来自生物软件网，谢谢 jerry(xujz602@sohu.com)和 Huang(candycat719@sina.com)辛苦的翻译工作。

翻译说明：

1 原英文稿中附有许多示例（如输出窗口，序列等），译文一般用“参见英文稿”表示在阅读此段时请参见英文中的图示。对于一些较小的示例（如等式的推导）译文中保留。

2 原英文稿中也给出了许多算法和程序的原始文献和网址（译者认为这是 BioEdit 的一个优点，如果想深入的学习不能不读一读原始的文献），译文中用 REFERENCE 表示请参考英文原稿。

3 译文中对专业的词汇采用以下办法处理，即一般采用国内已有人使用的译法，如果未见到则译者给出一种译法并在旁边列出英文（译文中各节的标题都是这样处理的）。前者如最简单的例子，Aligment 一词有“比对”，“对比”，“对排”等多种翻译，郝柏林院士建议译做“联配”，（见《生物信息学手册》p175）；方舟子译做“排列对比”（见《新语丝》网页）。本译文采用“联配”的译法；后者如 mask 一词，（文中专门有一节解释其含义），此词的普通含义有“面具，遮饰”，译文中使用“屏蔽”并在旁边写 mask。总之，此类词汇使用多了，自然明了其内在的含义。

4 偶尔译者会对某处略做解释（旁边用“译者注”表示）表示译者的理解，请注意。

5 翻译时在词汇的翻译和算法的理解上参考了以下资料：

A 《生物信息学手册》 郝柏林等著 上海科学技术出版社 2000 年

B 〈生物信息学——基因和蛋白质分析的实用指南〉 Andreas D.Baxebanis 等原著，李衍达等译 清华大学出版社 2000 年

6 由于译者占有的资料不多，水平有限，在译文中肯定有漏译，译的不全面甚至理解完全错误的地方(尤其是算法上)，敬请指正。Email me at: xujz602@sohu.com

关于 **BioEdit**
介绍
BioEdit 版本 5.0.6

汤姆 • 霍尔
当前版本制作于2001.12.2

BioEdit是一个生物序列编辑器，可在Windows 95/98/NT/2000中运行，它的基本功能是提供蛋白质、核酸序列的编辑、排列、处理和分析。1.0α版本是最早的、未完成的并有瑕疵的版本。1.0αⅡ版本也一直未完成，并有很多问题，但是比较前一个还是增加了一点东西，修正了一些问题。在2.0版本中，在增加和配置附加分析应用程序上增加了一个界面，使其能通过BioEdit得到一个图形界面。而且，还增加了位置排列的信息基础动态描影。

版本3中增加了疏水／亲水面、互交的2-D浮雕数据绘图和一些更多的序列操作法。版本4为绘制和注解质粒载体增加了一个图形界面。在4.7.1版本中，修改了处理序列信息和存储方法，而且增加了一个二进制文件格式，允许快速保存和打开大的排列。序列容量增加到20,000。在版本5中，增加了自动注解序列或手动使用所有的标准Genbank功能部件定义。而且，在Isis Pharmaceuticals公司的请求下，增加了序列排序和分型、组控制、注解行，以及残基和非残基字符的鉴别。BioEdit并不打算成为一个强序列分析程序，但是打算成为一个序列分析的友好用户界面，并连接其他在局域网和万维网上的更多的序列分析程序。它现在使用于大的排列（>2000序列）。文件界面最初模仿于一个非常好的程序－－Don Gilbert编写的SeqApp and SeqPup。印地安那州大学免费提供SeqApp (用于个人计算机) and SeqPup (用于交换平台)，地址是：ftp://iubio.bio.indiana.edu/molbio/seqpup/

GeneDoc是一个特别的排列程序，能够自由的在Windows 9x 和NT上使用。也是一个非常专业的程序，有很好的蛋白质排列注解和分析、描影和结构定义功能部件，就象一个反映排列的内在的进化树，而这些在BioEdit中是没有的。GeneDoc的网址是http://www.psc.edu/biomed/genedoc。

GeneDoc 有比 BioEdit 更好的描影和分类选项，有助于手工排列序列，还有更好的图形处理、缠绕和伸展的排列视图选项、动态共有序列和更平滑和更快速的排列卷曲和刷新。

BioEdit 是用 Borland's C++ Builder 编写的 C++程序。我是北卡罗来纳州大学微生物系的研究生，不是专业的程序员。这是我学习 C++语言的入门，必然是个非专业的设计（这不是我博士工作的一部分）。这个程序非常小而且很有效率。BioEdit 为序列排列、输出和一些分析提供容易的工具。

BioEdit功能
BioEdit的主要目的是，为那些不愿意被迫详细了解一个程序的使用方法的生物学家，提供一个有用的工具。BioEdit是直观的、菜单式的并有大量的图示，提供用户一个外部分析程序的图形界面。主要功能是提供明显的、容易使用的菜单选项。

5.0.6版本提供以下功能：
• 用于序列处理和编辑的，简单的图形界面

• 使用编辑选项，包括残基的“select and drag”（选择和拖动）和“grab and drag”（抓取和拖动），变量选择选项，鼠标点击插入和删除缺口，全框选择，全屏编辑中剪切、复制和粘贴，编辑窗口的自动刷新
• 固定序列框，保护排列中的固定残基
• 使用各种功能部件（内含子、外显子、促进子、CDS和所有标准GenBank功能部件类型）自动的和手动的注解序列。使用一个模板序列，自动注解同一排列中的其他序列。
• 序列分组，分为各个颜色编码家族，为同步手动排列锁定组成员
• 用户定义的适当功能部件（能够设定考虑任何功能部件，就像用于类似性描影、序列同一性矩阵和保存图表视图的核酸或氨基酸序列中的相关碱基，）
• 用户定义的基序搜索使用标准的Prosite命名法和IUPAC功能部件，允许搜索核酸或氨基酸序列，还有精确的文本搜索包括或忽略缺口。
• 程序行可以定义为DNA、RNA、核酸、蛋白质、未定义或注解。注解可以用于保存普通的注释或东西，就象二级结构模糊定义，但是不能保存计算。
• 根本的多基因树图阅读器，支持节点翻转和打印
• 链接多基因树图到排列，并保存到BioEdit格式排列文件
• 在一个排列末端添加另一个排列
• 配置附件应用程序界面，进入一个有BioEdit产生的图形界面的外部分析程序。在外部应用程序中，自动提供信息和找回文件。外部应用程序进入分开的调度单位，允许同步应用BioEdit。外部程序的输出文件可以自动被其他程序打开。
• 在ABI自动序列模型377、373、3700中显示、打印和编辑ABI痕迹文件，在版本2和3中有SCF文件，就象用Licor序列输出文件
• RNA比较分析工具，包括共变，可能配对和互交信息分析
• 使用鼠标指示的动态数据视图的互交信息输出2－D矩阵图表
• 关于互交信息矩阵行和框的互交式的1－D图表
• 用BioEdit或GanBank格式保存序列注解信息
• 通过氨基酸翻译，排列蛋白质编码核酸序列
• 在排列中搜索保存的残基（寻找好的PCR目标或帮助定义基序）
• 在核酸或蛋白质序列中搜索用户定义的基序，或用通配符搜索精确的文本，并选择包括或忽略缺口
• 用支持最多20,000序列每个文档进行循环存储器分配。最多可以成功测定四百六十万个碱基（*E. coli*基因组）。核糖体数据库中的原核细胞16SRNA排列（29 Mb, 6205个序列）将会被单独处理。在配置为Pentium 233 Mhz、80 Mb RAM的计算机中，用BioEdit计划文件格式，最多只需要10秒种可以写入一个16S RNA排列。
• 内部的读写GenBank、Fasta、Phylip和NBRF/PIR文件，用Don Gilbert’s ReadSeq导入／输出一些其他格式的文件
• 使用BioEdit计划文件格式，快速读写大排列文件
• 使用自动更新的排列蛋白质全标题和GenBank区域信息，进行ClustalW多序列排列（Des Higgins et. al.编写的内部界面、外部程序），就象排列来自于核苷酸序列的蛋白质视图时的核苷酸编码序列。
• 将残基块状复制到剪贴板，允许将全不排列或部分排列粘贴到文字处理器
• 基本序列处理（在文档之间复制／粘贴序列，翻译和还原编码，RNA→DNA→RNA，反转／互补，大写字母／小写字母）
• 多文档界面（最多同时打开20个文档，但是在其他打开的窗口不能设置限制）
• 六框翻译核酸序列为Fasta格式ORF表。

•用矢量图进行半自动质粒／矢量绘图和注解，自动酶切位点和位置标记，自动多接头视图，和用户控制绘图工具
•将质粒文件保存为可编辑的矢量图形文件（如位图），复制到其他图形程序，并可以打印。
• 氨基酸和核苷酸成分摘要和图表
• “Revert to Saved”（恢复保存）和“undo”（撤销）功能
• 编辑氨基酸和核酸序列
• 简单的指定色彩表编辑，蛋白质和核酸序列使用不同的色彩表
• 排列易感的描影法以信息为根据，其中包括排列位置
• BioEdit 能够读写GenBank, Fasta, NBRF/PIR, Phylip 3.2 和 Phylip 4格式，能够读ClustalW 和 GCG格式.
• 10个附加格式的导入／输出过滤器（使用Don Gilbert的ReadSeq）
• 导入/添加一个文件到最后的另一个文件上(不考虑文件格式)。
• 基本的多文本编辑器
•限制性内切酶图谱用于任何或所有形式的翻译、复酶和输出选项，包括酶的提供者和环状DNA选项。
• 游览限制性内切酶创造商
• 自动连接到你喜欢的网页游览器（如：Netscape 或 Internet Explorer）

程序和程序组的概述

BioEdit是用Borland C++ Builder 3.0编写的(开始时是用C++ Builder 1.0)。这是（曾经是）Borland公司的最新C++产品，它结合了Borland C++ 5和Delphi的可视要素库（VCL），允许用户界面的可视开发。
使用快速申请开发（RAD）环境的好处在于它能够容易的创造出大量的图形界面。它的缺点是编码不轻便。BioEdit只能在Windows 95, 98, NT and 2000中使用。我原来计划可以使BioEdit在Win16使用，但是自从Windows 3.x过时了以后，我就不再计划这样做了。

组织：BioEdit当前支持同时编辑最多50个文件。主要的控制形式包括打开文件的菜单，创建新文档，调整球形选项如色彩表，密码子表，分析参数选择，和一个窗口管理器。最初，每个文档有它自己的整套处理菜单，可以限制文档，然而，这被一个更传统的多文档界面所替代。BioEdit没有使用额外的物理存储器（除非编辑大的排列），但是它看起来像占用了很多资源。BioEdit每个文档最多可以有20,000个序列，但在序列大小上没有限制。在80MbRAM的233MHz的个人计算机上，可以很好的处理一个来自于核糖体数据库的完整的原核16S rRNA排列（6205个序列，每一个有3319个字符）。一旦用BioEdit格式保存，这个文件可以在几秒钟打开（用GenBank格式要几分钟才能打开）。

程序文件(BioEdit.exe)可以在主安装目录中找到。可能还有以下子目录：
• apps（附件程序、网页和网页书签）
通常，以下文件会出现在apps文件夹（按名称排列）：

accApp.ini (在首次安装时为accApp.def)
Bblast.html
BioEdit.html
blast_adv.gif
blast_form_0.gif
blastall.exe (在没有BLAST的版本中不出现)

blastcl3.exe (在没有BLAST的版本中不出现)
blast.txt
bookmark.txt
cap.doc
cap.exe
clear_inp.gif
clustalw.exe
clustalw.txt
cutter.html
Dnadist.doc
Dnadist.exe
Dnamlk.doc
Dnamlk.exe
Dos4gw.exe (PHYLIP 程序需要)
Expasy.gif
fastDNAml.doc
fastdnaml.exe
Fitch.doc
Fitch.exe
formatdb.exe (在没有BLAST的版本中不出现)
IdPlot.exe
isrecsmall.gif
Kitsch.doc
Kitsch.exe
mod_ad.gif
mod_submit.gif
nnpredict.html
Note.gif
PFSCAN_form.html
phi_blast.gif
PHIBlast.html
Phylip.map
Protdist.doc
Protdist.exe
Protpars.doc
Protpars.exe
psi_blast.gif
PSIBlast.html
Readseq.exe
ReadSeq.txt
scnpsit1.html
Siblogo.gif
smweb.gif

• database (是局部的BLAST数据库，安装的版本必须有BLAST工具).
BioEdit (全版本) 有以下文件在database文件夹：
Ecoli.phr
Ecoli.pin
Ecoli.psq
Ecoli_ORFs.txt (*E. coli* 开放读码框架的文本文件).

• help
BioEdit.cnt
BioEdit.GID (不是安装来的，出现在帮助文件第一次使用后)
Bioedit.hlp

• tables
Blosum62
codon.tab
color.tab
dayhoff
defcolor.tab
enzyme.tab
Gc.val
gonnet
Identify
match
Pam120
Pam250
Pam40
Pam80
Seqcode.val

安装文件夹通常包括以下文件：：
_deisreg.isr (安装相关文件)
_isreg32.dll (安装相关文件)
BioEdit.exe (BioEdit 执行文件)
DeIsL1.isu (安装相关文件)
RNaseP_prot.gb (蛋白质排列示例)
RNaseP_prot_genes.gb (DNA排列示例)
RNaseP_RNA.gb (RNA排列示例)
PBSSK_plus.pmd (质粒绘图示例)
bacterio.gb (附带GenBank 信息的蛋白质排列示例)
bacterio.bio (附带GenBank信息、图式注解、记号标记和序列族的BioEdit文件示例)
YopD.gb (附带GenBank信息的另一个示例文件)
TreeView.zip (Roderic D.M. Page编写的极好的系统进化树阅读器，完全安装才有)
TreeView.txt (记录TreeView的安装信息和配置BioEdit与tree-generating附件的连接)。

license.txt (BioEdit 许可证协议)
ReadMe.txt (总说明)

重要的是，文件夹和文件的名字不能更改，如果更改了BioEdit将不能正确安装。

将会有一个BioEdit.ini文件出现在你的Windows主目录下，它包含BioEdit的初始化默认值和参数选择。虽然这个文件可以手动编辑，但是我们推荐不要编辑和手动编辑这个文件。

当前被支持功能部件和已知问题的列表，请看BioEdit的功能和已知问题 / 局限性。

已知问题和局限性

BioEdit想要成为一个处理个别简单序列的多用途界面，带有适合于自动化多重排列选项的综合序列排列，最佳成对排列，并且着重于使手工排列更容易。随着时间的推移，增加了一些附件的功能（质粒绘图、限制性内切酶图谱、ABI和SCF查阅、RNA比较分析和其他功能中的图式注解）。然而，常用的查找功能、特殊化分析，如蛋白质二级结构、三级结构的预测、RNA结构的热动力学预测、排列性质的统计学分析、序列模式的概率或神经网络模型、排列和结构的预测，不包括在这个程序之内。

虽然用户可以配置命令行附件应用软件，有程序链接连到ClustalW、局域BLAST和BLAST client 3。但是在ClustalW程序或BLAST程序升级后，不能保证这些链接正确工作。虽然在BioEdit安装程序中提供的局域BLAST和Clustal程序将会继续工作，但在下一次NCBI决定改变它的委托人时，BLAST client 3将不能正常工作。我也不再一直支持这个程序。源代码将在稍后提供下载，但是会有一些紊乱，没有很好注释，限制于Borland C++ Builder（这是我毫无疑惑的发布源代码的原因。）

同样，自动网页链接为网页（如：BLAST、PSI-BLAST、PROSITE轮廓扫描网页）提供一个选择序列，它们的工作依赖于网页的局域HTML模板，BioEdit编辑的资源包括查询文本区域的选择序列。因为万维网的高度易变性，这些也许不能长时间正常工作。如果一些地址变化，或者HTML界面充分改变，这些将不再能正确工作。它们可能可以在BioEdit/apps文件夹中局部的被新的同名更新网页所替代，但是它们是否能正常工作将依赖于，网页中必需的URL定位是否被指定为绝对路径或相对路径，它们是否依赖于局域CGI或Java程序和其他潜在的问题。

