

PAUP 4.0 快速开始指南 Quick Start Tutorial

翻译者: XUWEI (bioxuwei@163.net)

生物软件网 (<http://www.bio-soft.net>) 提供

2003 年 12 月 1 日

下面的指南手册对 PAUP4.0 的基本用法提供了一个非常简洁的概述。这个指南将通过分析例子数据文件中的一个例子来一步一步教会你。这个例子可以在你提供给你的盘里,也可以在 <http://paup.csit.fsu.edu/data/primate-mtDNA-interleaved.nex> 上下载。这个指南为那些以前没有用过 PAUP 经历的人设计,如果你已经很熟悉 PAUP,那么你可以跳过这个指南,我们假定用户已对系统发育术语和特别的计算机操作系统很熟悉。如对 DOS 和 Unix 用户而言,我们已在你的路径上设定了 PAUP 二进制,只要键入命令“paup”便会运行 PAUP 程序。只要你用 PAUP4.0 越来越多,你就会发现对下面操作的描述又很多替代方案。显而易见的原因是:我们不可能在这个指南中描述所有的可能操作。但是,只要你的时间允许,我们还是鼓励你去开发其他的菜单和命令行选项。

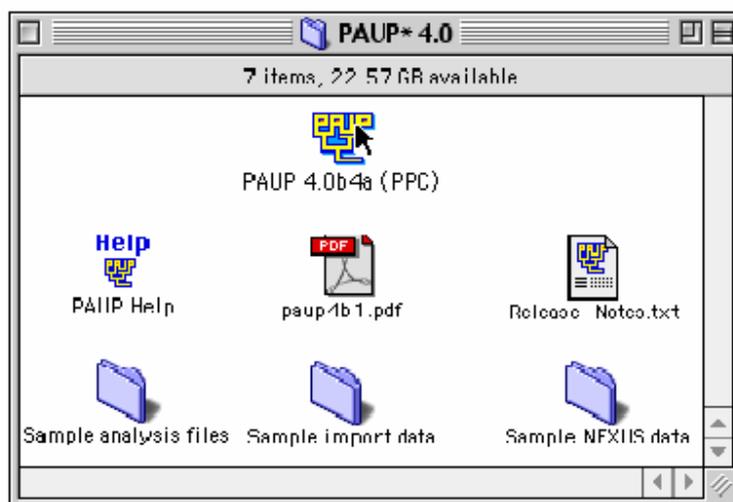
目前有 PAUP4.0 的好几个版本,这些版本分别适合 Macintosh, Windows 和 Portable 的这三个界面, Macintosh 界面允许你执行菜单命令和命令行,而在 Windows 和 Portable 界面(DOS 和 Unix)界面几乎完全由命令行驱动,很多菜单功能在 WINDOWS 界面有,而这些大部分包括归档和编辑操作,这个指南将用菜单选项和命令行语法来展示在不同环境下 PAUP 可能的运行情况。

为了排版和印刷上的方便,这个指南全篇采用了以下几点。第一,菜单,菜单选项和对话框或屏幕任何位置上的选项均用黑体 San serif 字体,如,文本 **File > Open** 意思是选择主菜单“**File**”再选择“**File**”菜单下的“**Open**”选项。第二,用户想在命令提示行或在对话框中键入的文本用无格式的固定了宽度的字体表示,如,指令“**输入: weights 2:1stpos**”表示只要出现“**输入:**”,都应当确切的键入它出现的。最后,界面特别的说明已分开列出并表述,但是所有其他的文本适合所有 PAUP 界面。

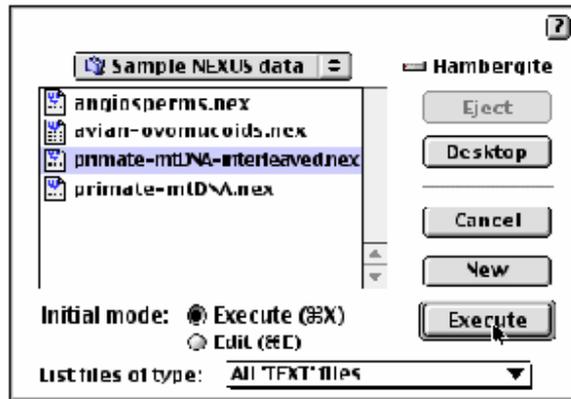
开始 PAUP

1 检查数据文件

Mac 和 Windows. 双击 PAUP 应用图标。



· 当第一次启动时, PAUP 将自动登陆到打开对话框



- . 选择“Sample NEXUS data”文件夹中的“primate-mtDNA-interleaved.nex”的文件。
- . 通过单击 **Edit** 来改变原来的 Execute 起始模式为 Edit (编辑) 模式, 这样来打开 PAUP* 的编辑器。

Portable . 改变路径到 Sample-NEXUS-files, 它包含在 PAUP 文档中选择你最喜欢的文档编辑器 (如 vi 或 emacs) 来打开文件名为 primate-mtDNA-interleaved.nex 的例子文件。

. 在这个编辑器里, 通过例子文件来滚动. 要注意这个文件被单词“begin”和“end”划分为多个文本区, “begin”后面的单词定义区型. 在这个例子里, 有以下几个区型: 分类, 元素, 假设, 和 paup. 但是, 有很多其他的 NEXUS 区型. 实际上, 一个 NEXUS 格式的优点是应用程序对那些他们不能识别的区能够简单的跳过去. 可以看 Maddison 等 (1997) 的文章来了解关于 NEXUS 格式更详细的讨论.

