

ARLEQUIN 使用说明

引言

Arlequin是一款优秀的人类遗传学数据分析软件，其名字来源于法语“Arlecchino”，是一个十七世纪意大利著名喜剧人物的名字。这个喜剧人物具有多个面目，可以根据需要，多个角色之间轻而易举的相互转变。Arlequin软件包如此取名，大概是为了说明此款软件能够满足遗传分析方面的需求。Arlequin软件包提供了许多方法和统计学检验来从遗传学和人口统计学数据（如大量的分子序列数据和传统的等位基因频率等）中挖掘信息。Arlequin软件有着友好的 Java图形操作界面，便于使用者操作。

Arlequin软件包由 Stefan Schneider、David Roessi 和 Laurent Excoffier 人完成。Arlequin软件包下载和升级的网址为：[http://l2p.unige.ch/~excoffier/Arlequin/](#)。下载后的 Arlequin 软件包基本由 Arlequin20_zip.exe 和 jre117-win32.exe 组成，在运行 Arlequin 程序之前，需要先安装 jre117-win32.exe。Arlequin20_zip.exe 是个自解压的程序，点击此程序将文件释放到所选择的目录，就可以运行了。在上述网址还提供了一个升级包 arlpatch2001.zip 修正了原软件里边的一些 bug，并提高了某些计算程序的精确性；下载后解压，直接运行即可。

Arlequin 功能概述：

Molecular diversity (分子多态性)

Mismatch distribution (错配分布)

Haplotype frequency estimation (单倍型频率估计)

Linkage disequilibrium (连锁不平衡)：检测不同位点上等位基因的非随机关联

Hardy-Weinberg equilibrium (哈温-伯格平衡)

Tajima's neutrality test (Tajima 中性检验)

Fu's neutrality test (Fu 中性检验)

Ewens-watterson neutrality test (Ewens-watterson 中性检验)

以上三个中性检验都是基于无限位点模型，适用于 DNA sequence 和 RFLP 单倍型。

Chakraborty's amalgamation test (Chakraborty 融合检验，检测人群的均一性或同质性，和中性选择等)

Minimum Spanning Network (MSN，最小扩张树或称之为最小支撑树，基于分子差异)

AMOVA (分子差异度分析，用以评测人群的遗传结构)

Pairwise genetic distances (遗传距离的估计)

Exact test of population differentiation (检测随机交配群体单倍型的非随机分布)

Assignment test of genotypes (通过估计等位基因频率将单个基因型分配到特定的人群中)

Arlequin 软件包功能强大，以上列出了本软件包一些基本的功能，下文将对这些功能进行详细的阐释和实例讲解。

Arlequin 输入数据的格式

Arlequin 软件包大致能接受以下五种数据格式：DNA sequences、RFLP data、Microsatellite data、Standard data、Allele frequency data。这些数据可以使单倍型 (haplotypic) 数据格式，也可以是基因型 (genotypic) 数据格式。对于 RFLP data 类型数据，“1”表示存在限制性位点，“0”表示不存在限制性位点，“-”表示限制性位点缺失。对于 DNA sequences 类型数据，“-”代表一个缺失的核苷酸，“?”代表一个未知核苷酸，R 表示 A/G (purine)，Y 表示 C/T (pyrimidine)，M 表示 A/C，W 表示 A/T，S 表示 C/G，K 表示 G/T，B 表示 C/G/T，D 表示 A/G/T，H 表示 A/C/T，V 表示 A/C/G，N 表示 A/C/G/T。

Arlequin 软件包输入文件的扩展名应该为 *.arp 配置文件的扩展名为 *.ars 在 Arlequin 软件包输

入文件中，#后可以输入任何字符，直至此行结尾。

1、Alequin输入文件的格式- profile

在一个Alequin输入文件的最开始是 Profile部分

[profile]

Title=" "

当前分析数据的名称或标题（双引号内可以为任何字符串）如：Title="ancient mtDNA data of QiDan"
NbSamples=

（指所分析数据中人群的个数，可以是 1-1000之间的任何整数） 如：NbSamples= 6。

DataType=

此部分用以说明所分析数据的格式。可以输入的字符有：
DNA, RFLP, MICROSAT, STANDARD, FREQUENCY。例如：DataType=DNA

GenotypicData=

此参数用来说明所分析的数据是单倍型数据还是基因型数据。可以输入的字符为：0 (haplotypic data) 和 1 (genotypic data) 例如：GenotypicData= 0

LocusSeparator=

此参数用以说明在不同的位点用何字符来分离等位基因。可使用的字符有：
WHITESPACE, TAB, NONE, 或除#、?、-, 以外的任何字符。例如：LocusSeparator= TAB

缺省值为：WHITESPACE

GameticPhase=

此参数用以说明配子片段的基因型是否已知。此处可使用的字符有：0 (gametic phase not known) 和 1 (known gametic phase) 例如：GameticPhase= 1

缺省值为：1

RecessiveData=

此参数用以说明所研究基因型数据是否为隐性等位基因。此处可使用的字符有：0 (co-dominant data 共显性) 和 1 (recessive data 隐性数据)。例如：RecessiveData= 1

缺省值为：0

MissData=

用来确定用什么字符来代表缺失的位点数据，这个字符要输入在“ ”或“ ’ ”之间。例如：MissData= “ § ”。

缺省值为：MissData= “ ? ”。

Frequency=

当单倍型或表现型的频率用绝对或相对的数值来表示时，用到此项。可填的参数有：ABS （绝对数值）、REL （相对数值）。绝对数值可以通过对样本数目的相对频率计算而得到。例如：Frequency=ABS。

缺省值：ABS

CompDistMatrix=

此参数用来说明距离矩阵数据是否来自原始的数据，还是直接就是数字的形式。可采用的字符有：
0（利用亚矩阵数据）和 1（通过单倍型信息计算距离矩阵）。例如：CompDistMatrix= 1

缺省值：0

FrequencyThreshold=

此参数用来界定输出文件中，单倍型频率数据的范围。可采用的数值有：从 0.01到 0.0000001的有理数例如：FrequencyThreshold=0.01

缺省值：0.00001

EpsilonValue=

此参数代表利用基因型数据用来估计单倍型频率和连锁不平衡的运算法则的收敛标准。此处可以用的数值为： 10^{-7} 到 10^{-12} 例如：EpsilonValue= 10^{10}

缺省值： 10^{-7}

2、Alequin输入文件的格式-Data section

Data部分的数据格式要求如下。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/118143105105006123>