想要配置命名行分析程序的界面很好的工作，可能不需要复杂的scripting语言。然而，因为这个界面及其选项的静态特点，可能有程序不能正确的通过BioEdit运行，虽然绝大多数接受命令行的程序可以被设置。总之，许多人可能宁愿为了更好的控制选项而从命令行运行程序。

BioEidt可以很好显示合适大小的排列。然而，对于一次打开的排列文档数量有限制，同样一个单一排列中的序列数量也有限制。现在，最多一次打开50个排列文档，一个排列中的最多序列数是20,000

序列数量的限制和序列长度是无关的。排列的绝对大小是有效的系统内存决定的。如果文档在系统中全部进入虚拟内存，编辑将会变得很慢。如果排列中有几千个rRNA基因，或者全

部基因组的序列列表，在Win95/98或NT系统中，至少需要64到128Mb的内存，在Win2000系统中，至少需要128Mb内存。

在排列矩阵N× M > 40,000,000 (N = 序列数，M=最长序列长度)时，Undo（撤消）选项自动失效。

BioEdit是由Borland C++ Builder编写的，是100% Windows基础。它是不可移植的。因为这个程序的大部分是图形界面，在UNIX或Mac中可能不好使用。

BioEdit使用手册

序列编辑 / 处理

手工序列排列

下面是基本的BioEdit排列文档窗口。如果你不喜欢现在的样子不要当心，字体、大小、背景颜色、残基颜色和标题窗口宽度都可以改变。鼠标箭头右下方的黄色条幅显示的是当前序列的绝对位置。这同样显示在控制栏的“Position”标题。选择关闭黄色条幅，就进入“View->show sequence position by mouse arrow”。

总的手工排序功能是：

在编辑窗口有三个可应用的基本模式：

选项可在“Sequence->Edit Mode”中找到。

Select / Slide mode(选择/调整模式)：用鼠标左键选择框住的残基。用鼠标来回的拖动选择。默认值是朝你滑动的方向忽略“unlocked gaps”，并在所选择的另一边开启新的“unlocked gaps”。为了移动所选择的全部序列的下游，不管缺口，在移动时按住“shift”键。你也可以在按钮板上切换合适的按钮（见后），改变默认值为：移动所选择的全部序列的下游。选定选项后，在滑动时用“shift”键忽略“unlocked gaps”。用“shift”键选择所有在现在选定的和新选择的残基。

“CTRL”键可以在当前选择上增加一个新的选择（例如，你也许想在三个互不相连的序列中选择残基）。

Edit mode（编辑模式）：在编辑残基模式中，你可以在文档的任何位置（除了标题）放置任何类型的光标。用箭头你可以在序列中走来走去。编辑有两种形式：插入和改写。当编辑器在编辑模式，可以看见在编辑模式的下拉菜单中有一个选项：

在其它两个排列模式,这个选项不会出现.

Grab & Drag mode(抓取/拖动模式)：从“mode”目录中选择“Grab & Drag”，或者切换“G/D”按钮（见后），你可以从屏幕上动态的抓取和拖动单个残基。用“shift”键移动整个残基序列的下游（或者在按钮板上切换成合适的按钮－－见后）。

Grouping of sequences（序列分组）： Sequences may be grouped into groups (or "families").序列可以进行分组（或分成“家族”）。 一个组的序列排列可以相互锁定，意味着手动调节（用可调整的残基插入或／和删除缺口）将自动同步调节一个锁定的组。This ***only*** applies to sliding resides (Select / slide mode or Grab & Drag mode), not to single insertions and deletions of gaps with right mouse clicks. For information on grouping sequences and locking the alignment of groups of sequences, see grouping sequences.
这只适合于可调整的残基（Select / slide mode或Grab & Drag mode），不能用鼠标右键进行单个缺口的插入和删除。想了解有关序列分组和其排列锁定的信息，看“grouping sequences”。

工具条 **/** 加速按钮：


锁定和开启全部序列的所有缺口。当打开一个排列，这个按钮是在开启状态，但是缺口是现在的，虽然它们过去被保存。在这个按钮被按下去后，才能进行改变。为了开启当前序列的所有缺口，你必须按这个按钮两次进行切换到这个状态（第一状态是锁定所有缺口）。


上个按钮的锁定状态。


按下这个按钮，可以用鼠标右键插入单个缺口。


用鼠标右键删除缺口。


在所有序列中插入缺口，除了在用鼠标右键点击这个按钮的位置。
在所有序列中插入缺口，除了在用鼠标右键点击这个按钮的位置。在选择位置没有缺口的序列将不会改变，但是有这个按钮在那儿，缺口将始终被删除。


转换鼠标左键和右键的默认值功能


切换“Grab & Drag”模式


按下这个按钮，可调整残基的默认值是忽略或扩展到下游缺口。使用“shift”键可以调整转换这个功能。


按下这个按钮，可调整残基的默认值是移动全部所选序列的下游，胜过忽略或扩展到下游缺口。使用“shift”键可以调整转换这个功能。


普通视图模式。当序列颜色显示时，残基根据当前的色彩表着色。这个选项用于序列是单色视图时。所有其他视图覆盖单色视图。


反转颜色视图模式。背景栏根据每一个残基的色彩表描影。残基的颜色是它们普通颜色的反转。


“排列的强度”——残基根据每一栏的信息内容灰度描影。


残基背景根据每一栏的信息内容描影。


把文档窗口中一致的和类似的残基描影。按下这个按钮，控制条上将会出现一个下拉菜单，可以控制隐藏的百分比开端。蛋白质排列的类似性隐藏的矩阵文件可以在“Alignment->Similarity Matrix”菜单中详细说明。


绘出功能部件，其上有层次的序列。


只绘出功能部件，没有序列。


根据当前的色彩表，序列彩色视图。


根据当前选择的序列颜色，序列单色视图。只用于“normal view”按钮也被按下。


用一个字符（默认值是‘.’）显示序列的同一性（默认值是top）.


如果按下前一个按钮，这个下拉菜单能够选择标记同一性的字符。


显示或隐藏交互信息检查器（只用于RNA分析）


引出色彩表编辑对话窗。


切换“ignore anchor points”模式。如果这个按钮没有按下，固定栏限制排列的范围。按下这个按钮，固定栏被忽视。

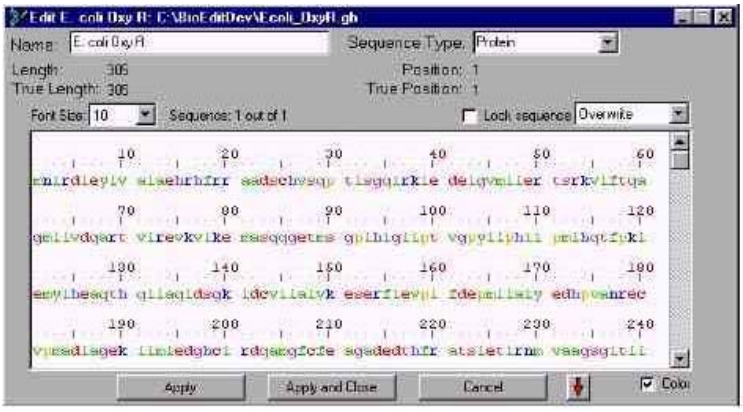
卷屏速度控制器：控制水平卷屏条（卷屏是因残基增加）。

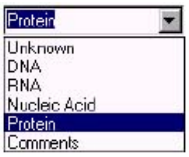

增加或移去位置标记旗

增加或移去一个栏的固定点

在编辑盒中编辑

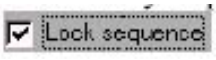
在一个文本窗口中，进行一个序列主要的编辑会十分方便。为一个序列开启一个编辑窗口，双击序列的标题，或选中序列并从“Sequence”菜单中选择“Edit Sequence”。为了使改变生效，必须按下“Apply”或“Apply and Close”按钮。取消将不会改变序列。在一个序列第一次编辑时，将会出现下面的窗口。





在“Sequence Type”下拉菜单中，下列选项是可用的。如果一个序列是“未知的”，蛋白质色彩表通常是彩色的，就像一个已经经过类似性底纹处理的蛋白质序列。

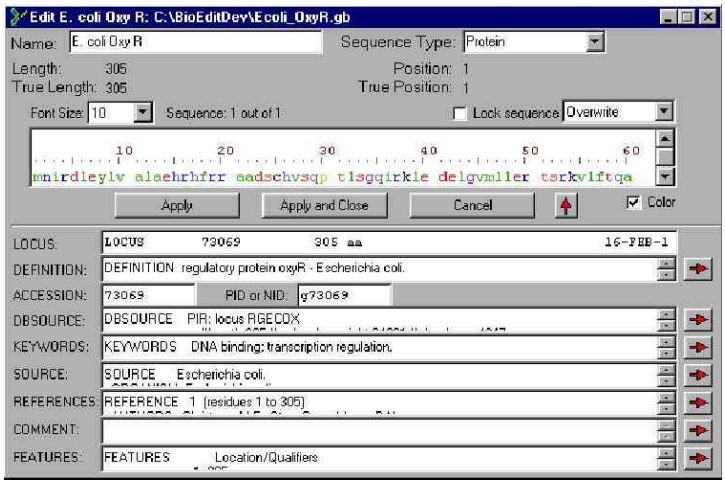
可以保留一个关于排列的每一行的屏幕信息的注解，但是不能计算类似性和同一性，不服从标准的处理，如翻译、互补、自动排列等。



在单个序列编辑器中，你可以用“lock sequence”选项选择锁定任何序列。

应用这个选项时，“selecting/dragging”，或抓取和拖动将不能使用。但是用鼠标右键增加或删除缺口始终可以使用。

按下  按钮，可以展开窗口看相关的GenBank的信息。窗口扩展如下：



 按钮可以用于提出在大的编辑窗口中的相关领域。

**注意：GenBank信息将只能用GenBank或BioEdit格式保存。

***注意：GenBank信息，包括“功能部件”领域，是内部独立于用户定义的图示注解。

窗口隐藏

一个文档可以进行“窗口隐藏”，就是双击窗口的标题栏可以隐藏标题栏。再次双击可以使其变回原来的大小。它也可以最小化和最大化。

增加一个新序列

通过以下方式增加新序列：

1.在“Sequence”菜单下选择“New Sequence”选项。序列可以像原始文本一样被键入或复制进序列窗口。按下“Apply”按钮，可以在文档中增加序列。

2.通过“Edit”菜单的“Copy Sequence(s)”和“Paste Sequence(s)”命令，复制或粘贴来自其他BioEdit文档的序列。同样，也可以使用当前菜单快捷键(默认值：Ctrl+F8复制，Ctrl+F9粘贴)。

全屏编辑

序列可以在全屏编辑，就像在一个文字处理器上一样。必须首先设定“Mode”选项为“Edit Residues”(BioEdit在安装后默认模式为“Slide Residue”)

在编辑模式下，你可以使用箭头在屏幕上移动、输入像在文本编辑器中一样。编辑有两种选项：插入模式和改写模式，它们类似于在文字编辑器中的功能。

选择序列

点击序列的标题可以选中序列。拖划出一个方框可以选中多个序列，或用“shift”键选择两个选择序列之间的所有序列。用“Ctrl”键加鼠标可以分别选择标题，或给选中的序列加上详细的标题。双击标题将会打开一个单序列编辑器。再次点击原先选中的标题，使其进入全

屏编辑模式。你可以编辑标题后，按下< return >或点击序列标题板的任何位置，使对标题的改动生效。

移动序列

想移动一个序列(或一些序列)，选中它(用鼠标左键点击它的标题，使其变亮)，把它拖放到你想要的位置。

Cut / Copy / Paste（剪切 / 复制 / 粘贴）

Copy（复制）：

编辑窗口的文本(序列残基)：用鼠标选择文本，并从“Edit”菜单选择“Copy”。不像文字编辑器，你可以复制你想选择的区域，而不是复制文本的全部行。这种方式复制的区域可以粘贴在任何能够进行文本编辑的程序中。

如果，只是如果，你没有选中在全部序列中任何残基，序列的标题将会以BioEdit序列结构形式复制到BioEdit的剪贴板，在选择“Paste Sequence(s)”时，全部序列将会被粘贴到文档。

全部序列：用鼠标选择序列标题，并从“Edit”菜单选择“Copy Sequence(s)”。标题被选中的序列将以Fasta格式被复制到Windows剪贴板。多于一个被选中的序列将以Fasta序列目录的形式复制到剪贴板中，并在BioEdit内部复制成一组全部BioEdit序列结构，能够被粘贴在任何BioEdit文档中。

注意：BioEdit剪贴板中包括所有序列相关数据（Genbank信息、图示注解），是在BioEdit同一步骤的内部（它们不能在独立的步骤之间转移）。为了在BioEdit排列文档之间复制序列，必须确定两个文档是在程序的同一步骤打开的，只有Fasta格式的序列可以被复制到普通的Windows剪贴板。

Paste（粘贴）：

在编辑窗中的文本：为了把一个序列粘贴入主编辑窗，界面必须是“Edit Residues”模式（见全屏编辑）。如果文本的一个区域被粘贴到一个序列，只有第一行（用回车键定义） 将会被粘贴。这避免了在粘贴文本进入序列时可能出现的问题，也避免了不注意的使错误的序列在其下。为了把文本的片段粘贴到排列的一个区域，片段必须一次一个的粘贴进序列。如果文档在“Slide Residues”或“Grab and Drag”模式，“Paste（粘贴）”的功能将会和“Paste Sequence(s)（粘贴序列）”的功能一样。（见后）

全部序列：从文档菜单到粘贴序列，从“Edit”菜单中选择“Paste Sequence(s)”。序列将会增加到文档的最后。它们可以移动到文档的任何位置。

“Cut（剪切）”和“Cut Sequence(s)（剪切序列）”：就象“Copy（复制）”和“Copy Sequences（复制序列）”一样，但是其功能是从文档中删除复制的信息。然而，只有在“Edit Residues” 模式下，残基才能从文档中删除。同样，当在没有选中任何残基的情况下使用剪切功能时，标题被选中的序列将以Fasta格式被复制到Windows剪贴板，并以序列结构的形式复制到BioEdit剪贴板中，但是它们不能从文档中删除。为了适当的从文档中剪切序列，可以选择“Cut Sequence(s)” 。

Minimizing an Alignment（排列的最小化）
当一个排列手工处理时，当序列定期的增加并手工排列到一个现有的排列中，缺口经常导致一个专栏出现在每一个序列中。为了在不改变现有排列的情况下移动缺口，选择“Minimize Alignment”。

Basic Manipulations / Sequence Menu（基本处理 / 序列菜单）
一些简单的序列处理可以通过BioEdit的一个菜单选项自动完成。这些选项在“Sequence”中。

在BioEdit中Masks（屏蔽）在这一点上有一点薄弱，主要用于RNA比较分析功能。关于在BioEdit中如果使用屏蔽，看“Masks”。

锁定和开启缺口：当残基在序列中滑动时，一个锁定的缺口将不能被压缩。为了锁定缺口，选择想要锁定的缺口后选择“Lock Gaps”。想要锁定序列中的所有缺口，选择序列的标题后选择“Lock Gaps”。想要锁定一个排列中的缺口，切换“lock/unlock”按钮进入锁定状态：

开启的缺口就是锁定状态的相反。想要开启一个排列中的所有缺口，切换“lock/unlock”按钮进入开启状态：

“Degap”选项可以移动所有开启的缺口。它也可以移动被选中标题的序列中的所有开启的缺口。
注意：‘～’和‘.’（表示开启的缺口），‘—’表示锁定的缺口。
这个惯例用于BioEdit中每一个窗口和功能。如果一个句点没有经过BioEdit加上的缺口特点，但也有一种加工过的缺口类型，为了程序的可计算性，宁愿使用BioEdit中的缺口特点。同样，一些程序可能使用一个句点去表示排列位置中没有，残基或缺口，但是只是在序列的开始或结尾。BioEdit不直接注意这种差别。序列行之前或之后的位置被加工成缺口，而且BioEdit假定每一个排列包含有真正的同源序列（尽管BioEdit也被设计成允许用户忽视程序的排列中心，并只使用它处理序列的数据）。

