

基于 Arlequin 软件的 HLA 单倍型分析研究*

刘悦越^{1,2}, 贾双双², 袁娟丽³, 杜加亮¹, 陈红兵³, 国 泰¹(1. 中国食品药品检定研究院, 北京 100050; 2. 中国医学科学院/北京协和医学院医学
生物学研究所, 昆明 650118; 3. 南昌大学, 南昌 330031)

摘要:目的 应用 Arlequin 软件进行 HLA 单倍型的分析。方法 将 HLA-A 和 B 两基因座位的等位基因分型数据进行格式转换, 编辑成 .arp 格式输入文件, 导入 Arlequin 软件进行分析计算。结果 用 Arlequin 软件进行 HLA 单倍型分析, 直接得出各单倍型及等位基因的频率和统计表, 分析得出 19 种 HLA-A 等位基因、39 种 HLA-B 等位基因和 128 种 A-B 单倍型。结论 Arlequin 软件可用于 HLA 单倍型的计算, 该方法具有简便、快速、广谱、可操作性强的特点。

关键词: 人类白细胞抗原; 单倍型; 等位基因; 基因多态性**DOI:**10.3969/j.issn.1672-9455.2016.13.004 **文献标志码:**A **文章编号:**1672-9455(2016)13-1752-03

Analysis of HLA haplotype by Arlequin software*

LIU Yueyue^{1,2}, JIA Shuangshuang², YUAN Juanli³, DU Jialiang¹, CHEN Hongbing³, GUO Tai¹(1. National Institutes for Food and Drug Control, Beijing 100050, China; 2. Institute of Medical Biology,
Chinese Academy of Medicine Science & Peking Union Medical College, Kunming 650118, Yunnan, China;
3. Nanchang University, Nanchang, Jiangxi 330031, China)

Abstract: **Objective** To investigate haplotype frequencies of HLA using Arlequin software. **Methods** The genetic type data of alleles in human leukocyte antigen (HLA) A and B loci were transformed to the file with .arp. Then the HLA haplotypes were calculated by Arlequin software. **Results** The frequencies of alleles and haplotypes were shown in table directly. A total of 19 HLA-A alleles, 39 HLA-B alleles and 128 HLA A-B haplotypes were observed. **Conclusion** In conclusion, the HLA haplotypes frequencies could be calculated by Arlequin software, which is a simple, convenient, fast, universal and operable way to analyze the haplotype.

Key words: human leukocyte antigen; haplotype; alleles; polymorphism

人类白细胞抗原(HLA)位于人第 6 染色体短臂 6p21.31, 是目前所知人体多态性最为丰富的基因系统。HLA 主要参与抗原提呈, 进而激活 T 淋巴细胞, 在特异性免疫反应中起重要的作用。HLA 的多态性使不同的个体对抗原的反应性和易感性不同, 因此 HLA 多态性与病原体易感、自身清除和疾病的发展之间的关系成为近几年的研究热点^[1-4]。然而, 群体中 HLA 不同的基因座位的等位基因还存在连锁不平衡的现象, 从而使得单倍型分析在群体遗传学和疾病基因关联性研究中日益受到重视^[5]。

目前常用于单倍型分析的软件有 Arlequin、Haploview 和 Phase 等, 但这些软件由于没有成熟的教程常常成为医学工作者的顺利使用的一个壁垒, 尤其是数据输入等环节有特要求, 使得研究者们无从下手。笔者通过实践应用摸索, 利用 Arlequin 软件进行 HLA 单倍型分析, 其具有简单、快速、精准的特点, 尤其是更为复杂的多个基因座位的单倍型连锁分析。本文通过实际操作案例与相关同行分享利用 Arlequin 软件在 HLA 单倍型分析中的应用。本文所有数据均基于 Arlequin3.5.1.3 版。

1 材料与方法

1.1 数据来源 以本实验室 96 份血样的 HLA-A、B 基因座分型数据为例, 对软件的操作过程进行介绍。

1.2 方法

1.2.1 Arlequin 软件的安装 首先进行 Arlequin 软件的安

装, 可以从 (<http://anthro.unige.ch/arlequin/>) 免费下载。其次进行两个支持软件安装: TextPad, 用于编辑 Arlequin 软件认可的数据形式。Rcmdr, 是 R 统计软件包的控制台版本, 用于统计数据的输出, 可以从 (<http://www.r-project.org/>) 下载。

1.2.2 基础数据输入及处理 为了便于后续的分析, 医学工作者往往将 HLA 各基因座位点的分型结果录入 Excel 表格进行整理记录。格式如下表 1, 每个样本上下两个结果分别表示两个等位基因的型别。进行软件分析必需首先对数据进行格式转换。第一步, 将上述 Excel 原始数据转换为文本文件数据, 可利用 Excel 另存为“文本文件”。第二步, 打开安装好的 TextPad 软件, 编辑 Arlequin 软件要求格式的输入文件。包括两个部分, 文件说明部分和数据部分, 如下图 1 所示。