2 执行数据文件

关闭例子文件并按以下操作:

Mac 和 Windows

- . 选择 File > Execute "primate-mtDNA-interleaved.nex"

Portable

- . 在系统提示行输入 paup. PAUP* 将显示该程序的一般信息并在 paup 提示行显示光标.
- . 输入: execute primate-mtDNA-interleaved.nex;

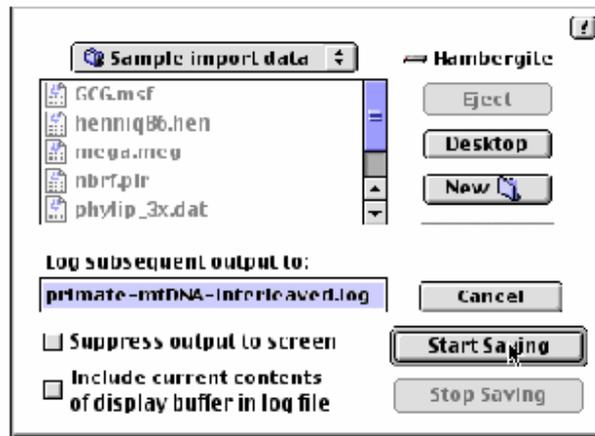
在执行这个例子文件后, PAUP 将显示关于这个数据的评论和一般信息. 如, 显示关于这个数据矩阵的维数和数据类型等等的部分, 接着是显示这个数据设置的来源. 到这里, 还没有进行数据分析; PAUP 只是简单的处理了这个数据, 并等待执行下一步的命令.

记录日志结果

通常, 你想要在硬盘上记录 PAUP 分析的结果, 这样使得你有你分析的结果的记录.

1 开始日志记录

Mac 和 Windows .选择 **File > Log Output to Disk...**



- Mac .在 **Log subsequent output to:** 下输入 **practice.log** ,单击 **Start Saving**
- Windows . 在 **Filename:** 下输入 **practice.log** 单击 **Start**
- Portable . 输入: **log start file=practice.log;**

2 停止记录

在 PAUP 决议的任何时候都可以开始和停止日志记录。如果要停止日志记录,按以下操作:

- Mac . 选择 **File > Log Output to Disk...** 并单击 **Stop Saving**
- Windows . 选择 **File > Log Output to Disk...**
- Portable . 输入: **log stop;**

总结数据

现在这个数据矩阵正被分析,你可以获得这个数据设置的基本总结数据。一开始,你可以显示关于这个例子数据设置元素的信息。

- Mac . 选择 **Data > Show Character Status...**
 - . 不要改变默认设置并单击 **OK**.
- Portable 和 Windows . 输入: **cstatus;**

PAUP* 将显示关于当前元素状态的总结 (如类型,权重,等等)。注意,如果打开记录,显示在屏幕上的总结信息也会被保存在日志文件中。你也可以选择去显示分类 (**tstatus**)、全部数据矩阵 (**showmatrix**) 和更多的总结,所有的这些命令都在 **DATA(数据)** 菜单下面。

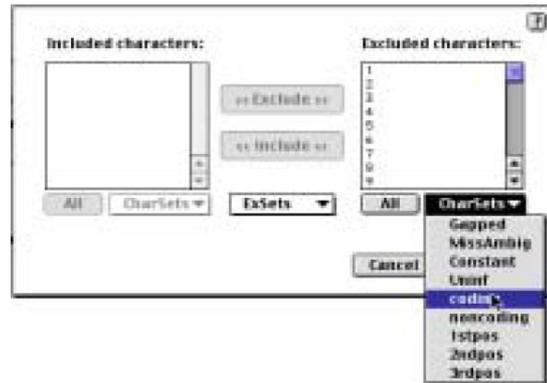
管理数据

PAUP提供了几个方法对包括在数据矩阵中的分类和元素的子集进行限制分析,如这个例子数据设置包括主要的线粒体DNA编码和非编码蛋白区域.假设我们只想分析这个数据中的编码区域.用**charsets**命令可以识别该例子文件中这些区域的元素。你可以通过一个单个的名称

来参考一组元素,这样使得元素设置程序变得简单.除了编码区域,你可以通过排除数据集中所有的元素来开始.