Sequence Menu（序列菜单） (不包括“mask”功能)

New Sequence（新序列）： 创造新序列，开启一个单一序列编辑器。

Edit Sequence（编辑序列）： 在单一序列编辑器中开启首次选择的序列。

Select Positions（选择位置）：开启一个对话框，允许在所有选中的序列中选择具体位置。

Open at cursor position（在光标处打开）： 如果文档处于编辑状态，光标同时出现，这个选项将在单一序列编辑器中打开光标当前所在位置的序列。

Rename（重命名）： 根据子菜单选项重命名序列标题：
Edit title：改变屏幕上序列的标题。
with LOCUS： 把所有选中的标题改为LOCUS栏内容。

with DEFINITION：把所有选中的标题改为DEFINITION栏内容。
with ACCESSION：把所有选中的标题改为ACCESSION栏内容。
with PID/NID：把所有选中的标题改为PID/NID栏内容。

Sort（分类）：根据下列标准进行序列分类：
By Title（标题）
By Locus（位置）
By Definition（定义）
By Accession（增加）
By PID or NID
By Reference（参考值）
By Comment（注释）
By residue frequency in a selected column（所选栏中的残基频率）

选择最后的选项（by residue frequency）时，必须选择单一的残基栏，根据有效的残基的频率分类。

Pairwise alignment（成对排列）： 两个序列的最佳排列
Align two sequences（排列两个序列）(optimal GLOBAL alignment最佳GLOBAL排列)：使用基于Smith和Wasterman最佳排列模式的全球排列运算法则，最佳排列两个序列。

Align two sequence（排列两个序列）(allow ends to slide允许末端滑动)：使用基于Gotoh修正的Smith和Wasterman最佳排列模式的全球排列运算法则，最佳排列两个序列，使其不能约束序列末端（每一个序列末端允许自由滑动到其它序列）。这种排列适用于快速识别小序列的序列阅读区的重叠区域，而不需要自动邻近装配程序。

Calculate identity/similarity for two sequences（计算两个序列的同一性和类似性）： 依据两个序列当前在文档中的排列（不要排列它们），计算其同一性和类似性（根据当前类似的矩阵）。

Similarity Matrix（类似矩阵）（用于成对排列和滑动）：这些矩阵只用于氨基酸序列。BioEdit不用矩阵设计核酸。
BLOSUM62：BLAST的默认矩阵。BLOSUM矩阵通常有利于数据查询和适度设想大的进化距离（很小的BLOSUM数值＝很大的进化距离－－只限于BioEdit提供BLOSUM62 matrix [intermediate]）
PAM40：适用于十分密切相关的序列（40PAM单位＝相关性很小的进化距离－－在PAM矩阵中，大的PAM数值＝很大的进化距离）
PAM80
PAM120
PAM250：适用于十分疏远的相关序列（大的PAM距离）
DENTIFY：简单匹配或错配的矩阵，用很大的数值（－10000）处罚错配
DAYHOFF：是一种PAM矩阵－－M.O. Dayhoff的最原始的PAM250矩阵（每个数值取最近的整数）
MATCH：简单匹配或错配的矩阵，用a –1处罚错配，用a +1记录匹配。

GONNET：一种1992年Gonnet推荐的改进的PAM250矩阵。

Features（功能部件） (功能部件注解功能)：

Automatically annotate from GenBank Feature Fields（来自GenBank功能部件区域的自动注解）：这个选项允许你根据原有的为这个序列存储的GenBank数据，增加功能部件。

Edit Features（编辑功能部件）：增加、修改或删除序列功能部件。

Annotate Selection（注释选择）：在所有序列中，增加一个将会跨越当前选择位置的功能部件。

注释选择序列像一个模板用于第一个序列

Sequence groups (or families)（序列组或家族）：序列分组或不分组，编辑当前序列组。

Edit Mode（编辑模式）：设定当前编辑模式。详见序列排列手册。

Mask（隐藏） (隐藏上面选项)

Toggle color（切换颜色）：切换单个序列颜色。是早前的版本留下来的，没有什么用途。

Gaps：

Lock gaps, Unlock gaps and Degap（锁定缺口、开启缺口和删除缺口）： 详见上文。

Insert multiple gaps（插入多个缺口）：排列窗口中，在当前选择的位置上插入一定数量的缺口。

Manipulations：不依赖于序列类型的简单处理。

lowercase and UPPERCASE（小写字母和大写字母）：只指示序列，不包括标题。

Reverse（反转）：反转任何序列

Remove numbers（移动数字）：用于指示。增加后，可以轻松的表示部分来自于GenBank格式的文本文件和网页的序列。

World Wide Web（万维网）

Automated links are provided to the following selected WWW search functions：以下万维网搜索功能，提供自动链接：

BLAST, PSI-BLAST和PHI-BLAST.

Prosite profile and pattern scans（层面和模式扫描）

nnPredict protein secondary structure prediction（蛋白质二级结构预测）

Nucleic Acid（核酸）

Nucleotide Composition（核苷酸组成）：绘制核苷酸组成图，给出一个包括G+C和A+T百分比和分子重量的概要。

Complement（互补）：DNA或RNA序列的互补。这个选项对于蛋白质序列无用，在五种标准碱基（A、G、C、T和U）以外的字符和嘌呤／嘧啶是没有影响的（和嘌呤“R”互补的是嘧啶“Y”）

Reverse complement（反向互补）：与互补相似，但是序列相反。

DNA->RNA and RNA->DNA：在序列中切换”T”和”U”

Translate（翻译）：翻译在方框1、2或3中序列，或者翻译现在选择的序列范围。密码子用空格分开。核苷酸序列显示在蛋白质序列的上方。翻译的序列根据参数选择，用三个字母或一个字母的氨基酸符号表示。如果被选中的序列部分已经被翻译，根据现在的参数选择，全部核酸序列或只有翻译的序列部分显示。一个汇总表可能显示在翻译序列下方，显示密码子出现在序列中的次数，以及根据提供的密码表每个具体氨基酸的密码子的频率。

Find Next ORF（寻找下一个开放读码框架ORF）：根据参数设定，从最后的选择点开始，搜索当前选择的序列中的ORF。

Create plasmid from sequence（从序列中创造质粒）：一个DNA序列可以直接修改进入质粒／载体。序列中自动执行一个限制性内切酶图谱。想要详细了解质粒，看“Plasmid drawing with BioEdit”

Restriction Map（限制性内切酶图谱）：在DNA或RNA序列中执行一个限制性内切酶图谱。

Sorted and Unsorted six frame translations （分类和不分类六个翻译框）：在所有六个文本框中，根据指定的起始密码子（ATG，“任何”或用户定义的），翻译核酸序列，并最大化和最小化ORF。分类翻译最多输出几千个ORF。为了进行一个未加工的全部基因组（或较大基因组）的翻译，使用不分类翻译（在不分类翻译中，输出的数据可以像文件直接打印，只需要很小的内存）。

Protein（蛋白质）：

Amino Acid composition（氨基酸成分）：给出一个蛋白质氨基酸成分的图和摘要，包括分子重量。

Hydrophobicity profiles（疏水面）：

使用Kypt和Doolittle（1982）的方法，选择疏水等级来计算疏水面平均数“Mean hydrophobicity”。

使用Eisenberg et. al.（1984）的方法计算疏水力矩“Hydrophobic moment”。在这儿不使用其运算法则中寻找跨膜的α螺旋的方法，而是每一个残基划分用户定义的序列片段中的疏水力矩（每一个残基表现用户定义片段的开端）。

Mean hydrophobic moment（疏水力矩平均数）：对于每一个残基，一个相同尺寸窗口的疏水力矩平均数用于计算每一个疏水力矩。

注意：我没有专家的权力去预言这些疏水面图。BioEdit不给疏水的和／或跨膜蛋白质片段下结论，这些图需要用户自己去判断。

想要了解这些图标的方法和意义的描述，以及疏水比例的参考和疏水分析运算法则，看“Hydrophobicity Profiles”

Translate or Reverse-Translate(翻译或反翻译)：根据BioEdit.ini文件中的密码表，把DNA或RNA翻译成蛋白质。默认值是“/tables”目录下的“codon.tab”。默认值是来源于J. Michael Cherry (cherry@frodo.mgh.harvard.edu)用GCG程序CodonFrequency编辑的*E. coli*密码子选择表。这种格式的密码子表可以使用，但是这个格式的密码子表必须被BioEdit承认。想要选择不同的表，看“Codon Tables”。蛋白质序列可以根据每个具体的氨基酸密码子参数，被还原成DNA碱基。显然，如果一个核酸序列被翻译成蛋白质和被还原回来，都将会丢失信息。

Translate in Selected Frame (Permanent)(被选框中翻译(持续的))：如果当前被选的框是＋1框（如果选中多个框，默认值是选择的始端），这个选择允许你翻译这个核酸序列。当

应用于一个蛋白质序列时，它通常像上一个选项一样，导致一个还原的反翻译。

Toggle Translation（切换翻译）：在核酸和编码蛋白质序列中切换核苷酸序列，允许用于任何视图的序列排列。具体看“Toggling between nucleotide and protein views”

Toggle Translation in selected frame（在被选框中切换翻译）：如果当前被选的框是＋1框（如果选中多个框，默认值是选择的始端），这个选项允许你切换翻译视图（不丢失核苷酸信息）。

Dot Plot (pairwise comparison)（点图（成对比较））：相互比较两个序列的矩阵，生成一个点图。

Customizing the View（用户自定义视图）

BioEdit当前支持下列视图选项：

- Background colors for sequence and title windows（序列和标题窗口的背景色）
- Default monochrome sequence and title colors（默认单色序列和标题色）
- Character fonts（字符字体）
- Font size（字体大小）
- View sequences in bold-face type（用粗体显示序列）
- View sequences in monochrome or color (editing is faster in monochrome)（单色或彩色显示序列（在单色中编辑速度快一些））
- Normal color view (residues colored)（普通颜色视图（残基颜色））
- Inverse (background colored)（反转（背景色））
- Strength of alignment（排列强度）：根据每一个位置包含的信息进行描影
- Strength of Alignment – Inverse（排列强度－－反转）：和Strength of Alignment相同，背景代替残基被描影。
- Identity/Similarity shading（同一性／类似性描影）：如果残基在框中的频率等于或超过用户定义的终止点，残基将会根据色彩表定义的颜色进行背景描影。核苷酸排列只能根据同一性进行描影。蛋白质可以根据同一性和类似性描影，类似性是根据当前选择的氨基酸类似性得分矩阵计算的。只有定义为有效残基和无注解序列，才能进行类似性和同一性计算。
- Sequences and Graphical Features（序列和图形功能）：用有层次的序列，在文档屏幕顶端，绘制图形序列注解
- Graphical Features（图形功能）：用卡通模式绘制图形序列注解，不画出残基字符。使用这种模式时，在窗口顶端出现一个比例因数的活动栏，使其能够在1：1和1：32768之间选择比例因数。
- Conservation plot（保存图表）：如果根据用户定义的标准（默认值是序列顶端）判断，残基是在相同栏中的同一种残基，残基将会根据用户的定义字符而绘图（默认值是一个周期）。如果你想要保存图表的新标准，用鼠标右键点击序列标题，可以改变序列标准（参数）
- Show or hide the mutual information examiner（显示或隐藏多信息编辑器（只用于RNA比较分析））
- Show or Hide the translation toggling control（显示或隐藏翻译切换控制）。由于空间的限制，这在共有信息编辑器控制中是相互独立的。
- Show sequence position by mouse arrow（用鼠标箭头显示序列位置）：在文档窗口的序列中移动鼠标，鼠标的绝对位置（忽略缺口）将会出现在序列视图窗口上的控制条中。位

置也可以通过鼠标箭头显示（包括标题的全部长度），这个选项可以使这个功能使用或停止。
• Split window vertically（垂直拆分窗口）：在文档窗口内部产生一个复制窗口，并和当前文档同步。文档窗口被一个垂直窗口拆分栏拆分（它是真正同步的文档）。两个窗口的垂直滑动位置是共用的，而水平滑动栏是相互独立的。窗口可以通过主文档的窗口拆分栏来调整大小。如果再次选择这个选项，窗口将会恢复正常。
• Split window horizontally（水平拆分窗口）：产生一个水平窗口，直接放置在原窗口的下方，导致原窗口底部和新窗口顶部的边界想一个窗口的拆分栏。如果再次选择这个选项，这个窗口将会消除。
• Save options as default （将选项保存为默认值）：当“Auto-update view options”关闭时，选择这个选项将把当前文档的视图选项保存为所有新建文档的默认值。
• Auto-update view options （自动更新视图选项）：当点击这个选项时，所有文档视图和参数改变将自动保存为新文档的默认值。
• Customize menu shortcuts（定制菜单快捷键）：开启一个对话框，允许改变菜单的快捷键为任何组合键。
• Hide control bar or Show control bar（隐藏或显示控制条）：为了适合一个框窗口屏幕上的更多序列，主控制条可以取消。如果隐藏控制条，“Show control bar”（显示控制条）可以使用。如果控制条是隐藏的，序列编辑方式可以通过“Sequence->Edit Mode submenus”来改变。菜单的默认值也将改变。

想要改变这些设置，在一个打开的文档的“View”菜单中选择适当的选项。想要使任何具体的文档的现在的视图选项成为所有后来开启文档的默认值，选择“Save Options as Default”。
这些视图选项可以通过“View”菜单，或通过点击排列窗口中合适的按钮来选择。

Color Table（色彩表）
所有的文档使用一个色彩表。这个色彩表文件是“color.tab”，可以在BioEdit的安装目录中的\tables目录中找到（详见Program Organization）。虽然，色彩表可以通过手动编辑，更容易的是在主应用程序控制条中的“Options”菜单下选择“Color Table”选项。

（编辑色彩表）：想要编辑色彩表，在主应用程序控制条中的“Options”菜单下选择“Color Table”选项。核苷酸和蛋白质的色彩表是不同的。想要改变残基的颜色，双击残基上的颜色盒，进入一个颜色对话框。想要增加或删除残基，点击“＋”或“－”按钮。在出现的窗口中，为了增加或删除残基而推动按钮。当增加残基时，它将像默认色一样变黑。这个颜色会变成你想要的颜色。

想要手工编辑色彩表，必须注意观察以下格式：
•用一条数据行指示每个色彩表，包括精确文本“/amino acids/”（氨基酸）或“/nucleotides/”（没有双引号）。
• 用“/////”（没有双引号）指示每一个表的末端。
• 用文件中的两条程序行详细指示每一个残基的颜色：
• Line 1:一个三个字节十六进制数字（或它的整数值）。这三个字节分别用蓝色、绿色和红色显示数值。
•Line 2:一个包括所有字符的列表显示所有将会是这种颜色的字符。如果文件中的一个字符颜色在别处被重新定义，最后一次定义的颜色是有效的。