1.2.3 文件说明部分操作要点 Title 表示本次分析的名称。NbSamples 指所分析数据的样本数或者人群的个数, 必须为整。GenotypicData 表示数据为单倍型数据, 还是基因型数据, 字符“0”表示为单倍型数据, 而“1”表示基因型数据。GameticPhase 表明配子片段的基因型是否已知, 字符“0”表示未知, 而“1”表示已知。DataType 表示所要分析的数据类型, 通常为 DNA、RFLP、MICROSAT、STANDARD 和 FREQUENCY, 同一个文件只能有一种数据类型, 本文中的实例为标准的 HLA 分型数据。LocusSeparator 为基因座分隔符, 用于分割不同基因座的型别数据的符号, 可以是除了“#”外的任何符号, 也可以没有, 默认值为“WHITESPACE”, 即为空格。Miss-

* 基金项目: 国家科技重大专项课题资助项目(2012ZX10004702, 2014ZX09304-316)。

作者简介: 刘悦越, 女, 助理研究员, 主要从事生物制品质量研究。△ 通讯作者, E-mail: guotai@nifdc.org.cn。

ingData 用于表示缺失数据的表示方法,任何字符联合单引号或双引号均可使用,默认值为“'?’”。RecessiveDat 表示某基因型数据是否为隐性等位基因,“0”表明“否”,而“1”表示“是”;RecessiveAllele 代表隐性等位基因的表示方法,可以用任何一串字符联合双引号表示,默认值为“null”。

表 1 HLA 各基因座位点的分型结果录入 Excel 格式

编号	频数	基因座	
		A	B
1	1	01:01	15:01
		11:01	49:01
2	1	01:01	13:01
		11:01	52:01
3	1	01:01	13:02
		30:01	57:01
4	1	02:01	07:02
		03:01	48:01
5	1	02:01	40:06
		02:07	46:01

```
[Profile]
Title="96 Algerians tested for HLA (A-B)"
NbSamples=1

GenotypicData=1
GameticPhase=1
DataIype=STANDARD
locusSeparator=WHITESPACE
MissingData=?
RecessiveData=0
RecessiveAllele="null"

[Data]

[[Samples]]
SampleName="96 Algerians tested for HLA (A-B)"
SampleSize= 96
SampleData={
1 1 01:01 15:01
11:01 49:01
2 1 01:01 13:01
11:01 52:01
3 1 01:01 13:02
30:01 57:01
4 1 02:01 07:02
03:01 48:01
5 1 02:01 40:06
02:07 46:01
6 1 02:01 35:01
26:01 52:01
7 1 02:01 38:01
26:01 56:01
8 1 02:01 15:18
29:02 45:01
. . 29:02 45:01
. . . . .}
```

图 1 TextPad 数据编辑格式

1.2.4 数据部分操作要点 数据从第一步保存的文本文件中可以直接复制粘贴,其中第 1 列表明型别的编号,第 2 列为频率,后两列为分别 A、B 两个基因座的型别数据。需要注意的是 SampleSize 必须与第 2 列频率的总和相等,数据末尾必须以“}”结束。编辑完成后,需要保存为.arp 格式文件。

1.2.5 软件分析 从 Arlequin 软件中打开上面处理后预存的格式为.arp 文件。操作步骤如下:第一步,在“Settings”项选择“Haplotype inference”调出对话框,界面如下图 2 所示。第二步,在对话框中勾选“Estimate haplotype frequencies by mere counting”,如果需要统计每个基因座的等位基因型别频数,则同时勾选“Estimate allele frequencies at all loci”。其他选项保留默认值。第三步,点击对话框右上方的“Start”键开始统计分析。结果以 HTML 文档的格式保存在与数据同根的文件夹中,名称与数据文件名称一致,并自动在浏览器中弹出。

2 结 果

2.1 Arlequin 软件的输出结果 根据以上步骤,分析结果显示本次共得出 128 个单倍型。通常医学工作者关心的部分是各单倍型的频数和相对频率,结果中会自动列出相对频率的结果。如果在分析是同时勾选了等位基因的频率估算,则在结果分别显示各基因座的基因型种类和频率。为了更为方便的分

析,可以将结果数据复制到 txt 文件中,再进行 Excel 表格格式转换。

2.2 等位基因型别和单倍型结果 分析结果显示,96 份样本中 HLA-A 共有 19 种型别,HLA-B 共有 39 种型别,单倍型 128 种。下表 2 为筛选出显示频率>5.0%的 HLA-A、B 等位基因和频率>1.5%的单倍型。