1 排除元素

Mac • 选择**Data > Include-Exclude Characters...**



- 在 **Included characters:**下面点击**All**并点击 **>> Exclude >>**
- 在 **Excluded characters:** 下面选择 **CharSets : coding**并点击**<< Include <<**
- 点击 **OK**

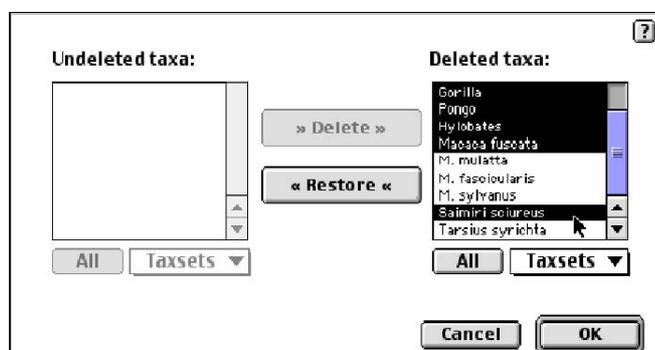
Portable 和 windows • 输入: `include coding/only;`

2 删除分类

你也可以限制你所有的分析为五个人科动物种和三个其他用于外围集团分类的灵长类种.这五个人科动物(*Homo sapiens*, *Pan*, *Gorilla*, *Pongo*, 和 *Hylobates*)可以用例子文件中的taxset命令来确定,用相同的方法,charset命令使得你可以通过单个名称来参考一组元素,taxset使得你通过单个名称来参考一组分类.

Mac • 选择**Data > Delete-Restore Taxa...**

- 在**Undeleted taxa:**下点击**All**并点击**Delete**
- 在**Deleted taxa:**下面选择**Taxsets**并然后选择**hominoids**
- 拉下应用命令键(**⌘**)选择 *Lemur catta*, *Macaca fuscata*和 *Saimiri sciureus*



- 点击 Restore 和 OK

Portable 和 *Windows* • 输入: `undelete hominoids lemur_catta macaca_fuscata saimiri_sciureus/only;`

要注意在命令行中分类名中的空格一定要用“_” (字符下划线) 或包含在单引号中, 也有PUAP不注意分类标签中的元素语法格. 最后, 要意识到当各自用exclude和delete命令 (或菜单相等的命令) 来排除或删除分类时, 你实际上没有修改数据文件. 也就是, 下一次执行这个例子数据时, 所有的元素和分类设置仍将在内.

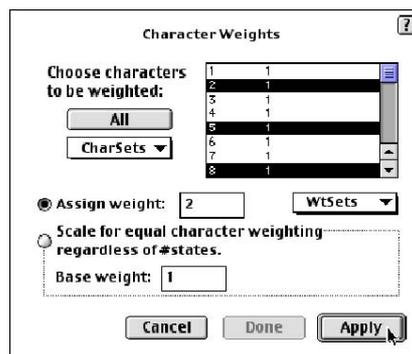
定义假设

在你开始分析之前, 这里有一个很好的机会了解你的数据矩阵中元素的有关信息, 这些信息可能提示这些元素可能有不同的加权值. 如, 通常发生在密码子第一位的取代频率要低于第三位. 这个简单的解释是密码子第一位的取代通常导致一个氨基酸的取代, 但是, 第三位的改变却不会改变氨基酸的转录. 你要把这些信息包含在下面的分析中, 可以通过采用在密码子第一位的高权重的取代来解决. 用charset命令已经确定例子文件中的密码子位点.

1 增加元素权重

Mac • 选择 **Data > Set Character Weights...**

- 在 **Charsets** 菜单下选择 **1stpos**
- 在 **Assign weight** 框中输入 **2** 并单击 **Apply** 然后单击 **Done**



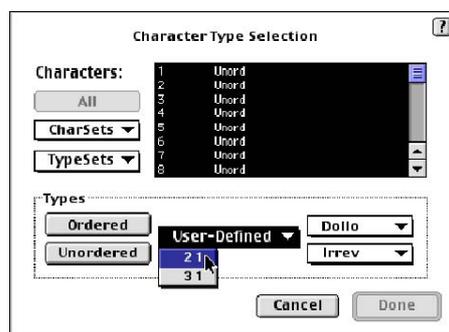
Portable 和 *Windows* • 输入: `weights 2:1stpos;`

2 设置元素类型

PAUP默认所有转换耗费都相同,在这部分,你将调用一个元素类型,它被分配了比转录较高的权重的颠换.更详细的说,我们将假定颠换(从嘌呤(A或G)到嘧啶的改变(C或T))的耗费是转录的两倍,一个在这个分析包含这个假设的方法是:建立转录/颠换“步骤矩阵”.这样的步骤矩阵已在例子文件中定义.为了应用转换耗费到当前考虑的所有元素中去,可以按以下操作:

Mac • 选择 **Data > Set Character Types...**

- 在 **Characters:** 下面点击 **All** 按钮
- 在菜单 **User-Defined** 下 **2 1** 并点击 **Done**