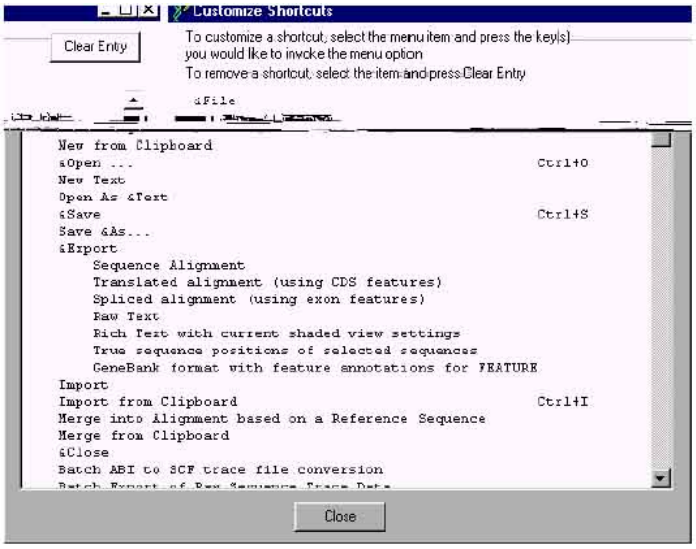
注意：手动编辑色彩表不是必需的，也不推荐。

如果色彩表被破坏，它可能导致程序启动失败或在编辑色彩表时程序失败。如果发生了这种情况，你可以删除色彩表并建立一个新的，一次一个残基（你将在启动和选择色彩表时出现一个错误。但是，在按下“Save Table”（保存表格）按钮时，程序将建立一个新的色彩表）。这是单调乏味的，因此在BioEdit的/tables文件夹下有一个“defcolor.tab”文件。如果色彩表被破坏，你可以复制“defcolor.tab”文件，并将文件名改为“color.tab”。

Customizing menu shortcuts（用户自定义菜单快捷键）

首选菜单快捷键可以编辑任何菜单项目或次级菜单项目（不能作用于三级菜单）。然而，快捷键只能在排列文档窗口中定制。例如，如果Ctrl+Y被设置为“copy”（复制）的快捷键，但是在文本编辑器中，Ctrl+C是复制的快捷键。

想要设置快捷键，选择“View->Customize Menu Shortcuts”。想要设置快捷键，只要打开感兴趣的菜单项目，用鼠标选定，压下你想确定使用的键。想要完全消除快捷键，加亮菜单项目，压下“Clear Entry”。

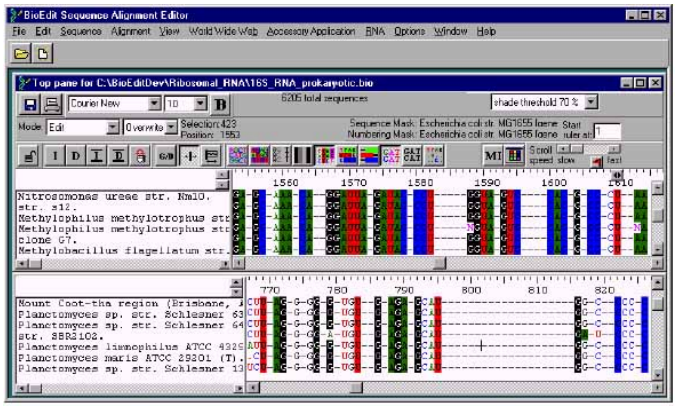


Splitting the window view（拆分窗口视图）

它可以方便的同时编辑排列的两个不同部分。BioEdit允许两种方式将文档拆分成两个同步的窗口，一种方式是垂直拆分，另一种是水平拆分。

想要垂直拆分窗口，选择“View->Split Window Vertically”。下图是原核细胞16S rRNA部分排列的拆分视图。两部分使用一个垂直拆分栏，但是在水平方向拆分是相互独立的。窗口可以用鼠标调整打小。

想要水平拆分窗口，选择“View->Split Window Horizontally”。下图是另一个原核细胞16S rRNA部分排列的拆分视图。两个窗口保持联系，但是有独立的垂直和水平拆分栏。



Sorting Sequences（序列分类）

一个排列文档中的序列将根据以下标准分类：

- Title（标题）
- LOCUS（位置）
- DEFINITION（定义）
- REFERENCES（参数）
- COMMENT（注释）
- ACCESSION（增加）
- PID/NID
- residue frequency in a selected column（被选栏中的残基频率）

想要序列分类，选择“Sequence->Sort-><sort type>”

Graphical Feature Annotations（图形化的功能部件注释）

这有时可以方便得到关于序列特定要素的信息（如外显子、内含子、螺旋、图形等等），有利于快速而简便的进行参考，而不用使用外部的来源，像笔记本、其他文件、文献或万维网。因此，在5.0.0版本中增加了序列功能部件的图形化注释。注释可以手动完成，或自动来源于现有的GenBank格式的FEATURES数据。当鼠标箭头移动到序列残基上时，在序列中横跨任何位置的功能部件的名称和描述，能够像工具提示一样显示在排列窗口的右边。标准的GenBank格式功能部件类型，是序列类型内部保存的基本要素。在定义每一个功能部件的描述时，如果一个是根据GenBank标准，BioEdit的功能部件注释功能能够为注释一个

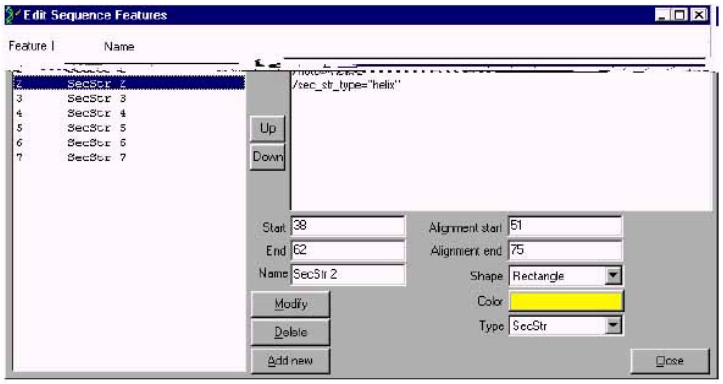
序列或一系列序列提供一个方便的途径，并能用标准GenBank格式输出，用于使用用户定义的图形化功能部件代替GenBank的FEATURES区域。这是一个用于Sequin或BankIt序列提交的有用起点。

在BioEdit中，当手动的在序列中增加一个功能部件时，“真实”的位置被计算并假设用先例替代排列中的绝对位置，而且，所有未来的排列调节与这些“真实”位置相关。例如，如果在一个功能部件中删除5个缺口，功能部件的末端回缩5个排列中位置，但是功能部件残基的绝对数量没有改变。如果在一个功能部件中删除3个碱基，功能部件目前的末端位置将会回缩3个位置。同样，当自动增加来自于GenBank信息的功能部件时，符合每一个功能部件的真实起点和末端的排列中的位置将被计算，并更新于反映当前所有要素的排列情况。BioEdit允许通知功能部件的标题、颜色、形状（矩形、椭圆形、菱形或箭头形）、方向（只是形成一个不同方向的箭头）、类型（“未定义”或67种标准GenBank功能部件类型中的任何一种）、一个描述的储量空间（长度上没有限制）。

Adding, modifying and deleting sequence features manually（手动增加、修改和删除序列功能部件）

有两种方法手动增加或修改序列功能部件：

1.点击使序列标题增亮，并选择菜单选项“Sequence->Features->Edit Features”。只能增亮一个序列标题，否则BioEdit将不知道哪一个序列编辑功能部件。将会显示下列对话框（如果没有任何功能部件增加到序列，路径将会是空的）：



想要增加一个功能部件，将功能部件的标题填写入“name”（名称）框，在“Desc”框中增加一个描述，在简单序列、非排列序列或者一个容易测定排列位置的排列序列中指定起点和末端位置，你可以指定这些插入点，BioEdit将为你指出真实位置。如果起点和末端你都填写，BioEdit将忽略排列位置并将在指定的真实位置上重新计算它们。注意：如果你想要功能部件反映方位，而方位是颠倒的，指定的起点位置将是一个高数，末端位置将是一个低数。另外，按下“Color”按钮选择颜色（你将会见到一个颜色对话框），选择形状，在“Type”中指定类型。做完这些以后，按下“Add New”（增加一个新的）在序列中增加一个新的功能部件。注意，如果两个功能部件在位置上有交迭，在序列中进一步向下的功能部件将连接在下一个功能部件的顶部。想要改变序列中功能部件的位置，点击一个或更多功能部件的标题使其增亮，按下“Up”（向上）或“Down”（向下）按钮。

想要修改现有的功能部件，用鼠标左键点击功能部件的标题，并像增加功能部件一样做同样的工作。按下“Modify”按钮，而不是像增加功能部件一样按下“Add New”按钮（按下“Add New”按钮将有效地复制功能部件）。通过一次性选择所有功能部件的标题，你可以同时修改多个功能部件的要素。所有被选择的功能部件中共有的要素将会显示。如果修改任何要素，并按下“Modify”按钮，修改将会应用于所有功能部件。

想要删除一个或更多的功能部件，点击功能部件使其增亮，按下“Delete”按钮。

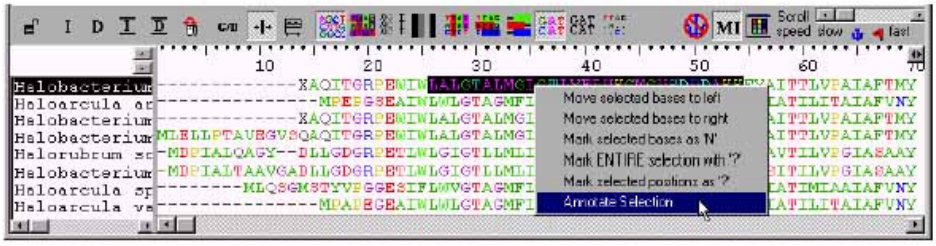
当增加或修改功能部件结束时，按下对话框右边的“Close”按钮。

2. 从排列窗口增加或修改一个功能部件：

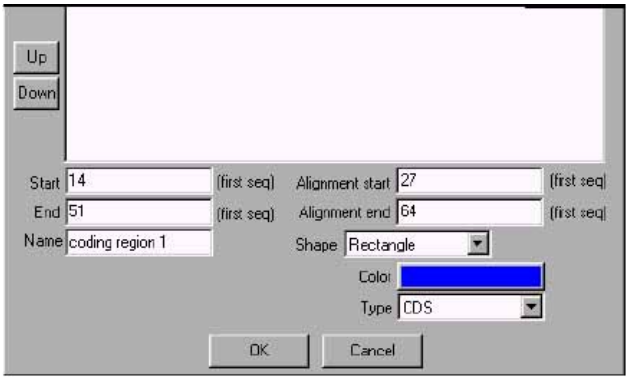
使用鼠标右键点击出现的上下菜单，在排列窗口中增加或修改功能部件。为了使用这个菜单，必须关闭所有鼠标右键激活的排列功能部件。这意味着下列四个按钮不能按下：

如果这些按钮中任何一个被按下，排列窗口的右键点击将会根据按下的按钮增加或删除缺口。

想要增加新的功能部件，在排列窗口中，点击你希望的功能部件在序列中的位置，并使其亮度增加。下一步，在增亮区域点击鼠标右键，并在出现的上下菜单中选择“Annotate Selection”。



将会出现下列对话框：



其他的选项等于手动增加功能部件，但是你必须根据排列窗口中的选择指定功能部件位置。你可以通过选择跨越多个序列的残基块来注释一个选择块。既然这样，相同的功能部件将被应用于每个序列中同样的排列位置，每个序列中的“真实”位置将根据排列位置独立更新。

想要修改现有的注释，在注释的任何位置鼠标右键点击，选择“Update Annotation”（更新注释）（只可见于你右键点击的注释，和这个区域只有一个注释时），会出现一个上面的对话框。按下OK将会更新注释，而不是增加一个新的。如果你只是鼠标右键点击注释，对话框中的位置将反映当前功能部件的起点和末端。然而，如果在右键点击前你选择一个新的注释，对话框中的位置将反映选择部位在排列中的位置，假定你想要更改功能部件的位置。你可以通过选择一个块、鼠标右键点击、选择“Update Annotation”这个方法，同时更新在一些序列中同位置的注解。

Annotating sequences automatically from existing GenBank FEATURES data（根据现有的**GenBank**的**FEATURES**数据自动注释序列）

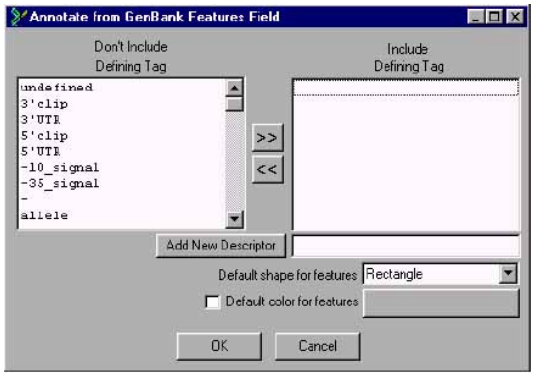
你可以根据现存的GenBank格式FEATURES数据自动的在BioEdit中增加功能部件。想要看见是否在序列中有FEATURES数据，双击序列标题，按下  按钮展开窗口，看“FEATURES”框。如果框中的数据格式内容和下列相同，期待的数据格式是自动注释：

GenBank FEATURES 区域中的格式化例子：

```
FEATURES                    Location/Qualifiers
     source                 1..247
                            /organism="Halobacterium salinarum"
                            /db_xref="taxon:2242"
     NonStdResidue          1
                            /non-std-residue="PCA NH3+"
     SecStr                 10..31
                            /note="helix 1"
                            /sec_str_type="helix"
... etc.
```

想要详细了解BioEdit将要寻找的全部标签的目录，可以看程序中使用的对话框，或者看“GenBank Format”。

想要自动注释序列，点击你要注释的任何序列的标题使其增亮（它们有GenBank的FEATURES数据），选择“Sequence->Features->Automatically annotate from GenBank Feature Fields”。将出现下列对话框：



搜索可用的标签在左边。想要在目录中增加搜索的标签，选择你想要包含的标签，按下“>>”按钮（用“<<”按钮，你可以将任何标签移回到另一边）。通过在“Add New Descriptor”（增加新的描述符）键入你自己的标签，你可以增加自己的标签，但是所有在左边的框中的标准GenBank标签将也会被应用。

你可以为所有功能部件选择默认的颜色，或者你可以让BioEdit自动选择颜色（随后它们还是可以编辑）。如果BioEdit选择颜色，所有同类型的功能部件将会是同色的。如果你选择默认颜色，所有的功能部件将不管类型都是一种颜色。你也可以选择默认形状。所有功能部件的默认值是矩形。可用的形状选项是：矩形、椭圆形、菱形和箭头形。如果选择箭头，起点和末端位置将是决定功能部件方向的重要因素。

当BioEdit增加功能部件时，它搜索全部FEATURES数据，寻找上面对话框指定的指定标签。如果它找到一个（在正确的位置被格式化），将创造一个新的功能部件。标题将是功能部件的类型加上反映目录中功能部件当前数量的数值（如：“exon 1”、“intron 1”、“exon 2”等等）。描述将是所有文件中根据标签被描述的数据。例如，一个CDS功能部件将会有以下的名称和描述：

Name:
CDS 2
Description:
/label=b0014

/gene="dnaK"
/product="DnaK protein (heat shock protein 70)"
/note="o638; 100 pct identical to DNAK_ECOLI SW: P04475"
/codon_start=1
/transl_table=11
/translation="MGKIIGIDLGTTNSCVAIMDGTTPRVLENAEGDRTTPSIIAYTQ
DGETLVGQPAKRQAVTNPQNTLFAIKRLIGRRFQDEEVQRDVSIMPFKIIAADNGDAW
VEVKGQKMAPPQISAEVLKKMKKTAEDYLGEPVTEAVITVPAYFNDAQRQATKDAGRI
AGLEVKRIINEPTAAALAYGLDKGTGNRTIAVYDLGGGTFDISIIEIDEVDGEKTFEV
LATNGDTHLGGEDFDSRLINYLVEEFKKDQGIDLRNDPLAMQRLKEAAEKAKIELSSA
QQTDVNLPYITADATGPKHMNIKVTRAKLESLVEDLVNRSIEPLKVALQDAGLSVSDI
DDVILVGGQTRMPMVQKKVAEFFGKEPRKDVNPDEAVAIGAAVQGGVLTGDVKDVLLL
DVTPLSLGIETMGGVMTTLIAKNTTIPTKHSQVFSTAEDNQSAVTIHVLQGERKRAAD
NKSLGQFNLDGINPAPRGMPQIEVTFDIDADGILHVSAKDKNSGKEQKITIKASSGLN
EDEIQKMVRDAEANAEADRKFEELVQTRNQGDHLLHSTRKQVEEAGDKLPADDKTAIE
SALTALETALKGEDKAAIEAKMQELAQVSQKLMEIAQQQHAQQQTAGADASANNAKDD
DVVDAEFEEVKDKK"