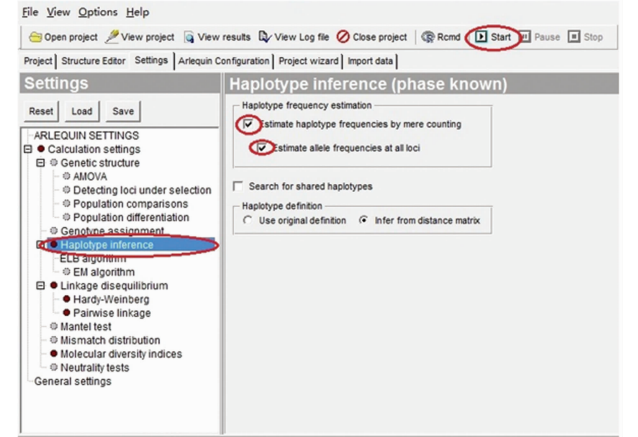


图 2 Arlequin 软件分析界面

表 2 等位基因和单倍型结果

型别	频数(n)	频率(%)	型别	频数(n)	频率(%)
A * 02	52	27.08	A * 11-B * 13	10	5.21
A * 11	39	20.31	A * 02-B * 15	9	4.69
A * 24	23	11.98	A * 02-B * 40	8	4.17
A * 30	18	9.38	A * 24-B * 40	8	4.17
A * 33	17	8.85	A * 02-B * 13	7	3.65
B * 13	30	15.63	A * 02-B * 07	6	3.13
B * 40	28	14.58	A * 11-B * 40	6	3.13
B * 15	17	8.85	A * 30-B * 13	5	2.60
B * 52	15	7.81	A * 02-B * 46	5	2.60
B * 51	13	6.77	A * 11-B * 15	5	2.60
B * 48	10	5.21	A * 01-B * 13	4	2.08
B * 35	10	5.21	A * 02-B * 35	4	2.08
			A * 31-B * 51	4	2.08
			A * 11-B * 51	4	2.08
			A * 32-B * 52	4	2.08
			A * 33-B * 58	4	2.08
			A * 02-B * 38	3	1.56
			A * 02-B * 49	3	1.56
			A * 11-B * 55	3	1.56
			A * 24-B * 48	3	1.56
			A * 33-B * 44	3	1.56
			A * 33-B * 52	3	1.56

3 讨 论

HLA 的统计分析方法有很多,包括 SAS、Haploview、Phase、Arlequin 等,但是大部分软件在安装、数据输入、操作分析等环节需要计算机编程等专业背景,为医学生物学工作者带来不便。然而 Arlequin 软件具有免费使用和简便易操作的特点,对于没有隐性基因的数据,软件采用最大似然原理计算单倍型的频率,与其他软件或程序基础的统计方法相同^[6-7],但使用界面友好,优势明显。

本文研究在该群体中共观察到 19 种 HLA-A 型别,39 种

HLA-B 型别,呈现丰富的基因多态性。HLA-A * 02、A * 11、A * 24、A * 30、A * 33、B * 13、B * 40、B * 15、B * 52、B * 51、B * 48 和 B * 35 的频率高于 5%,与我国汉族人群的分布相似^[8-9]。A 和 B 位点在理论上可以组成 741 种单倍型,但实际观察到 128 种,呈明显的连锁不平衡。频率最高的三条单倍型是 A * 11-B * 13、A * 02-B * 15 和 A * 02-B * 40,它们在上海 4 082 例汉族人群的研究中也高于 1%^[10]。

事实证明,利用 Arlequin 软件能够快速简便地对 HLA 单倍型进行分析。本文以具体数据作为实例,从软件安装直至数据分析详细阐述了软件的应用,为使用者提供了一定的借鉴参考。

参考文献

[1] Grimaldi V, Sommesse L, Picascia A, et al. Association between human leukocyte antigen class I and II alleles and hepatitis C virus infection in high-risk hemodialysis patients awaiting kidney transplantation[J]. Hum Immunol, 2013, 74(12):1629-1632.

[2] Miao F, Sun H, Pan N, et al. Association of human leukocyte antigen class I polymorphism with spontaneous clearance of hepatitis B surface antigen in Qidong Han population[J]. Clin Dev Immunol, 2013, 2013(12):145725.

[3] Pithukpakorn M, Roothumnong E, Angkasekwinai N, et al. HLA-DRB1 and HLA-DQB1 are associated with adult-onset immunodeficiency with acquired anti-interferon-gamma autoantibodies [J]. PLoS One, 2015, 10 (5):

e0128481.