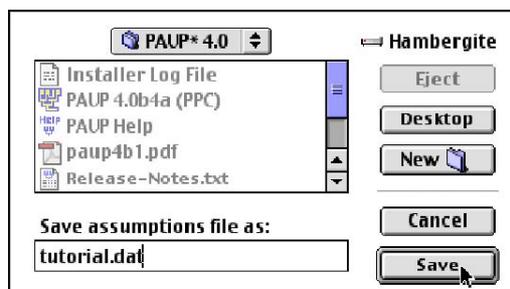


Portable 和 Windows • 输入: `ctype 2_1:all;`

3 保存当前假设

到这里,你已经排除元素,删除分类,加权元素和定义元素转化类型.如果因为某一个原因,你要废除你的分析和关闭PAUP,你将选择所有菜单选项或重复输入先前的命令来回到你所到的地方,一个解决这个潜在的耗时任务的方法是保存你的假设到一个文件中,这样可以在迟点时间重新打开.

- Mac
- 选择 **Data > Save Current Assumptions...**
 - 在文件名中输入 `tutorial.dat` 并单击 **Save**



Portable 和 *Windows* • 输入: `saveassum file=tutorial.dat;`

4 重新打开假设

重启PAUP并执行开始操作的文件
`primate-mtDNA-interleaved.nex`,按以下步骤重新打开先前的
假设设定

Mac 和 *Windows* • 选择 **File : Open...** 并选择 **tutorial.dat**

- 改变 **Initial mode(起始模式)**:由**Edit**到**Execute** 并点击 **Execute**

Portable • 输入: `execute tutorial.dat;`

你现在将返回到你所开始的地方.为了保证假设是有效的,从
Data菜单或在命令行用`cstatus`选择**Show Character Status....**你将
得到下面的输出

```
Character-status summary:
Current optimality criterion = maximum parsimony
285 characters are excluded
Of the remaining 693 included characters:
  All characters are of user-defined type "2 1"
  462 characters have weight 1
  231 characters have weight 2
  296 characters are constant
  155 variable characters are parsimony-uninformative
Number of (included) parsimony-informative characters = 242
```

查找树

1 定义最佳标准

PAUP* 4.0 具有用不同的最佳标准来分析数据的优点;这些标准有简约性法,可能性法和距离依靠法.在这个手册的好几章节和大量已发表的文献中专注于比较这些最佳标准的性能.在这里我们只是仅仅说每个标准都有各自的长处和限制,而不花费时间讨论目前标准各自相关的优点.一开始,我们默认采

用最大简约性法标准来查找最优树.在这个指南的后面,你可以用其他标准查找树.

Mac • 选择 **Analysis > Parsimony** (注: 简约性法是默认设置并可能已被选定.)



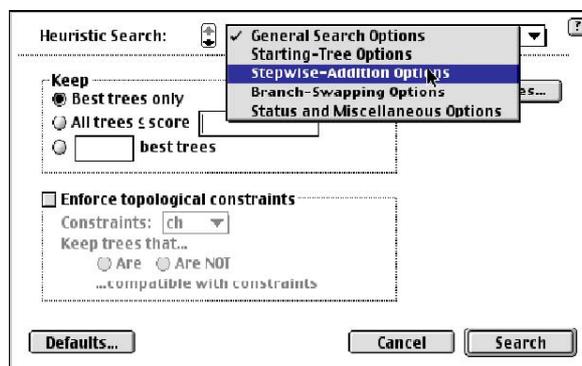
Portable 和 Windows • 输入: `set criterion=parsimony;`

2 定义查找策略

PAUP* 提供了两个基本类方法来查找最优树;它们是精确的和启发式的.精确的方法保证找到最优树,但是却可能需要大量的计算时间来分析大批量的数据设置.启发式方法不一定找到最优树,但是一般需要较少的计算时间.即使当前的数据相对少,你开始将采用启发式查找.

Mac

- 选择 **Analysis > Heuristic Search...**
- 在项目菜单下选择 **Stepwise-Addition Options**



- 把 **Addition sequence** 由 **Simple** 更改为 **Random** 并点击 **Search**

Portable 和 Windows • 输入: `hsearch addseq=random;`

一旦查找开始PAUP将显示查找过程中关于选项和假设的全面信息.如果你已记录日志结果,这些信息将保存到日志文件中,当查找结束时,PUAP将显示查找结果的全面信息.

Mac 和 Windows • 点击 **Close** 来关闭查找状态对话框.

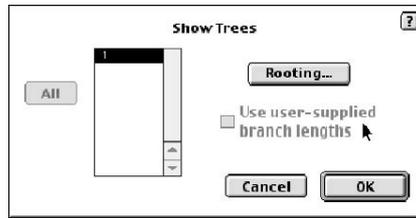
打印树

1 显示树

根据你的屏幕上的输出结果,将有一个单树正在内存中,按以下操作来显示树.

Mac

- 选择 **Trees > Show Trees...**



- 点击 **OK**, 这个单个最大简约性树将显示出来.

