10	282	235	83.33	0.285	0.428
>15	282	240	85.11	0.289	0.433
>20	280	242	86.43	0.291	0.437
>25	279	246	88.17	0.294	0.441
>30	277	249	89.89	0.297	0.446
>35	276	252	91.30	0.298	0.449
>40	273	252	92.31	0.298	0.448
>45	270	250	92.59	0.293	0.442
平均 'Mean'	278(4)	238(18)	85.61(11.10)	0.284(0.021)	0.427(0.032)
正态性检验 ** Normal Test**	0.659	0.769	0.573	0.990	0.984
	0.778	0.595	0.898	0.281	0.288

注: \* 括号中的数值为标准差; \*\* 正态性检验上行为 Z 值,下行为 P 值。

Note: \* Numbers in parenthesis are the standard deviation; \*\* The up line for Normal test is the Z value and the below line is the P value.

异结果均有所差异。但有一个共同趋势：随着读带标准值的增加，所得的多态位点数、多态位点率、Nei's 基因多样性指数、Shannon's 信息指数、组内基因多样性指数、总基因多样性指数、基因分化系数和遗传距离均增加，当其增加到一定水平时，又开始下降，而基因流和基因一致度刚好相反，随着读带标准值的增加，其值降低，当降至一定水平时，又开始升高。用不同读带标准所得的遗传变异结果均呈明显的正态分布 ( $P > 0.05$ )。以总灰度和总灰度百分比划分读带标准，所得到遗传变异结果的平均值都十分接近，总灰度百分比的 Nei's 基因多样性

指数平均为 0.279，而总灰度的 Nei's 基因多样性指数平均为 0.284；总灰度百分比的 Shannon's 信息指数平均为 0.422，而总灰度的 Shannon's 信息指数平均为 0.427(表 2)。两组随机划分的钉螺平均基因一致度在总灰度百分比数据中为 0.956，在总灰度数据中为 0.958；两组的平均遗传距离在总灰度百分比数据中为 0.045，在总灰度数据中为 0.043(表 3)。

### 讨 论

在以影像为基础的分子生物学实验中，如 Northern 印迹试验、蛋白质印迹试验和凝胶电泳试验等，

表 3 两组钉螺不同读带标准的遗传变异结果  
Table 3 Genetic variation of different band-reading standards between two subpopulations of snails