[4] Amanzadeh A, Amirzargar AA, Mohseni N, et al. Association of HLA-DRB1, DQA1 and DQB1 alleles and haplotypes with common variable immunodeficiency in iranian patients[J]. Avicenna J Med Biotechnol, 2012, 4(2):103-112.

[5] Jafarzadeh A, Bagheri-Jamebozorgi M, Nemati M, et al. Human leukocyte antigens influence the antibody response to hepatitis B vaccine[J]. Iran J Allergy Asthma Immunol, 2015, 14(3):233-245.

[6] 李婧, 潘玉春, 李亦学, 等. 人类基因组单核苷酸多态性和单体型分析及应用[J]. 遗传学报, 2005, 32(8):879-889.

[7] 郭晋. 医学遗传资料统计分析方法的研究与 SAS 实现 [D]. 北京:中国人民解放军军事医学科学院, 2010.

[8] 冯云飞, 宿军, 杨春晴, 等. 潍坊地区血小板捐献汉族人群 HLA-A、B 和 HPA1-17 系统基因多态性研究[J]. 中国输血杂志, 2015, 28(1):59-62.

[9] 王芳, 廖群, 黄霞, 等. 重庆汉族人群 HLA-A、B、DRB1 高分辨等位基因多态性研究[J]. 重庆医学, 2014, 43(26):3455-3460.

[10] 杨剑豪, 刘嫵, 郑皆炜, 等. 4 082 名上海骨髓库汉族无关供者 HLA-A、B、DRB1 高分辨等位基因及单体型多态性研究[J]. 中国输血杂志, 2012, 25(12):1262-1267.

(收稿日期:2016-02-10 修回日期:2016-04-14)

(上接第 1751 页)

检测 S/CO 值为 23.97±1.869,该值高于 21.210。通过对定值质控品的检测,证实 ELISA 检测 S/CO 值为 15.314~21.210 时,可估计胶体金法检测结果。

综上所述,在急诊手术前或急诊侵入性检查前采用胶体金法进行 HBsAg 检测,存在一定程度的漏检风险,应引起实验室和临床的关注。

参考文献

[1] Seeger C, Mason WS. Hepatitis B virus biology[J]. Microbiol Mol Biol Rev, 2000, 64(1):51-68.

[2] 刘杨, 彭道荣. 乙肝表面抗原实验室检测方法的进展[J]. 国际检验医学杂志, 2015, 36(17):2546-2548.

[3] 马连学, 李艳菊, 魏薇. 化学发光微粒免疫法和酶联免疫吸附法检测乙肝表面抗原的结果对比分析[J]. 中国实用医药, 2014, 3(9):89-90.

[4] Sonneveld MJ, Rijckborst V, Boucher CA, et al. A comparison of two assays for quantification of Hepatitis B surface Antigen in patients with chronic hepatitis B[J]. J Clin Virol, 2011, 51(3):175-178.

[5] 周迎春, 陈辉, 关平, 等. 3 种方法测定低浓度乙型肝炎病毒表面抗原效果评价[J]. 检验医学与临床, 2007, 4(11):1070-1071.

[6] 王海刚, 丁磊, 潘航, 等. 3 种 HBsAg 检测方法的比较[J]. 国际检验医学杂志, 2014, 35(12):1614-1615.

[7] 陈远林, 秦立新, 张仁生, 等. 住院患者血液传染性相关指标检测的意义[J]. 中华医院感染学杂志, 2007, 17(4):388-390.

[8] 龚光明, 周红芳, 李桃, 等. 医护人员血源性职业暴露监测分析与预防策略[J]. 中华医院感染学杂志, 2012, 22(15):3303-3305.

[9] 马春华, 赵跃, 李艳萍, 等. 某院工作人员血源性职业暴露基线调查及分析[J]. 重庆医学, 2013, 42(10):1139-1141.

[10] 姜永梅, 许永进, 商玲, 等. 急诊护士血源性职业暴露状况及风险防范[J]. 中华医院感染学杂志, 2013, 23(11):2714-2716.

[11] 厉挺, 强明珠, 陶爱女, 等. 消化内镜医源性感染的危险因素及控制对策[J]. 中华医院感染学杂志, 2013, 23(19):4761-4762.

[12] 马连学, 李艳菊, 魏巍. 化学发光微粒免疫法和酶联免疫吸附法检测乙肝表面抗原的结果对比分析[J]. 中国实用医药, 2014, 9(9):89-90.

[13] 何宗忠, 裴宇容, 魏东, 等. 电化学发光测定低浓度乙肝病毒表面抗原的临床意义[J]. 中国实验诊断学, 2013, 17(11):2056-2058.

[14] 朱宇清. 乙型肝炎表面抗原胶体金免疫层析法血清快速测定的性能评估[D]. 上海:复旦大学, 2012.

(收稿日期:2016-03-04 修回日期:2016-05-07)