Portable 和 Windows

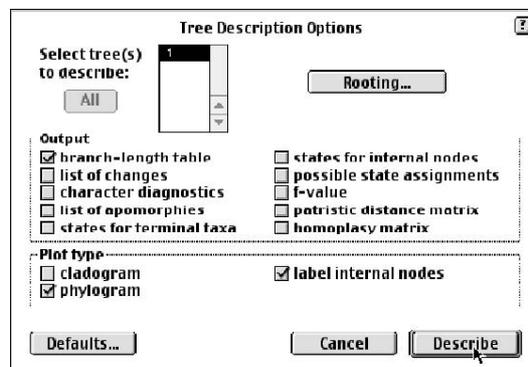
- 输入: `showtrees`;

2 描述树

`showtrees` 命令画出分类次序的一个简单分支图.举例说,你想知道树的分支长度的有关信息. 要得到树的详细图,按以下操作.

Mac

- 选择 **Trees > Describe Trees...**



- 在**Output**下面选择 **branch-length table**
- 在 **Plot type**下面不选定**cladogram** ,选择**phylogram**
- 点击 **Describe**

Portable 和 Windows •输入: `describetrees 1/plot=phylogram brlens=yes;`

3 打印低决议树

Mac 和 Windows • 一个快速得到显示在屏幕上的树的方法是打印显示缓冲的内容,但是,打印显示缓冲会打印出树和在树显示以前的其他所有屏幕显示内容.因此,我们建议你首先清除显示缓冲的内容然后重新显示树,这个操作可以用showtrees或describetrees命令来完成.

•选择 **File > Print Display Buffer...**

Portable

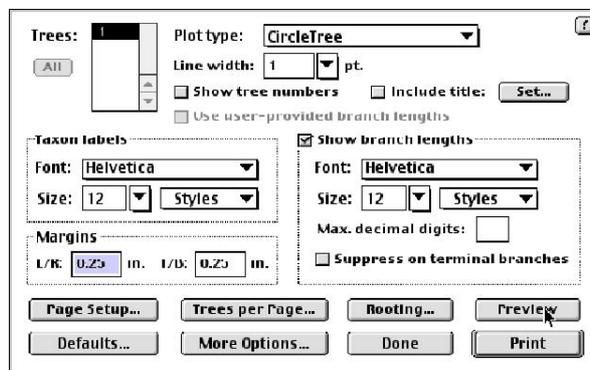
• 要用portable版本来打印,你可以在showtrees命令发布后打印控制台窗口,也可以选择部分日志文件来打印,在你的屏幕上显示的任何树都保存在日志文件中

4 打印高决议树

只有 Macintosh 界面才能够打印高决议树

Mac .选择 **Trees > Print Trees...**

Portable 和 Windows



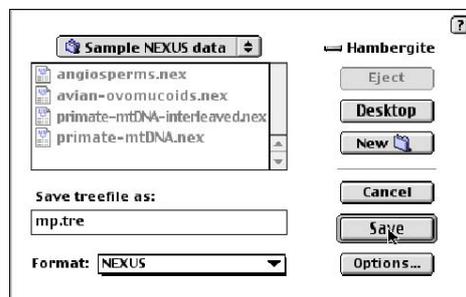
- 选择 **Plot type > CircleTree ...**
- 点击 **Show branch lengths** 检查框并点击 **Preview**
- 如果你要打印你选定的树,选择 **Done** 然后点击 **Print**

如果你使用Windows或Portable版本的PAUP,并要打印高决议的树,你必须使用一个第三方的树打印软件包,Rod Page编写的Tree View程序可以打印和保存NEXUS格式的高决议树.Tree View可以运行在Macintosh或PC机上,并可以从 <http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>. 免费下载.

保存结果

PAUP*能够用几种不同的格式保存树,这些格式有 PICT (只适合Mac), NEXUS, Freqpars, Phylip和 Hennig86. 要把树保存为NEXUS 格式,按以下操作:

Mac • 选择**Trees > Save Trees to File...**



- 在**Save Trees as:**对话框里选择 `mp.tre` 文件类型,并点击**Save**

Portable 和 Windows • 输入: `savetrees file=mp.tre;`

设置最优标准为距离依靠法

PAUP* 提供一个较大范围配对距离的方法,包括简单绝对差异到修正距离基础上的较为复杂的模式.配对距离能够用表格来总结,或者是用于除权配对法 (UPGMAM) 和邻位相连法构建树.另外,PAUP能用最

小进化和最小平方功能在距离依靠法中评价树.下面的部分将介绍这些方法.