读带标准 Standard of band-reading	第 1 组 First subpopulation		第 2 组 Second subpopulation		组内基因 多样性指数 Gene diversity within populations (Hs)	总基因 多样性 指数 Total gene diversity (Ht)	基因流 Gene flow (Nm)	基因分 化系数 Coefficient of gene differe- ntiation (GSt)	Nei's 无偏估计 Nei's unbiased measures	
	Nei's 基因 多样性指数 Nei's gene diversity(H)	Shannon's 信息指数 Shannon's information index(I)	Nei's 基因 多样性指数 Nei's gene diversity(H)	Shannon's 信息指数 Shannon's information index(I)					基因 一致度 Genetic identity	遗传 距离 Genetic distance
全部 All	0.207	0.309	0.202	0.301	0.204	0.216	7.209	0.065	0.971	0.030
总灰度百分比 Proportion of total grey (%)										
>0.1	0.277	0.413	0.247	0.366	0.262	0.279	7.456	0.063	0.961	0.039
>0.2	0.279	0.417	0.264	0.393	0.272	0.290	7.483	0.063	0.959	0.042
>0.3	0.272	0.409	0.262	0.391	0.267	0.287	6.700	0.069	0.946	0.056
>0.4	0.264	0.398	0.260	0.390	0.262	0.282	6.534	0.071	0.954	0.047
>0.5	0.255	0.386	0.260	0.390	0.257	0.278	6.212	0.075	0.944	0.057
>0.6	0.246	0.375	0.259	0.392	0.253	0.272	6.666	0.070	0.949	0.052
>0.7	0.244	0.373	0.263	0.398	0.254	0.273	6.502	0.071	0.956	0.045
>0.8	0.233	0.356	0.256	0.389	0.244	0.263	6.302	0.074	0.958	0.044
>0.9	0.226	0.347	0.253	0.386	0.240	0.258	6.481	0.072	0.959	0.042
平均 'Mean'	0.250	0.378	0.253	0.380	0.252	0.270	6.755	0.069	0.956	0.045
	(0.024)	(0.034)	(0.019)	(0.029)	(0.019)	(0.021)	(0.463)	(0.004)	(0.008)	(0.008)
正态性检验 ** Normal test**	0.381	0.437	0.976	1.226	0.730	0.762	0.781	0.544	0.482	0.437
	0.999	0.991	0.297	0.099	0.660	0.607	0.576	0.928	0.974	0.991
总灰度 Total grey (×1000 pMols)										
>5	0.248	0.367	0.239	0.356	0.244	0.259	7.744	0.061	0.966	0.034
>10	0.259	0.385	0.249	0.370	0.254	0.272	6.974	0.067	0.960	0.041
>15	0.264	0.394	0.255	0.378	0.259	0.277	7.177	0.065	0.960	0.041
>20	0.266	0.396	0.259	0.386	0.262	0.280	7.139	0.066	0.959	0.042
>25	0.265	0.396	0.266	0.396	0.265	0.284	7.194	0.065	0.950	0.052
>30	0.262	0.394	0.277	0.413	0.270	0.289	6.912	0.068	0.956	0.045
>35	0.263	0.396	0.280	0.418	0.271	0.291	6.943	0.067	0.946	0.055
>40	0.259	0.391	0.283	0.422	0.271	0.291	6.679	0.070	0.954	0.048
>45	0.255	0.386	0.284	0.425	0.269	0.288	6.922	0.067	0.956	0.045
平均 'Mean'	0.255	0.381	0.259	0.387	0.257	0.275	7.089	0.066	0.958	0.043
	(0.018)	(0.027)	(0.025)	(0.038)	(0.021)	(0.023)	(0.284)	(0.002)	(0.007)	(0.008)
正态性检验 ** Normal test**	0.963	1.117	0.542	0.495	0.777	0.800	0.749	0.702	0.572	0.571
	0.312	0.165	0.946	0.967	0.581	0.543	0.629	0.709	0.899	0.900

\* 括号中的数值为标准差; \*\* 正态性检验上行为 Z 值, 下行为 P 值。

\* Numbers in parenthesis are the standard deviation; \*\* The up line for Normal test is the Z value and the below line is the P value.



目前许多研究主要关心的是实验定性结果,如条带或斑点是否出现,即“有”和“无”,或与别的点比较,确定其相对强度。这些试验一般产生的条带和斑点数目相对较少,结果分析时一般也不需要准确的定量数据或对影像信息进行准确的定量,一些简单的定性结果就可以满足实验的分析需要。但随着人类基因组知识的增加和近年来分子生物学的迅速发展,使得人类有能力同时检测数以千计的基因表达水平(如微阵列分析<sup>[14]</sup>)或数以百计的表达蛋白(二维凝胶电泳试验<sup>[15]</sup>),这些需要检测和理解分子在数量上的细微变化以及提高图像信息定量的灵敏性。因此对这些实验结果的简单定性方法,就无法满足实验的分析需要。近年来随着图形分析软件(如 Glyko BandScan 软件)发展与应用,使得分子生物学实验中的图像信息数量化的准确性有了较大的提高。本次研究先用 Glyko BandScan 图形分析软件,将 AFLP 电泳图谱信息数量化,然后用不同的读带标准去读带,结果发现不同的读带标准所得到的遗传变异结果均有所不同。当读带标准偏低或偏高,反映钉螺种群遗传多样性指标(如: Nei's 基因多样性指数、Shannon's 信息指数、基因分化系数和遗传距离)的结果也较少,而当读带标准处于中间水平时,这些遗传多样性指标的数值较高,而基因流和基因一致度刚好相反,随着读带标准值的增加,其值降低,当降至一定水平时,又开始升高。用不同读带标准所得的遗传变异结果都呈明显的正态分布。因此可以用不同读带标准所得结果的平均值来代表该钉螺种群的实际遗传变异情况,也许比用肉眼和使用一个读带标准所得的结果更为准确。但这种方法在划分读带标准时,需要反复结合电泳图片的实际情况和统计结果,往返几次才能确定合理的读带标准划分,这也许会给数据分析增加一定的难度。将 40 只湖北阴性钉螺随机划分为 2 组,所得的平均基因一致度在总灰度百分比数据中为 0.956,在总灰度数据中为 0.958;平均遗传距离在总灰度百分比数据中为 0.045,在总灰度数据中为 0.043,进一步验证了用 Glyko BandScan 图形分析软件将 AFLP 电泳图谱信息数量化的准确性以及不同读带标准划分的合理性。用以总灰度划分读带标准和以总灰度百分比划分读带标