BioEdit将指定为“互补”的功能部件的末端（低数）放在左边，起点在右边（高数）。因为功能部件可以通过一个“join”命令指定多个位点，所以可以通过每一个单独起点 / 末端位置设定来创造一个独立的功能部件。因此第一个功能部件将会有全部的描述区域。后来通过“join”命令创造的功能部件将会有一个 “join #<number> to <feature type> <number>”的描述（例如“join #4 to CDS 2”）

Annotating other sequences based upon an annotated template（根据注释模板注释其它序列）

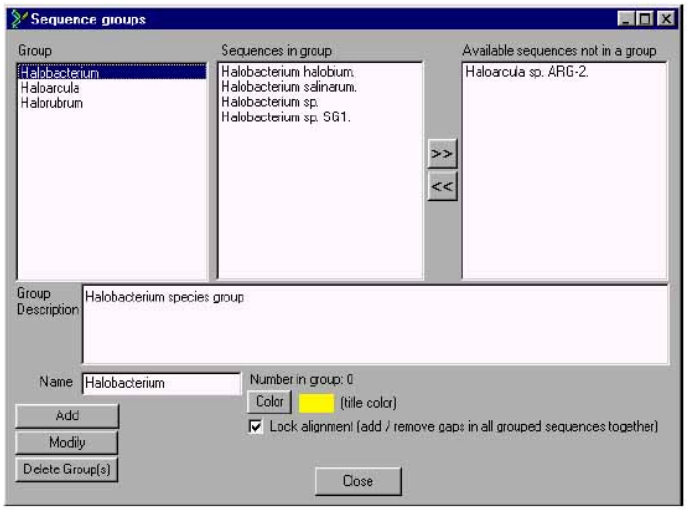
如果你处理一个同源序列的排列，将会是一个好机会，有利于你了解，你对其在序列中的功能感兴趣的功能部件，在生物学相关排列中序列中将会线状排列。所以，因为根据大多数，或者不是绝大多数像RNA或蛋白质螺旋、功能基序、内含子、外显子和CDS残基等等的功能部件排列序列，它可能只是必须根据感兴趣的功能部件注释一个序列。一旦一个序列适当的排列，将根据被注释序列中的排列位置功能部件，注释所有其他。这样，即使功能部件目前真实的位置和长度在序列中有不同，但是它们生物学排列中的相关位置将被线形垂直排列，于是注解正确的排列序列变得比单独注释每个序列更加容易。用这种方法创造的功能部件的真实位置将被BioEdit自动计算。

想要根据使用的另一个像模板的已注释序列，来注释序列，首先移动已注解序列到排列的顶端，选择所有你想要注释的序列的标题，选择“Sequence->Features->Annotate selected sequences using the first sequence as a template”

Grouping sequences into groups or families（序列分组或组成家族）

点击序列标题使其增亮成为分组指示颜色，序列可以通过分组来反映它们之间的相关性。同样，分组序列的排列可以为了前排列的同步排列调节，相互锁定，密切相关的序列可以根据新的数据或增加的序列进行调节。

想要编辑序列分组，选择“Sequence->Sequence groups (or families)”。将显示下列对话框：



你可以通过在“Name”中键入想要的组名，并按下“Add”按钮来进行分组。将会创造一个新的组，但是没有任何序列在其中。为了在一个组中增加序列，在“Group”目录中选择组的标题，从标题为“Available sequences not in a group”的右边目录中选择想要的序列标题，并按下“<<”按钮。每一组都有一个描述和颜色。排列窗口中的标题背景将根据组色来着色。你可以通过在“Group” 目录中点击序列使其增亮，并按下“Delete Group(s)”，来删除分组。

Verbal confirmation of sequences（序列的逐字检验）

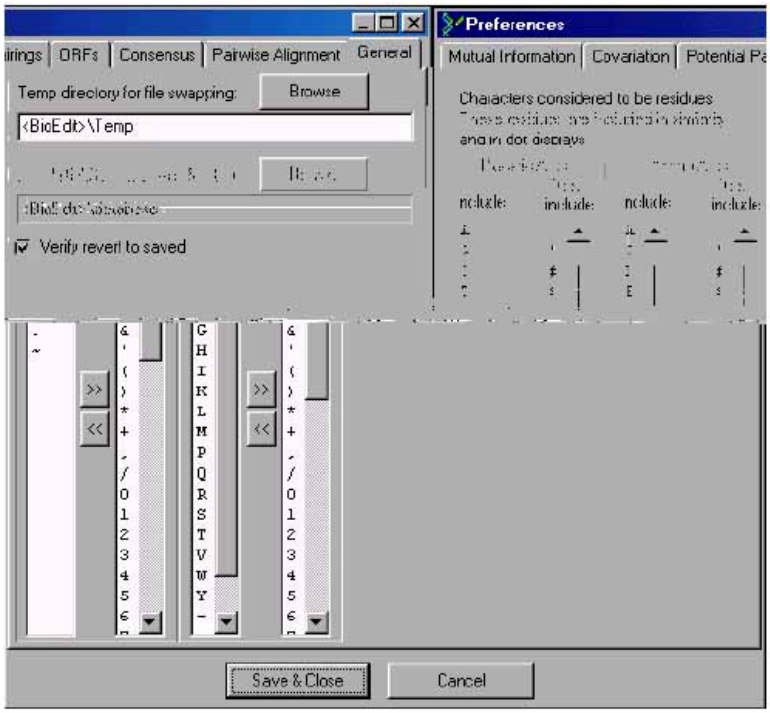
如果你在单一序列编辑器中，手动键入一个小序列，例如一个存储在文件中的用于合成的初级序列，它有时别人帮助你再次阅读序列，就像你一个碱基一个碱基的检验。如果没有人阅读你的序列，BioEdit将在单一序列编辑器中缓慢的重新阅读序列，进行阅读的碱基亮度增加（氨基酸序列也可以阅读）。

想要在单一序列编辑器中再次阅读序列，选择“Edit->Read Sequence Back (Press escape to cancel)”。

注意：这只能应用于单一序列编辑器。在单一序列编辑器中打开一个序列，双击它的标题，或点击它的标题增加其亮度，并选择“Sequence->Edit Sequence”

Valid residue characters vs non-residue characters（有效的残基字符 **VS** 非残基字符）

一个研究员可能希望使用序列中的不是常规定义的字符，它们是不明确的，或者只是有一个位置，但是不知道是一个残基还是一个缺口。因为这个原因，有一个选项可以明白的对字符进行定义，可以定义为适合进行类似性描影等计算和产生单位矩阵。对于氨基酸序列和核酸序列有不同的目录。想要看见或改变当前的关于“真实”残基的设置，选择“Options->Preferences-> General”。下列屏幕将会显示：



核酸字符的默认设置是AGUCT-~.，氨基酸的默认设置是ACDEFGHIKLMNPQRSTVWY-~. 。默认值中包含缺口字符，但是可以通过在左边框中选择它们，并按下“>>”按钮将其移出左边框，来删除它们。不管缺口字符是否像有效残基一样进行描影计算，“-”、“~”和“.”字符将被内部认为是缺口。同样，尽管缺口可能为了计算而被包括（它们可能被认为是一个错配，并且是由两个同源序列位点的不同而导致的，一个包含碱基，另一个没有［或者另一个是一个插入点］），缺口始终不能进行同一性描影，因为它们不是真正的物理实体。记住，所有的字符都是分别经过有效残基VS非残基字符处理过的，所以图国所有缺口字符经过检验（-、~和.）它们所有必须同时存在于氨基酸和核酸的有效残基目录中。

Locking a sequence to prevent accidental edits（锁定序列以防止意外的编辑）

可以锁定一个序列，防止在排列窗口中或在单一序列编辑窗口中，在序列中滑动残基或插入、键入字符。如果是已被分组的序列，手动排列一个序列（只通过滑动，而不是通过鼠标右键点击增加或删除缺口），这个组中的其他序列也同样闭锁，如果并只是如果序列组的组排列是锁定的。想要锁定序列，通过双击标题或增加标题亮度，在单一序列编辑器中打开序列，并选择“Sequence->Edit Sequence”。在编辑框中点击 Lock sequence 框，按下“Apply and Close”按钮

Anchoring a column to protect aligned regions（锚定排列栏以保护排列的残基）

这有时可以帮助锁定一个排列栏，而不用忧虑意外的移动序列位置，尽管可以关闭当前窗口。BioEdit因此允许锚定必要的栏以保护排列中的残基，使你不会弄乱它。想要锚定一个栏，按下“add / remove column anchors”按钮，用鼠标点击你想要锚定的栏。想要锚定一个全部区域，增加一个栏锚定你要保护的区域的每一边。如果你想要确定，没有排列可能在这个区域内，只是在每一边增加了一个锚定，不要选择所有序列标题，选择区域中的所有残基（通过鼠标拖动区域内的直条，增加区域的亮度），选择“Sequence->Gaps->Lock Gaps”

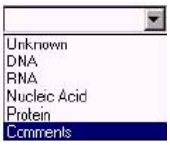
锁定区域内的所有缺口。区域可以被有效地锁定，直到锚被删除（或者按下“ignore anchors”按钮）。

想要删除锚，按下  按钮，点击一个现存锚以删除它。

如果你想要在一个锚定区域内调整排列,但是不想重新安排所有的锚，你可以按下“ignore column anchors”按钮  ,进行你的调整。在调整完成后，重新点击“ignore column anchors”按钮，使锚重新启动。

Comments（注解）

任何序列都可以产生一个注解，只在排列窗口中产生一个空间，但是不参与描影或计算，也不像一个真实序列一样被计算。而且，一个注解被内部处理，就像另一个序列在一些情况下被简单忽略和在主排列文件窗口中显示斜体。任何有效的ASCII字符可以在注解中被键入。想要创造一个注解简单的创造一个新的序列（“Sequence->New Sequence”），在单一序列编辑器中，将它的“Type”变为“Comment”：

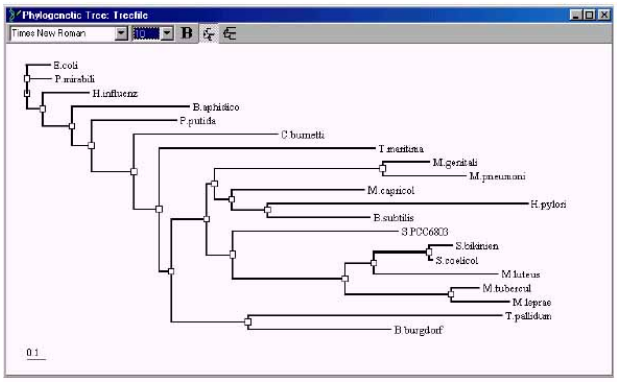


Phylogenetic Tree Viewer（系统树图）

BioEdit5.0.6版本包含一个非常根本的系统树图，能够打开和观看phylip格式树的文件。同样，多树可以直接连接排列文件（最多50个树可以连接到一个排列），系统树信息连同当前的节点和分支模式，一起保存为BioEdit文件模式。树图同样允许交换节点（在不改变发展史的情况下），保存、打印、编辑标签、观看带有或不带有距离信息的树图。然而，只有矩形的进化树图当前是可用的。在二者选一的格式选项中，我推荐使用TreeView，它能够用于http:://taxonomy.zoology.gla.ac.uk/rod/rod.html的Roderic Page。可以随BioEdit安装TreeView1.5.2版本，在BioEdit安装文件夹中可以找到TreeView.zip文件。

想要在BioEdit中打开一个phylip树，只要在程序的任何位置选择“File->Open”。BioEdit可以自动的判断出它是一个树图文件，并适当的打开它。

下面是一个树图显示在BioEdit中的样子：



你可以在任何节点点击鼠标，节点有一个小正方形（□）在它的连接处，可以交换树周围的节点。当保存树的全部分支模式和距离时，这将颠倒所有下游节点和枝叶的位置（在每一个分支末端的最后标签）。

想要编辑标签，用鼠标点击屏幕上的标志。标签将会进入编辑模式，将会变为完全被选择。你可以键入你想要用来替代的标签。做完后，可以用鼠标选择树图窗口中不同的放置位置，或者按下“enter”键。想要取消编辑，按下“Esc”键。

注意：树图有时会在一个节点连接两个以上的分支。我已经注意到使用Phylip程序编写的树图有时有三个分支直接从第一个节点连出。BioEdit树图允许一个以上的分支从每一个节点连出（最多10个，事实上，也比较安全），但是当一个来源于文件的树直接在BioEdit树图视窗中打开时，如果树的节点有多于两个分支，它将通过在每一点创造一个距离为0的额外节点，将有多于一个分支的节点自动修改为一个完全的二进制树。树的布局不会改变。在打开的树之上，树被重复，每一个分支移动到自己节点的两个节点以后，新的节点（来自于亲代，距离为0），直到没有有两个分支的节点存在。然而，当将一个树导入BioEdit排列中时，将不会执行这个转变，树将会直接导入。它在相同的视图框中观看，但是原来的节点组织被保留。

你可以从“File”菜单中选择“File->Save”来保存树图。当前版本的BioEdit只限于打开和保存Phylip格式的树。在Phylip格式中，上述的树像这样：

((P.mirabili: 0.13368,((B.aphidico: 0.6262,(((T.maritima:
1.14167,((((M.genitali: 0.24742,M.pneumoni: 0.43983): 0.88981,(M.capricol:
0.70024,(H.pylori: 1.37587,B.subtilis: 0.53651): 0.19415): 0.0886):
0.04525,(S.PCC6803: 0.87437,((M.tubercul: 0.14643,M.leprae: 0.30498):
0.56324,(M.luteus: 0.65897,(S.bikinien: 0.1209,S.coelicol: 0.01772): 0.29437):
0.14817): 0.59556): 0.14169): 0.18393,(T.pallidum: 1.3449,B.burgdorf:
0.75431): 0.40702): 0.06668): 0.13184,C.burnetti: 0.76309): 0.22955,P.putida:
0.45219): 0.10167): 0.15512,H.influenz: 0.24691): 0.08603): 0,E.coli:
0.12297);

BioEdit视图视窗只同时支持一个树，而且如果打开一个有多个树的树图文件，只有第一个树会被载入。然而，当在一个排列文件中导入树时，所有的树（最多50个）都可以被载入排列（像单独的树一样）。

树图视窗规定树的格式为视图窗的当前尺寸，现在不支持多页、全屏显示或手动指定尺寸，所以它只适合于小的树。同样，打印也是原始的，只有打印页的范围。而且，不能复制到剪切板中。为了产生树的图像，我推荐TreeView，它能像Windows的图元文件一样，将树图复制到剪切板中。

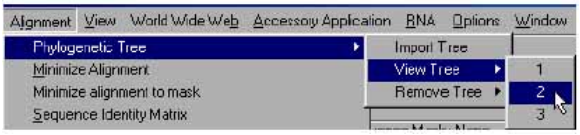
Importing Phylogenetic Trees into an alignment（将系统树导入排列）

它可以有时方便的使用系统树，来显示排列中序列的相关性。因为这个原因，BioEdit5.0.6版本允许你在排列中导入一个或更多的系统树（只要它们是Phylip格式的），并在一个BioEdit格式的排列文件中保存这些树。你可以在一个文件中最多保存50个树。通常，只有一个树是希望的，但是可能有一系列等同的树产生与过于简单的方法，或者可能你想要通过树来显示一个排列小组中序列的联系。

想要在BioEdit排列中导入一个树，打开排列（File->Open），选择“Alignment->Phylogenetic

Tree->Import Tree”。菜单将变成这样：

你可以提示指定导入树图文件。想要观看导入的树图，选择“Alignment->Phylogenetic Tree->View Tree-> (tree number)”。例如，如果你有三个树图和一个排列联合，菜单就会像这样：