1 设置最优标准

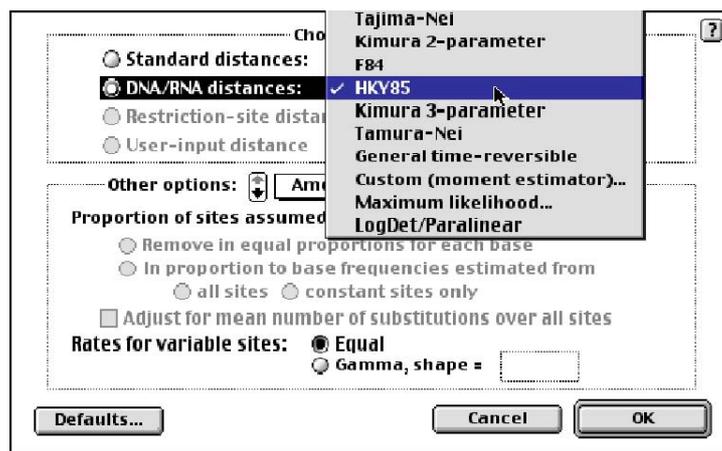
Mac • 选择 **Analysis > Distance**

Portable 和 Windows • 输入: `set criterion=distance;`

2 显示距离

首先你必须选择PAUP能够计算范围内的距离尺寸.在这个指南中,你将选择 Hasegawa, Kishino和 Yano (1985)距离.他们估测转换/颠换比和碱基频率.

Mac • 选择 **Analysis > Distance Settings...**



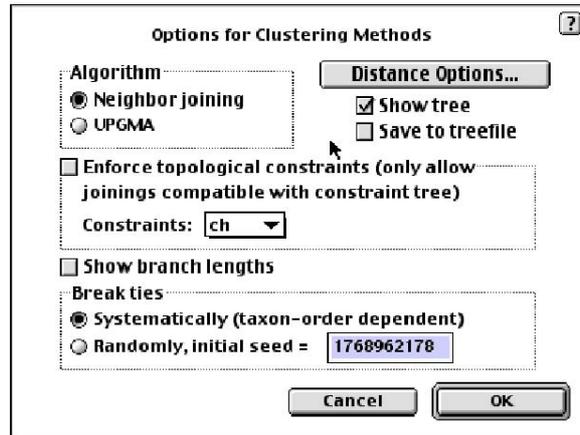
- 在距离设置对话框,把**DNA/RNA distances**由 **Uncorrected ("p")** 更改为 **HKY85** 并点击 **OK**
- 选择 **Data > Show Pairwise Distances**

Portable 和 Windows • 输入: `dset distance=hky85;`
• 输入: `showdist;`

3 构建邻位相连树

接下来,你将用 HKY85 距离构建邻位相连树

Mac • 选择 **Analysis > Neighbor Joining/UPGMA...**



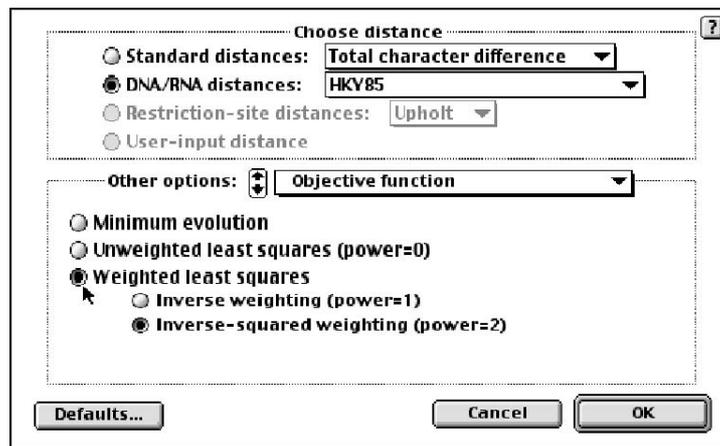
• 点击 **OK**

Portable 和 Windows

• 输入: nj;

4 建立最小平方树

Mac • 选择 **Analysis > Distance Settings...**



- 更改**Other options:** 为 **Objective function**
- 选择**Weighted least squares**
- **Inverse-squared weighting (power=2)** 是**Weighted least squares**下的默认设置.如果不是,现在选择它并单击**OK**.
- 选择**Analysis > Heuristic Search...** 来开始最小平方查找

- 在启发式查找对话框中点击**Search**

Portable 和 *Windows*

- 输入: `dset objective=lsfit power=2;`
- 输入: `hsearch;`

设置最优标准为可能性法

为结束这个指南,你将用最大可能性法标准来查找.在最大可能性法下面,一个对核酸取代较清楚的模式用来评价树.在可能性法分析中,选择一个合适的核酸取代模式是非常重要的步骤,但是这个超出了本指南的范围.为了节省时间,我们已经选择了合适模式,但是我们鼓励你去看 Swofford et al. (1996)关于在最大可能性法标准下选择模式的讨论.你将使用在这个指南中早些时候保存的简约性树来获得给定数据最佳模式参数设置.以后你将用相同的模式和参数设置评估来查找最大可能性树.