准,所得到遗传变异结果的平均值均十分接近,因此对电泳图谱信息数量化的数据进行遗传学统计分析时,可结合电泳图片的实际情况选择其中一种来划分读带标准。综上所述,先将电泳图谱信息数量化,再以不同的读带标准去处理与分析数据的模式,是一种较为合理且准确的 AFLP 电泳图谱信息处理与分析的方法。

#### 参 考 文 献

- [1] Vos P, Hogers R, Bleeker M, et al. AFLP: a new technique for DNA fingerprinting[J]. Nucl Acid Res, 1995,3:4407-4414.
- [2] Otsen M, den Bieman M, Kuiper MT, et al. Use of AFLP markers for gene mapping and QTL detection in the rat [J]. Genetics, 1996,37:289-294.
- [3] Roos MH, Hoekstra R, Plas ME, et al. Polymorphic DNA markers in the genome of parasitic nematodes[J]. J Helminthol, 1998,72:291-294.
- [4] Schreiner T, Prochnow-Calzia H, Maccari B, et al. Chimerism analysis after allogeneic bone marrow transplantation with non-radioactive RFLP and PCR-AFLP using the same DNA [J]. J Immunol Methods, 1996,196:93-96.
- [5] 周艺彪,姜庆五,赵根明. AFLP 标记技术在湖北钉螺遗传多样性中的应用研究[J]. 中国血吸虫病防治杂志, 2005,17:34-38.
- [6] Bloom GC, Gieser P, Lazaridis EN. Linking image quantitation and data analysis[A]. Looney SW. Biostatistical methods[C]. Totowa, New Jersey: Humana Press,2002. 15-27.
- [7] 常青,周开亚. 分子进化研究中系统发生树的重建[J]. 生物多样性,1998,6:55-62.
- [8] Luch M. The similarity index and DNA finger printing[J]. Mol Biol Evol, 1990,7:478-484.
- [9] Vekemans X, Beauwens T, Lemaire M, et al. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size[J]. Mol Ecol, 2002,11:139-151.
- [10] Nei M. Analysis of gene diversity in subdivided populations[J]. Proc Nat Acad Sci USA,1973,70:3321-3323.
- [11] Lewontin RC. The apportionment of human diversity[J]. Evol Biol, 1972,6:381-398.
- [12] Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals[J]. Genetics, 1978,89:583-590.
- [13] Nei M. Molecular evolutionary genetics[M]. New York: Columbia University Press,1987.187-192.
- [14] Gieser P, Bloom GC, Lazaridis EN. Introduction to microarray experimentation and analysis[A]. Looney SW. Biostatistical methods [C]. Totowa, New Jersey: Humana Press,2002.29-49.
- [15] Seillier-Moiseiwitsch F, Trost DC, Moiseiwitsch J. Statistical methods for proteomics[A]. Looney SW. Biostatistical methods[C]. Totowa, New Jersey: Humana Press, 2002. 51-85.

(收稿日期:2005-01-15 编辑:盛慧锋)