你可以将你的文件保存为BioEdit格式，而且你的相关的树图将会随文件而保存。记住，如果文件没有保存，“Revert to Saved”（恢复保存）选项将同样删除任何没有保存在文件中的树图。

你可以通过“Alignment->Phylogenetic Tree->Remove Tree”选项，删除树图。

你也可以在树图视窗中打开一个树图，选择将其连接到一个打开的排列文件中，如果很容易就看见树图，可以确定它是正确的。想要做到这一点，从程序的任何位置上选择“File->Open”命令，打开树图，确定你已经打开了一个排列文件，从树图视窗选择“File->Associate Tree With Alignment”。你将见到一个对话框，其中有当前所有打开的排列，从中你可以选择合适的排列。

File formats（文件格式）

File formats read and written by BioEdit（BioEdit读写的文件格式）

BioEdit 5.0.0版本可以读写以下格式：

- BioEdit
- Genbank
- Fasta
- NBRF/PIR
- Phylip 3.2 / 2
- Phylip 4
- 另外，BioEdit 4.7.0版本和以上，可以读ABI model 377 autosequencer文件。序列是提取的，痕迹显示在屏幕上，可以打印颜色。BioEdit 4.7.7版本和以上允许编辑可编辑序列。当前版本同样可以读SCF trace文件（版本2和3）、ABI 373和3700文件。
- BioEdit 4.7.7版本和以上，同样可以读ClustalW和GCG-格式文件，但是不能写。

这些格式以外，提供一个外部导入 / 输出筛选（Don Gilbert’s ReadSeq），允许导入和输出以下格式：

- IG/Stanford
- EMBL
- GCG (single sequence only)
- DNAStrider

- Fitch
- Zuker (import only)
- Olsen (import only)
- Plain or raw (single sequence only)
- PIR/CODATA
- MSF (multiple sequence format)
- ASN.1 (NCBI)
- PAUP/NEXUS

ReadSeq公用程序文件可以在BioEdit安装目录下/apps文件夹的ReadSeq.txt文件中找到。当打开序列时，在BioEdit中自动使用这个公用程序。如果打开的文件不是BioEdit能够读的文件格式，你可以用ReadSeq打开它。如果ReadSeq可以打开它，它就可以像GenBank格式文件一样导入BioEdit，而且它将被像文本文件一样打开。想要将一个文件保存为这些格式中的一种，选择“File->Export->Sequence alignment from an open document”。

BioEdit Project File Format（BioEdit的Project文件格式）

BioEdit为了快速的打开和保存大的排列文件（大于20Mb的文件尺寸－－最大达到100Mb或更大），提供了一个专门的二进制排列。读和计算大的排列格式（如GenBank）的未处理过的文本时，速度变得非常慢，因为它们不能指示程序有多少序列在文件中，以及有多大。

每一个序列结构包括标题、序列、序列类型和所有同样的GenBank区域包括BioEdit中的GenBank文件。另外，在BioEdit 5.0.0版本或以上，图时序列注释、序列分组信息、一致序列信息、序列锁定状况和位置标旗将被保存。

GenBank Format（GenBank格式）

BioEdit编写的GenBank文件具有以下最小的格式：

```
LOCUS              Escherichi                 119 amino acids
DEFINITION         Escherichi                   119 amino acids
ORIGIN
1     MVKLA FPREL RLLTP SQFTF VFQQP QRAGT PQITI LGRLN SLGHP RIGLT
51    VAKKN VRRAH ERNRI KRLTR ESFRL RQHEL PAMDF VVVAK KGVAD LDNRA
101 LSEAL EKLWR RHCRL ARGS
//
LOCUS              Proteus_mi             119 amino acids
DEFINITION         Proteus_mi             119 amino acids
ORIGIN
1     MVKLA FPREL RLLTP KHFNF VFQQP QRASS PEVTI LGRQN ELGHP RIGLT
51    IAKKN VKRAH ERNRI KRLAR EYFRL HQHQL PAMDF VVLVR KGVAE LDNHQ
101 LTEVL GKLWR RHCRL AQKS
//
etc...
```

LOCUS、DEFINITION和ORIGIN等关键字在检定GenBank文件时被查找。

GenBank文件同样可以包含附加信息。以下区域可以包括在任何GenBank序列入口，而且在打开GenBank文件时被查找。

LOCUS:序列的轨迹（在基因组中通常的位置）。这个区域通常是一个单一链，包含有Locus的名称、序列长度和经常提交的数据。BioEdit以前的版本将LOCUS当作序列的标题。
DEFINITION:序列的描述，经常是在线的。在没有BioEdit特定的“TITLE”区域时，DEFINITION区域被用于默认值的标题
TITLE:这是BioEdit中的特殊区域，能够在读GenBank格式文件时被其他程序所忽略。这个区域允许你保存序列标题，这不同于LOCUS 或 DEFINITION 区域入口。这可以使用户定义的标题用于通过Entrez下载的序列，而不需要改变序列标题的原始数据。TITLE区域不是标准GenBank文件的一部分，只适用于BioEdit。如果这个区域在试图用另一个程序打开序列时，有问题，像打开文本一样打开序列，在其他程序使用文件以前删除这个区域。
ACCESSION:序列的GenBank编号
PID or NID:蛋白质或核酸ID.
DBSOURCE:来自获得的序列的数据
KEYWORDS（关键词）
SOURCE:序列的来源（通常来自于获得的有机体）。这个区域通常包含ORGANISM子域，一个给出有机体描述的区域（通常是分类学的分类）。
REFERENCES:与序列提交相关联的参数
COMMENT:多种信息。一个关于与序列相关联的用户定义信息的方便的位置。
FEATURES:序列功能部件包括翻译、促进子、更多来源信息等等。
ORIGIN:标记真实序列数据起点。用 // 指出序列的末端。

LOCUS、DEFINITION和ORIGIN区域是GenBank文件想要被BioEdit承认所必需的。其他区域时可选择的。当GenBank文件保存时，如果LOCUS或DEFINITION区域是空的，它们将会用序列的标题和长度来填写。因此，LOCUS 和DEFINITION可以是相同的。其他空的区域不会写进序列的入口。

打开文件时，每一个区域像一个单一文本框一样被阅读。子域不能正式的被确认，所以可能存在于原始文件中的任何“不寻常”的格式（例如非标准的间隔）将会出现。然而，保存一个GenBank文件时，像在NCBI Entrez GenBank或GenPep中一样，特殊的子域名将被寻找和间隔。以下子域将被寻找：
REFERENCES field(s):（参数区域）
reference number (format = REFERENCE <num> <description>)
AUTHORS
TITLE
JOURNAL
MEDLINE
REMARK
STRAIN

FEATURES field:（特征区域）
BioEdit以前的版本只是寻找GenBank中少量的FEATURES标签。5.0.0版本和以上寻找以下所有67种标签：

3'clip
3'UTR
5'clip
5'UTR
-10_signal
-35_signal
allele
attenuator
CDS
C_region
CAAT_signal
conflict
D-loop
D_segment
enhancer
exon
Gene
iDNA
intron
J_segment
LTR
mat_peptide
misc_binding
misc_difference
misc_feature
misc_recomb
misc_RNA
misc_signal
misc_structure
modified_base
mRNA
mutation
N_region
old_sequence
polyA_signal
polyA_site
precursor_RNA
prim_transcript
primer_bind
promoter
Protein
protein_bind
RBS
Region

repeat_region
repeat_unit
rep_origin
rRNA
S_region
satellite
scRNA
SecStr
sig_peptide
Site
snRNA
source
stem_loop
STS
TATA_signal
terminator
transit_peptide
tRNA
unsure
V_region
V_segment
variation

任何在区域中的数据将会保存，然而，在REFERENCE和FEATURES区域中，保存在副标题以下的数据将不会像希望的一样被间隔。
想要编辑详细区域，请看“Editing in an Edit Box”

Fasta Format（Fasta格式）
在BioEdit中编写的Fasta/Pearson文件有以下格式：

>Escherichi 119 amino acids
MVKLAFPRELRLLTPSQFTFVFQQPQRAGTPQITILGRLNSLGHPRIGLT
VAKKNVRRAHERNRIKRLTRESFRLRQHELPAMDFVVVAKKGVADLDNRA
LSEALEKLWRRHCRLARGS
>Proteus_mi 119 amino acids
MVKLAFPRELRLLTPKHFNFVFQQPQRASSPEVTILGRQNELGHPRIGLT
IAKKNVKRAHERNRIKRLAREYFRLHQHQLPAMDFVVLVRKGVAELDNHQ
LTEVLGKLWRRHCRLAQKS
etc...

“>”字符后有一连串标题内容，和一个完整的字符链。“>”字符被用来确认Fasta文件。

NBRF/PIR format（NBRF/PIR格式）
在BioEdit中编写的NBRF/PIR文件有以下格式：

>P1;Escherichi
Escherichi 119 amino acids
MVKLAFPREL RLLTPSQFTF VFQQPQRAGT PQITILGRLN SLGHPRIGLT
VAKKNVRRAH ERNRIKRLTR ESFRLRQHEL PAMDFVVVAK KGVADLDNRA
LSEALEKLWR RHCRLARGS*
>P1;Proteus_mi
Proteus_mi 119 amino acids
MVKLAFPREL RLLTPKHFNF VFQQPQRASS PEVTILGRQN ELGHPRIGLT
IAKKNVKRAH ERNRIKRLAR EYFRLHQHQL PAMDFVVLVR KGVAELDNHQ
LTEVLGKLWR RHCRLAQKS*
etc ...

“>P1;”表示蛋白质序列；“>DL;”表示核酸序列。
序列写在10个字符 / 框的框中。序列的末端用星号表示。
NBRF 文件用 “>P1;”或“>DL;”表示，其后紧跟标题。

Phylip 3.2/2 format（Phylip 3.2/2格式）
在BioEdit中编写的Phylip 3.2 / Phylip 2文件有以下格式：

3 136 I
Escherichi MVKLAFPREL RLLTPSQFTF VFQQPQRAGT PQITILGRLN SLGHPRIGLT
VAKKNVRRAH ERNRIKRLTR ESFRLRQHEL PAMDFVVVAK KGVADLDNRA
LSEALEKLWR RHCRLARGS- ---------- ------
Proteus_mi MVKLAFPREL RLLTPKHFNF VFQQPQRASS PEVTILGRQN ELGHPRIGLT
IAKKNVKRAH ERNRIKRLAR EYFRLHQHQL PAMDFVVLVR KGVAELDNHQ
LTEVLGKLWR RHCRLAQKS- ---------- ------
Haemophilu MLKVVKVYLH NHNSQFLVVK LNFSRELRLL TPIQFKNVFE QPFRASTPEI
TILARKNNLE HPRLGLTVAK KHLKRAHERN RIKRLVRESF RLSQHRLPAY
DFVFVAKNGI GKLDNNTFAQ ILEKLWQRHI RLAQKS

所有在Phylip格式的序列有同样的长度。文件的第一行表示序列的数量和每一个序列的长度。“I”在这儿特指它是Phylip 3.2格式，而不是Phylip 4。每一个序列都写在标题之后的10个字符 / 框的框中。标题有10个字符长，序列和标题间隔3个间隔。

Phylip 4 format（Phylip 4格式）
在BioEdit中编写的Phylip 4文件有以下格式：

3 136
Escherichi MVKLAFPREL RLLTPSQFTF VFQQPQRAGT PQITILGRLN SLGHPRIGLT
Proteus_mi MVKLAFPREL RLLTPKHFNF VFQQPQRASS PEVTILGRQN ELGHPRIGLT
Haemophilu MLKVVKVYLH NHNSQFLVVK LNFSRELRLL TPIQFKNVFE QPFRASTPEI
VAKKNVRRAH ERNRIKRLTR ESFRLRQHEL PAMDFVVVAK KGVADLDNRA
IAKKNVKRAH ERNRIKRLAR EYFRLHQHQL PAMDFVVLVR KGVAELDNHQ
TILARKNNLE HPRLGLTVAK KHLKRAHERN RIKRLVRESF RLSQHRLPAY

LSEALEKLWR RHCRLARGS- ---------- ------
LTEVLGKLWR RHCRLAQKS- ---------- ------
DFVFVAKNGI GKLDNNTFAQ ILEKLWQRHI RLAQKS

所有在Phylip格式的序列有同样的长度，并且交叉存取。文件的第一行表示序列的数量和每一个序列的长度。每一个序列都写在标题之后的10个字符／框的框中，并且在每一个序列中交叉存取50个残基，写入每一个框中。标题在第一个框的前面。标题有10个字符长，序列和标题间隔3个间隔。所有的框都向右间隔13个间隔。

ABI Autosequencer Trace Files（ABI自动痕迹文件）
BioEdit 4.7.0版本和以上可以杜ABI模式377痕迹文件。但是我不熟悉旧的ABI文件或.SCF文件，所以当前不支持这些文件。解释ABI文件需要的多数信息可以从ABIView网页中（作者是David H. Klatte）得到。打印输出的标题信息使用一个十六进制的编辑器表示，来源于David Klatte的信息被当作是起点。

想要在BioEdit中打开一个ABI文件，只需要像打开其他文件一样。像排列文件和质粒文件一样，文件的格式将会被自动检测（如果文件的扩展名是.abi，你可以使用*.abi过滤器）。打开一个ABI文件时，（可编辑的）序列将会被选取到一个新的序列／排列文档，而且痕迹将会显示在一个单独的窗口。一个ABI文件包含有一个副本序列，允许编辑序列和保存原始的碱基顺序。非编辑序列在痕迹窗口中显示，并位于第一次打开的痕迹之上。下面的例子显示了随BioEdit安装而存在的sample.abi文件，用窗口平铺方式打开：

可以用鼠标选择痕迹的任何部分，可以从痕迹窗口复制部分序列。全部序列可以用原始文本形式或Fasta格式复制或输出。

进行窗口大小调整时，垂直缩放比例可以成比例的调整。全部痕迹可以通过Zoom菜单放大，水平缩放比例可以单独的从Horizontal Scale菜单调整。
想要编辑序列，你必须首先通过选择“View->Editable sequence”，将序列转换为可编辑序列。保存编辑过的序列将不会改变非编辑序列，选择“View->Non-editable sequence”可以在任何时候显示非编辑序列。非编辑序列显示通常是第一次打开的ABI文件时的默认项。然而，可编辑序列可以提取到一个序列文档之中。选择“Edit->Revert edited to non-editable sequence”，可编辑序列将会在任何时候恢复成非编辑序列。

选择“File->info”，你可以显示一些相关的文件标题信息。
选择“View->Reverse complement”，可以完全回复痕迹和序列。

输出痕迹看起来和输出ABI Prism一样。多数情况下，只要从“File”菜单中选择“Print”，就可以根据你想要的缩放比例进行格式化的打印。然而水平和垂直的缩放比例可以通过在“File”菜单下的“Print Scaling”菜单更改打印。可以选择一系列的边框形式，或者选择“other”后可以详细制定任何准确的缩放比例。

Saving sequence annotation information（保存序列注释信息）
BioEdit将会保存在标准GenBank格式文件中的信息。
以下区域将会包括在任何GenBank或BioEdit文件中：
LOCUS
DEFINITION
TITLE (BioEdit 独有－－不是标准的GenBank格式)
ACCESSION
PID or NID
DBSOURCE
KEYWORDS
SOURCE
REFERENCES
COMMENT
FEATURES

除了被保存在“FEATUERS”区域的文本信息之外，序列的GenBank区域可以使用来自GenBank FEATURES区域的标准标签，独立进行手动或自动的图示注释。然而，图示注释信息将只会被保存在BioEdit文件格式。
以上区域的描述，详见“GenBank file format”。
图示序列注释的描述，详见“Graphical Feature Annotations”。

注意：其他序列信息、标题和长度将植被保存在GenBank和BioEdit格式。它可以轻松的保存绝大多数的GenBank格式的序列文件，并且通过特殊转换才能使用其他格式。