1 设置最优标准

Mac

- 选择 **Analysis > Likelihood**

Portable 和 *Windows*

- 输入: `set criterion=likelihood;`

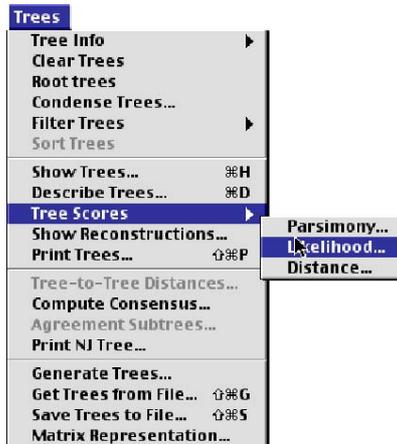
2 评价简约性树

我们已采用了 Hasegawa, Kishino 和 Yano (1985) 的用微克分布比率进化序列. 假定我们采用 PAUP 的简约性拓扑和数据来评价转换/颠换比率, 碱基频率和内部位点比率异质性.

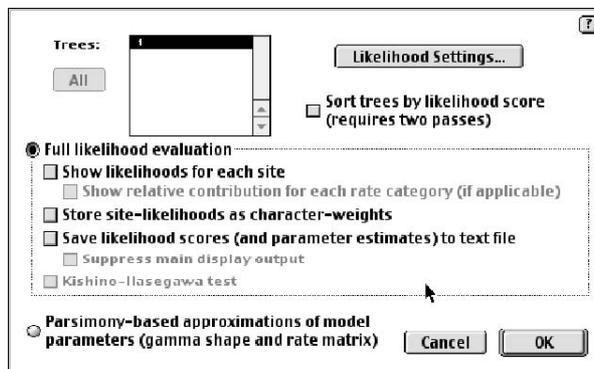
Mac

- 选择 **Trees > Get Trees from File...**

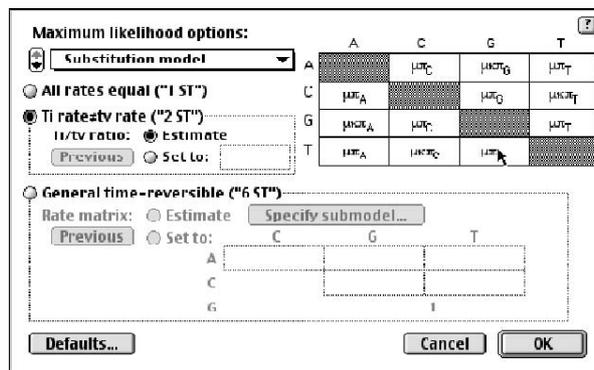
- 点击 **Yes** 来清除警告你没有保存树的对话框.
- 选择文件 **mp.tre** 并点击 **Get Trees**
- 选择 **Trees > Tree Scores > Likelihood...**

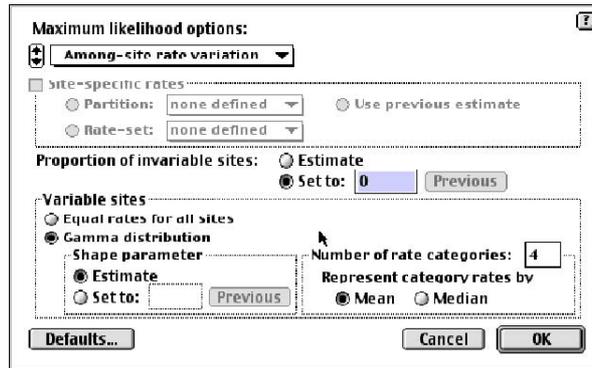


- 在树评价对话框中点击**Likelihoods settings...**

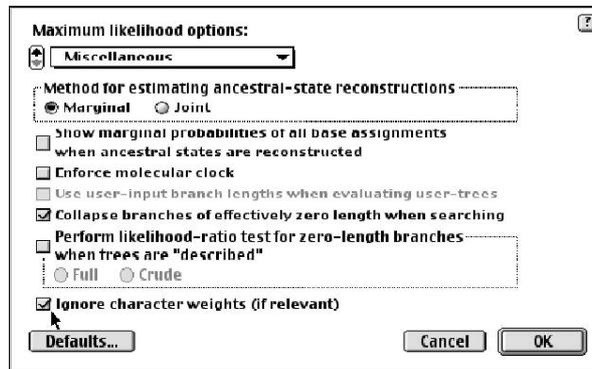


- 在 **Maximum likelihood options:** 下面选择 **Substitution model**
- 把 **Ti/tv ratio:** 由 **Set to: t** 更改为 **Estimate**
- 在 **Maximum likelihood options:** 下面选择 **Among-site rate variation**
- 在 **Variable sites** 下面选择 **Gamma distribution**
- 在 **Shape parameter** 下面, 把比率由 **Set to:** 改为 **Estimate**





- 点击 **OK**
- 返回 **Maximum likelihood option:** 菜单并选择**Miscellaneous**
- 选择 **Ignore character weights** (如果相关)并点击 **OK**



- 在树评价对话框中重新选择 **OK**

Portable 和 *Windows* • 输入: `gettrees file=mp.tre;`
 • 按Enter来消除没有保存树的警告

• 输入: `lscores 1 / wts=ignore nst=2
 ratio=estimate rates=gamma shape=estimate;`

根据你所有的计算机不同,可能要花费几秒到几分钟使得PAUP来处理当前内存中树的分支长度和取代模式参数的优化.当PAUP结束优化处理,将输出关于用简约性法查找的树的拓朴结构的负对数可能性并给出评估的模式参数值.

3 设置可能性模式参数

在启发式查找开始之前,你必须安装模式参数到那些先前评估的步骤.如果这些项目留待评估,PAUP将估测启发式查找过程中的每种拓扑重排参数.因为PAUP在启发式查找中可以产生成千上万的拓扑重排,使得评价项目设置需

要明显的增加时间来完成查找.通常,一个非常有效的在最大可能性方法下评价模式参数和树拓扑的方法是成功的评价树查找中产生的新的树的模式参数(Swofford et al. 1996).更详细的说,如果在可能性标准下找到的拓扑结构不同于评价参数上的拓扑结构,你就要在新的拓扑结构下重新评价参数并用新的设置参数重新查找.在这个指南中,你将完成一个在一个拓扑上的参数评价并随后应用这个参数来查找的反复.原则上,你一直继续直到得到相同的拓扑.

Mac • 选择 **Analysis > Likelihood Settings...**

- 在**Maximum likelihood options:** 下面选择 **Substitution model**
- 通过点击**Previous**按钮来选择先前步骤的评价值来设置 **Ti/tv ratio:**

• 在 **Maximum likelihood option:** 下面选择 **Among -site rate variation**

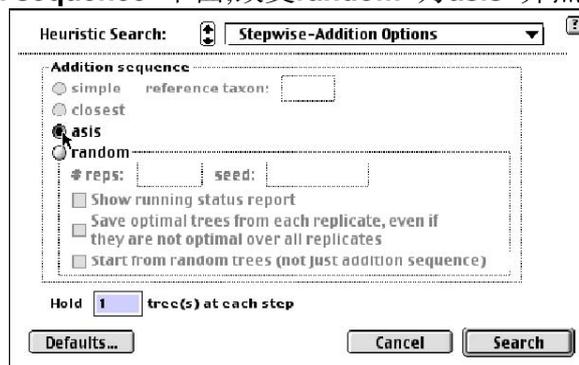
- 通过点击**Previous**按钮来选择先前步骤的评价值来设置 **Shape parameter**
- 点击 **OK**

Portable 和 Windows • 输入: `lset tratio=previous shape=previous;`

4 开始查找树

Mac • 选择**Analysis > Heuristic Search...**

- 在**Heuristic Search**下面,下拉菜单,选择**Step-wise addition options**
- 在 **Addition sequence** 下面,改变**random** 为**asis** 并点击 **Search**



Portable 和 Windows • 输入: `hsearch addseq=asis;`
同样,查找所需时间依赖你所使用的计算机.

在批文件中提交命令

分析者也可以用非交互式批方法来操作.当你知道你的分析需要大量时间完成时,这种方法特别有效.在下面的例子中,所有需要完成上述描述的例子分析的命令都包含在一个"paup"模块中,在开始的paup模块中加入了一套命令,它们用来消除在启发式查找完成后的提示对话框及几个其他的警告(仅仅在Mac 和 Windows中).要想以批模式运行模块,首先拷贝下面给定的文本到一个文件中,并把这个文件保存在和文件 `primate-mtDNA-interleaved.nex file` 相同的目录下.现在就像执行 `primate-mtDNA-interleaved.nex` 文件一样执行这个文件.

```
Begin paup;
  set autoclose=yes warntree=no warnreset=no;
  log start file=practice.log replace;
  execute primate-mtDNA-interleaved.nex;
  cstatus;
  include coding/only;
  undelete hominoids lemur_catta macaca_fuscata

saimiri_sciureus/only;
  weight 2:1stpos;
  ctype 2_1:all;
  Set criterion=parsimony;
  hsearch addseq=random;
  Showtree;
  describetrees 1/ plot=phylogram brlens=yes;
  savetrees file=mp.tre replace;
  set criterion=distance;
  dset distance=hky85;
  showdist;
  nj;
  dset objective=lsfit power=2;
  hsearch;
  gettrees file=mp.tre;
  lscore 1/wts=ignore nst=2 tratio=est rates=gamma

shape=estimate;
  set criterion=likelihood;
  lset tratio=previous shape=previous;
  hsearch addseq=asis;

                                end;
```

展望

这个指南包括PAUP* 4.0的简单概括的基本用法.PAUP中所有命令和选项的简要描述和列表都有命令参考文档.在这里,我们鼓励你去开发你自己的数据设置或其他PAUP包括的例子数据设置.你也可以在我们的网上在线找到关于怎样使用PAUP的信息.我们的网址是: <http://paup.csit.fsu.edu/>.