Reading Files saved with BioEdit with a Macintosh program
（读出在**Mac**程序中的**BioEdit**保存的文件）
Mac计算机和PC计算机使用的回车符不同。如果你需要使用一个文件（如SeqApp或DNA Strider），而这个文件是用Mac程序中的BioEdit编写的，可能需要首先在一个文本处理器（像Microsoft Word或WordPerfect）中打开这个文件，然后再次保存它以产生正确的回车符。BioEdit将来可以正确读出在Mac或UNIX上编辑文件。

Toggling between nucleotide and protein views（在核苷酸视窗和蛋白质视窗之间切换）
想要控制翻译处理缺口的方式，从“View”菜单中点击“Toggle Translation Control”选项，使排列文档的控制条中可以看见“Force contiguous codons”和“Ignore gaps that split codons”。

当处理编码蛋白质的核苷酸序列时，BioEdit允许在核苷酸序列和蛋白质序列之间来回切换，每一个视窗都反映两个视窗中缺口的插入或删除。在从蛋白质视窗中切换回来时，核苷酸信息将被保存。

想要在核苷酸和蛋白质视窗中来回转换，首先序列5'末端修剪成起始密码子，确定编码区在Frame 1中。选择切换的序列，并从“Sequence”菜单中选择“Toggle Translation”，或者从“Alignment”菜单中选择“Toggle Translation”（这将导致所有序列自动选择和切换）。序列可以在任一视窗被排列。另外，如果蛋白质视窗是被Clustal排列的，以下的核苷酸序列将更新为合适的缺口。

注意1：保存一个排列时，如果序列被切换在蛋白质视窗中，是核苷酸序列而不是蛋白质序列被保存。

注意2：这个功能只有在起始序列是核苷酸序列时才有用。如果起始序列是蛋白质序列，在关闭后这个功能将不能使用，因为蛋白质的编码区将不能通过单独检测氨基酸序列而知道。

注意3：处理核苷酸序列中的缺口又三种有用的方式：中心分裂密码子、单独出现或成对出现。蛋白质序列中的缺口将对应编码核苷酸序列中的三种缺口。然而，如果在核苷酸序列中有一个或两个缺口，或者缺口直接位于密码子中（在Frame 1 中），就会有一个问题。BioEdit依赖选择选项，使用三种方法中的一种来处理这个问题：

想要改变翻译切换选项，必须点击“View”菜单下的“Translation Toggle Control”选项。点击这个菜单项后，将会有两个点击盒出现在排列窗口的顶端面板的右边。有用的选项是：
1.强迫所有缺口三个一组出现，并且只出现在密码子之间（不出现在密码子之中）。在这种方式，如果缺口是被引入一个密码子内部，核苷酸下游在全部密码子产生蛋白质过程中左移。如果这导致一个或两个缺口（或如果一个缺口是手动放置在两个密码子之间），缺口将延伸到三个位置，产生一个氨基酸类型的缺口。如果它们不组成一个三联密码子，这将导致缺口位置自动改变。这使通过蛋白质翻译简单排列核苷酸序列变得容易。
**在点击“Force contiguous codons”点击盒时，这个模式被激活。

2.忽略分裂密码子的缺口。这个模式胜过试图“修复”序列，任何出现在密码子中的缺口或不能产生一个氨基酸缺口的缺口，在蛋白质的翻译中被简单的忽略。然而，它们还是被保留在核苷酸序列中，在从蛋白质切换回核苷酸时，它们还将在那儿。
**在点击“Ignore gaps that split codons”点击盒时，这个模式被激活。
Mode 1 and 2 不能同时激活。

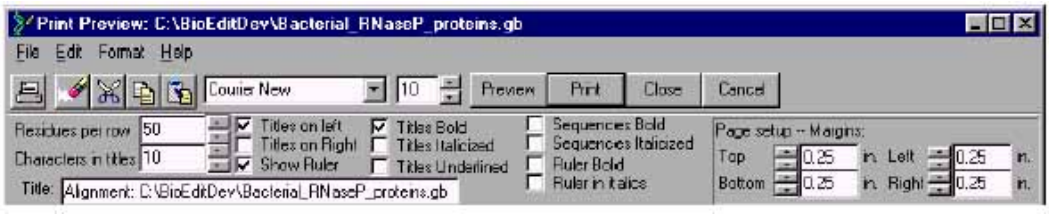
3.以上两种模式都不是。当以上两种点击盒都没有被点击时，这个模式被激活。这个模式不会尝试修复序列，但是缺口在翻译中不被忽略。任何不是连续三个的缺口将会导致移码。一个出现在密码子中的缺口（在Frame 1中）将会翻译成“X”，而不是一个有效的氨基酸。

**当以上两种点击盒都没有被点击时，这个模式被激活

Printing（打印）
想要打印排列，从“File”菜单选择“Print Alignment as Text”。将会出现一个预览界面。

预览和一个大的文本编辑器合为一体，你可以在屏幕上编辑排列。如果标题是指定的，它将打印在排列的起点。按下预览按钮使选定格式的排列重新进入预览窗口。如果做完了在屏幕上的所有编辑，确定设置了基础格式（每行的残基、标题的字符等等），因为再次按下预览按钮会将键入改写入预览窗口。

预览界面是清楚易懂的：



注意：预览窗口将显示最好的预览形式，将让你知道是否超过每行残基数量的上限。如果你超过了上限，序列的每一行将会卷曲。如果你试图打印超过右边界的残基时，这就会发生。如果发生这种情况，减少每行的残基数、字体大小或者右边界，或者横向打印。

Exporting as raw text（未处理文本输出）

BioEdit提供一个将排列转换为合适间隔的未处理文本的功能。想要用非处理文本文件输出排列，从“File”菜单选择“Save As ...”选项，并且在保存对话框中从“Save as type”选项中选择“Text Files”。会出现一个对话框，询问要保存的每行残基的数量和每个标题的字符：

Exporting as Rich Text（**Rich Text**格式输出）

一个类似性和同一性描影的排列可以用Rich Text格式输出，保留突出显示的残基，只要文件是显示在支持Rich Text格式文件突出显示的Word 97或更新的或其他文字处理器。排列的描影是与当前描影图形视窗的设置一致的，而且这个格式可以直接从描影图形视窗输出。想要直接用Rich Text格式输出一个排列，选择“File->Export->Rich text”，输出当前的描影视窗设置。

Shaded graphic view of alignment（排列的描影图形视窗）

想要用同一性和类似性描影来显示一个排列，在一个已打开的文档中，选择你想要包括的序列标题，从“File”菜单选择“Graphic View”。这和打印预览是相同的，但是允许被同一性和类似性描影的残基存在序列中，而且允许你显示排列的任何子集。以下是通常被建议的选项：

- 描影残基的变量阈值百分比（同一性和类似性共用一个阈值）
- 显示或隐藏标尺
- 显示或隐藏在序列数据左边或右边的标题
- 显示或隐藏在左边或右边的位置值
- 每行残基的变量值（20到2000）
- 每个标题的字符变量数（5到30）
- 标题可以被加粗、斜体和／或加下划线
- 序列可以被斜体和／或加粗
- 可以使用矩阵积分选择
- 任何字体都可以使用，而且可以正确的间隔。然而一些字体在这个视窗中看起来有些奇怪，

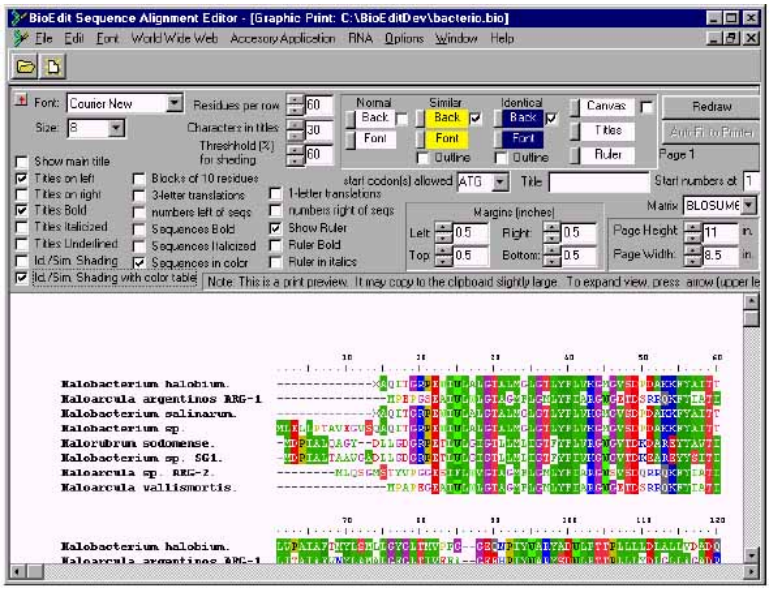
排版字体可以很好的工作。想要更多的控制字体，从“Font”菜单选择“Character Font”。
•页面的背景颜色、同一性和类似性的颜色，以及不同的、相同的和相似的残基字符的颜色。
•可以增加标题，把它放在每一页的开始。在这个版本中，如果排列多于一页，会出现问题，所以最好让标题区域空白。
• 排列可以绘成标准色彩表的颜色（允许打印彩色排列）
• 根据同一性和类似性描影（通过“Threshold (%) for shading”定义阈值）可以使用用户使用控制面板定义的颜色或者是当前排列色彩表中的颜色。
• 序列行可以组成10一组（像Phylip文件）或者像联系、完整的行一样打印。
• 翻译可以被显示在3个字母一组或者1个字母一组的核酸序列下面。
• 被描影的序列可以用Rich Text文档输出，保存突出显示的色彩（Word 97或以上版本，或者相同版本，是显示Rich Text中突出显示必需的）。

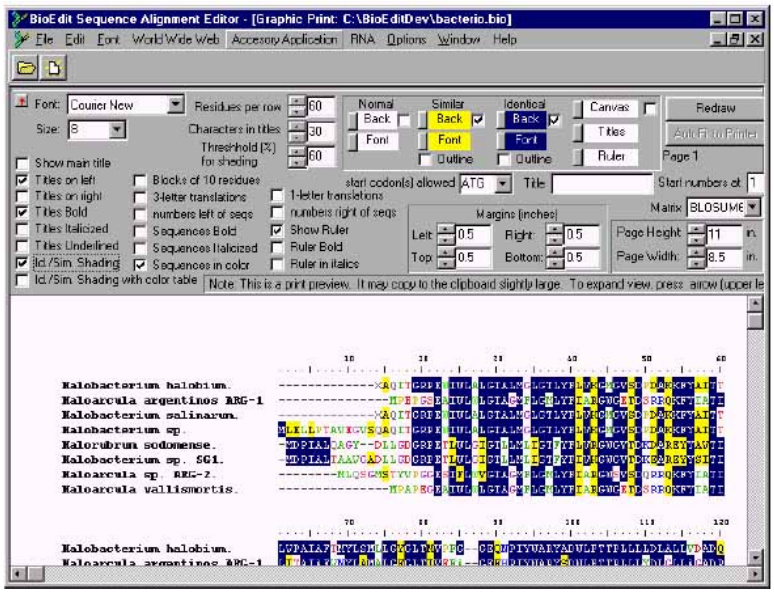
在当前视窗进行一定的更改时，必须按下“Redraw”按钮才能使屏幕上的更改生效。许多更改是自动更新的。想要改变颜色，可以双击色彩盒的标签，或者按下它左边的按钮。
想要将页面复制到剪切板中，并粘贴到其他应用程序中，从“Edit”菜单中选择“Copy”。页面可以当作位图或者Enhanced Windows Metafile(EMF)一样被复制。当成图元文件复制时，允许直接粘贴到应用程序中，如利用PowerPoint制作包含有被描影排列的幻灯片，或者粘贴到一个排版程序中制作成出版图表。一个图元文件提供向量图，能够利用输出设备的全部分辨率。现在，只有完整的一页可以进行复制。在以后的版本中，如果大家需要，可能会提供注释和选择功能。

长的排列可以使用多页，然而，在BioEdit现在的版本中有很严重的问题。现在，在图形视图窗口，页面的垂直定义值是用户用英寸指定的页面高度。当图形视窗垂直滚动时，当前的页面显示在表格的右上方。到一页的末端时，下一页显示在同一个位置。不能连续逐页显示。同样，如果一个排列超过一页，需要一些图像编辑将它编辑成图像文件。页面的高度最大是100英寸，但是这将占用巨大的内存，这是不被推荐的（每一页都是位图图像）。想要制造一个超过一页的图像，你可以增加页面高度，复制图像到剪切板中。

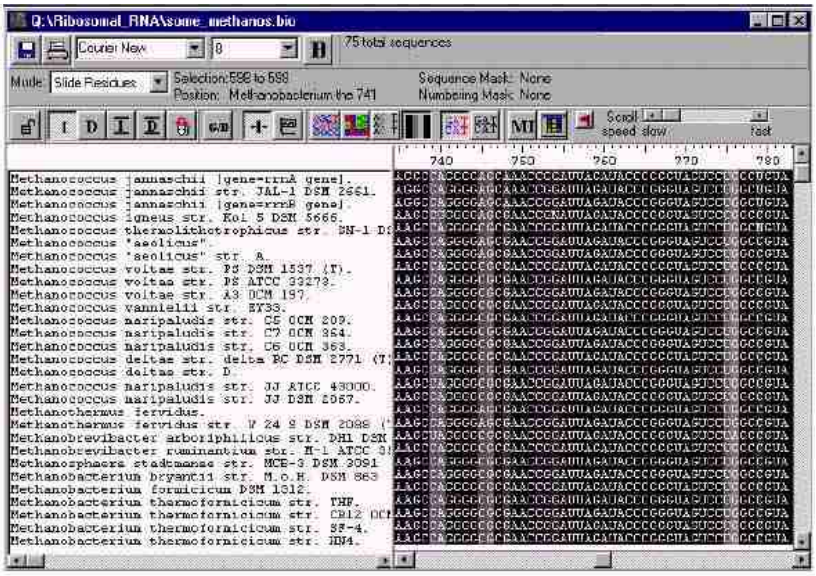
BioEdit以前的版本只能决定页面的高度。当前的版本增加了详细的页面设置（打印设置、页边距和页面大小设置），还可以修剪超过右边界的图形。这将适合（也许不是总合适）在当前设置下的打印机。同样，图形视窗中的页边距将精确反映每一页的打印预览情况。

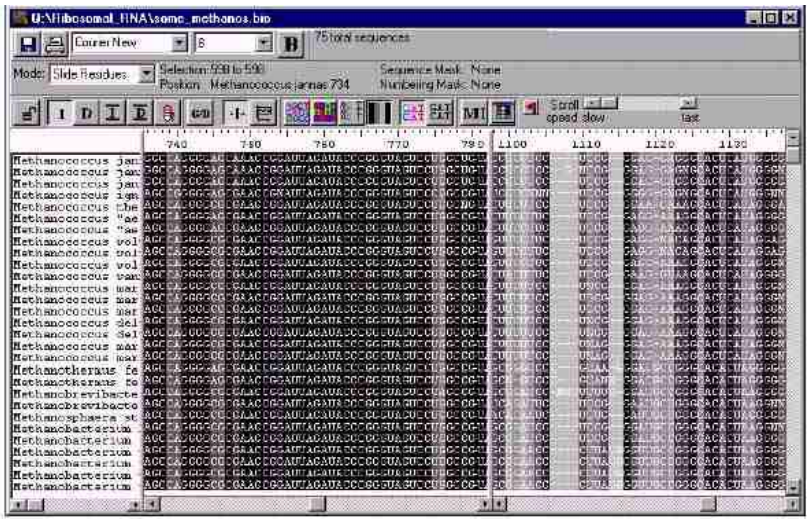
以下是来自于具有代表性的适盐性太古代的细菌视紫红素蛋白序列的两个描影视窗：





Information-based shading in the alignment window（排列窗口中的基本信息描影）
下面的视窗显示一个来源于太古代产甲烷菌的75 16S rRNA序列的基本信息描影。显示的这个区域是从保存的区域中通过平均信息量 / 基本信息搜索精选出来的。
这个视窗和下方的分裂窗口视窗比较，显示出相同的排列中的这个区域和没有很好保存的区域的区别。





Restriction Maps（限制性内切酶图）

BioEdit提供两种方法产生核苷酸序列的限制性内切酶图。一种内在的限制性内切酶图功能允许产生序列最多为65,536个核苷酸的限制性内切酶图。实际上，只能检测大约35Kb，而且在速度慢的计算机上会要消耗很长的时间。你也可以通过万维网直接链接到WebCutter限制性内切酶图上。

WebCutter：点亮你想要图谱的序列标题，从“World Wide Web”菜单中选择“Auto-fed WebCutter Restriction Mapping”。

BioEdit: 点亮你想要图谱的序列标题，从“Sequence”菜单选择“Restriction Map”。以下选项将会显示在一个界面窗口：

-- Display Map（显示图谱）：显示或省略序列的全图谱，互补链显示每个酶的酶切位点。默认值：yes

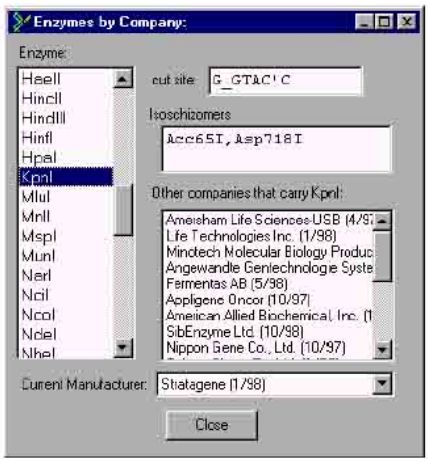
-- Alphabetical by name（按照字母顺序排列名称）：显示关于所有内切酶、它们的识别序列、切割频率和所有位置（5’末端开始是1）的列表。默认值：yes

-- Numeric by position（位置数）：关于酶切位点的列表。默认值：no

-- List of unique sites（唯一位点列表）：在全部序列中只有一个酶切位点的内切酶列表。

默认值：no
-- Enzymes that cut five or fewer times（切割5次或更少的酶）。默认值：yes
-- Summary table of frequencies（频率汇总表）：关于所有正确选择的内切酶和它们切割序列的次数。默认值：no
-- Enzymes that do not cut（不能切割的内切酶）。默认值：yes
-- 4-base cutters（4-碱基内切酶）：You may choose to omit enzymes that cut at a 4-base recognition sequence. 你可以选择省略能够切割4-碱基识别顺序的酶。想要包括这些酶，必须点击这个选项。默认值：no（不包括4-碱基内切酶）
-- 5-base cutters（5-碱基内切酶）：与4-base cutters相同。
-- Enzymes with degenerate recognition sequences（非严格识别序列的酶）：许多内切酶识别序列是不严格的。有时你可以排除它们。默认值：包括
-- Large recognition sites（大的识别位点）：通常用于克隆，只有共同的6-碱基识别酶被使用。如果你不想图谱被额外的信息弄乱，不要点击这个选项（就像4-碱基和5-碱基内切酶）
-- All Isoschizomers（同裂酶）：The enzyme list file used is the GCG-format file available from ReBase.酶的列表文件通常是可用于ReBase的GCC-格式文件。在文件中一些酶的切割位点和其它酶相同。想要只显示一个特殊识别位点的一个内切酶，不要选择这个选项（默认值＝不选择）。如果选择这个选择，图谱将会很大。所有你选择想要包括在内的内切酶的同裂酶，需要通过来自图谱界面的内切酶表的检测（按下“View Current Enzyme Table”按钮）。
-- Three frame translation（翻译）：显示沿着排列中的序列翻译（5’端到3’端的由左到右的翻译）
-- Translation of complement（互补翻译）：互补链的翻译方向相反。
-- Numbering（编号方式）：是酶切位点的核酸的号码，而不是识别位点的起点。
-- The interface window is not an MDI child and is designed to stay on top of the application.界面窗口不是MDI的产物，被设计成继续停留在应用程序的顶端。在产生限制性内切酶图谱时，窗口不显示，但是选择的选项保留默认值，直到应用程序关闭和再次打开。想要显示内切酶列表，界面窗口必须关闭或者最小化。
--可以用不同的内切酶文件替代BioEidt中的文件，但是它必须是GCC格式，必须命名为“enzyme.tab”，必须在\tables\文件夹中。

Restriction Enzyme Browser（限制性内切酶浏览器）
从核酸序列中得到内切酶谱时，显示酶的生产公司是很有用的。
通过在内切酶图谱中选择制造厂商和按下  按钮，可以手动浏览内切酶。你也可以通过选择“Options”菜单中的“View Restriction Enzymes by Manufacturer”选择，在任何时候检查内切酶。
显示以下对话框：



在这个例子中，所有来源于Stratagene的限制性内切酶显示在左边的列表中，KpnI的亮度增加。KpnI的识别序列显示在顶端，同裂酶显示在它的下方，其他提供KpnI的公司显示在同裂酶的下方。BioEdit使用ReBase提供的gcgenz表，限制性内切酶数据在万维网的地址是：http://www.neb.com/rebase/ 。可以从ReBase下载最新的gcgenz表，将其命名为“enzyme.tab”，并且替代在BioEdit安装文件夹中“tables”目录下的旧文件。

注意：表必须是gcgenz格式的。你可以从tables文件夹中打开“enzyme.tab”文件查看格式，或者查看“Restriction Maps”。限制性内切酶表格文件名必须是“enzyme.tab”，而且必须在BioEdit的“tables”文件夹里。

Codon Tables(密码子表)

BioEdit使用的唯一密码子表，它产生于GCG程序的CodonFrequency。BioEdit默认的密码子表是J. Michael Cherry (cherry@frodo.mgh.harvard.edu)编写的E.coli密码子用法表。

Six-frame translation(六框翻译)

在文档窗口中，点亮DNA序列的标题，从“Sequence”菜单中选择“Sorted Six-Frame Translation”或“Unsorted Six-Frame Translation”，可以将所有六联阅读框翻译为所有可能的开放阅读框。将会出现一个对话框，要你详细指定最小ORF尺寸、最大ORF尺寸和起始密码子。

Minimum ORF size(最小ORF尺寸)：只有密码链长度等于或大于最小值，才会报告。
Maximum ORF size(最大ORF尺寸)：只有密码链长度等于或小于最大值，才会报告。如果这一项是空白的，不限制ORF尺寸。
Start codon(起始密码子)：选择ATG或者任何你想要的三联密码子。只有起始于这个密码子的密码子链，才会报告。如果选择“ANY”，密码子链将基本上从终止密码子到终止密码子。

分类翻译和未分类翻译之间的区别：
分类翻译(Sorted)：ORF将在起始位置报告。负框序列根据他们的末端位置分类。可以分类和翻译的序列最多大约在10,500个以上。确切的数字不清楚。如果分类翻译太大，将会超过可分类序列的极限。如果出现这种情况，BioEdit将告诉你，当前它可以翻译的序列。多序列可以翻译到同一个ORF列表中，而且适合BLAST数据编辑。

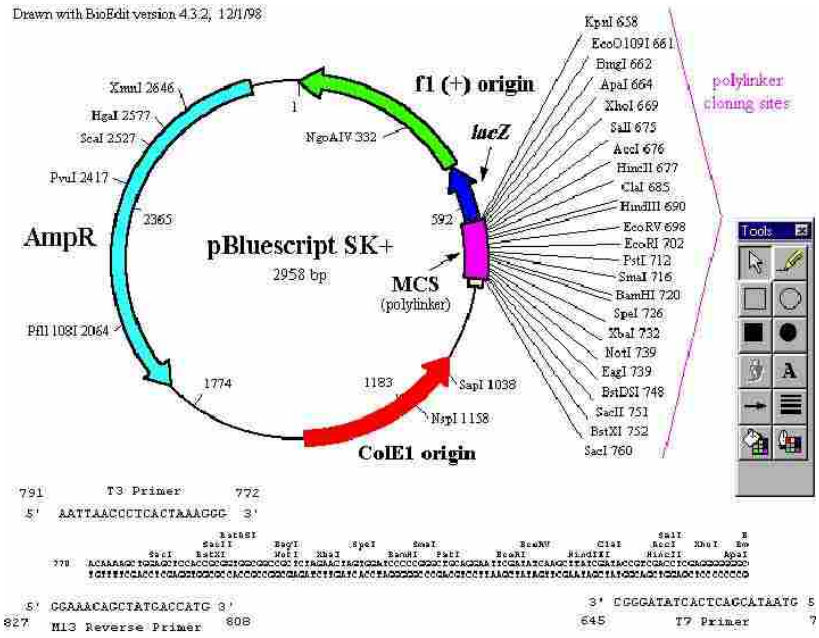
未分类翻译(Unsorted)：序列将按次序报告：遇到的终止密码子、同时通过全序列的六联框。密码子链在它们被遇到时，就写进了一个文件，因此不需要在存储器中分类。可以产生长列表。现在，一次只能翻译一个序列。

当前执行的是不固定的ORF辨认。序列简单的翻译成原始的密码子链。

ORF标题结构如下：<序列标题>: <起始碱基> 到 <末端碱基>: 框<框数> <序列长度>

Plasmid drawing with BioEdit(BioEdit的质粒绘图)

BioEdit提供简单质粒绘图的工具，以及快速简易注释。以下pBluescript SK+的载体图是BioEdit在几分钟内绘制好的。这个细节图的特征是从“Stratagene”提供的图谱中复制过来的。可是，绘制一个新的载体是简单的。

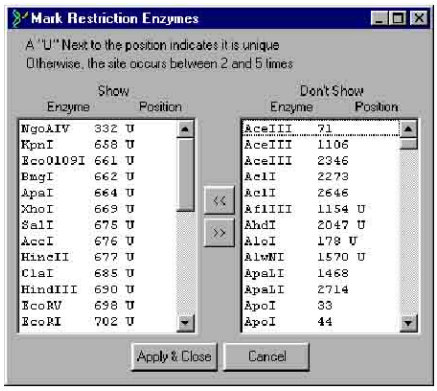


使用BioEdit质粒绘图功能，序列可以通过自动的位置标记，自动修改成环形质粒。特征、多连接位点和限制性位点可以通过使用对话框增加。当将一个序列进入质粒图时，在背景上出现一个限制性内切酶图谱，所以可以通过对话框选择可以增加限制性位点。它们自动增加到当前的位点。质粒功能提供简单的绘制和标记工具。然而，这些需要改良和延伸。标签和绘图可以通过鼠标移动和缩放。想要编辑目标性质，双击目标。

想要从一个DNA序列产生一个质粒，从“Sequence”菜单中，或者从“Sequence”菜单中“Nucleic Acid”子菜单中选择“Create Plasmid from Sequence”选项。选择这个选项时，限制性内切酶图谱将会使用通常商业化的，储存在存储器中的限制性内切酶。质粒第一次产生时，它显示成有10个位点标记的圆圈，中央是标题。

Restriction sites:(限制性位点)

想要增加限制性位点，从“Vector”菜单中选择“Restriction Sites”选项。将会显示一下对话框：

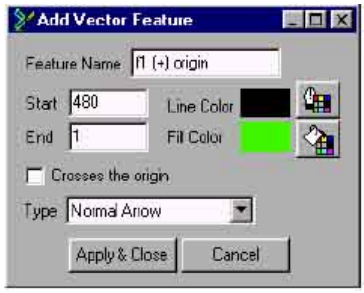


想要显示图谱中的限制性内切酶，从右边(“Don't Show”中)选择任何想要的酶，用  按钮将它们移动到左边。按下“Apply & Close”时，这个位点就会增加到图谱中。指定的酶如果只有一个酶切位点,就会在酶切位点上出现一个“U”。如果没有“U”，将会显示第一个酶切位点。想要移动图谱中酶的位置，在“Show”中增加选择的酶的亮度，按下  按钮将它们移动另一边。

Positional marks（位置标记）：
点击“Vector”菜单中的“Positional Marks”选项，可以出现以下对话框：

可以通过移动位置标记到“Show”中，单独增加位置标记，或者设定应用的分割标记数量。想要没有标记，选择“Divide into:”中的下拉菜单顶端的“None”。

Features（特征）：
想要增加一个特征，如抗生素抵抗标记，从“Vector”菜单选择“Add Feature”。将显示以下对话框：

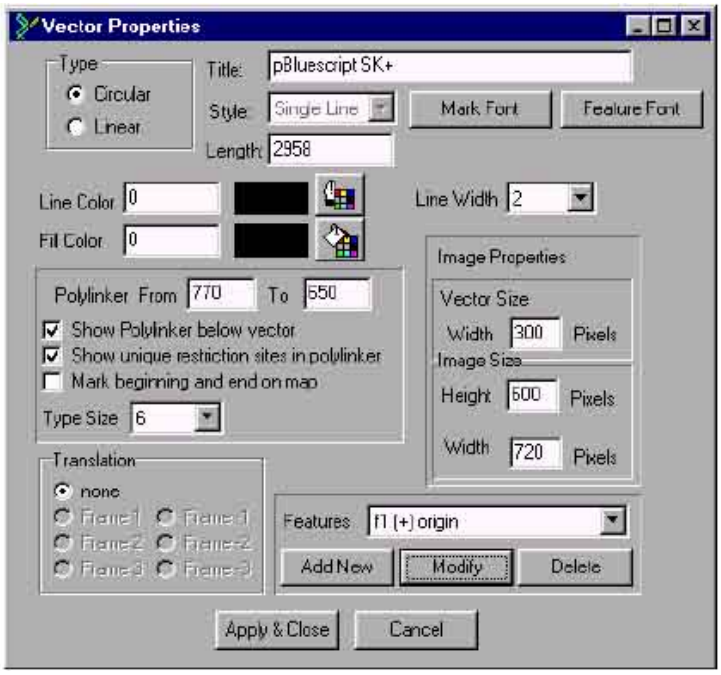


选择的类型是“Normal Arrow”、“Wide Arrow”、“Normal Box”和、“Wide Box”。在上面例子中的所有特征是“常规”宽度的。如果特征是一个箭头，箭头的方向将是从起点

位置到终点位置。

增加特征或酶时，他们各自的标记增加在外面，中心是可能的尺寸。标记可以被选择工具选择、移动、编辑和缩放。

General Vector properties（普通的载体属性）：
载体的属性可以通过选择“Vector”菜单中的“Properties”来更改：



可以通过指定起点和末端位置，来增加多接头按钮。多接头显示为“Courier New”字体。

在这个对话框中，特征可以被编辑、增加或者删除。想要编辑或删除一个现存的特征，在“Features”下拉式菜单中选择特征，并点击合适的按钮。点击“Add New”按钮，可以增加一个新的特征。

现在只有一个圆形、单链质粒是有效的。在以后的版本中中将会改进。

“Font”按钮改变指示的默认字体。特征标记的字体将可以单独改变，但是位置标记不能单独改变。

Drawing tools（绘图工具）：
使用非常简单的绘图工具，其功能和其它程序中的标准绘图工具大部分或小部分相同。目标的次序可以通过“Arrange”菜单来改变，目标可以通过“Arrange”菜单来分组或者解除分组。

注意：缩放分组目标的比例不是很好使用。

想要编辑目标的属性，双击目标，或者选择目标并且从“Edit”菜单选择“Object Properties”

Cut / Copy / Paste（剪切 / 复制 / 粘贴）:
当一个目标复制到质粒图中，它的结构也复制到内存中，使其可以在BioEdit中使用，而且目标的位图或者目标也复制到剪贴板中。目标可以向位图图像一样粘贴到其它应用程序中。

Printing（打印）:
打印时，图像可以根据打印机设定的分辨率输出成一个位图。然而，打印界面不是十分先进。左边距和顶端空白可以在“Print Setup”对话框中（来自于File菜单）详细制定。现在，不支持缩放比例输出到打印机，来定义打印尺寸。打印图的尺寸是由打印机的分辨率和屏幕分辨率的比率决定的。屏幕的分辨率是800×600。

Moving the vector（移动载体）:
用鼠标第一次选中的载体可以在页面中移动，用鼠标拖动它可以将它放置在新的位置。所有页面上的标签将相应移动。

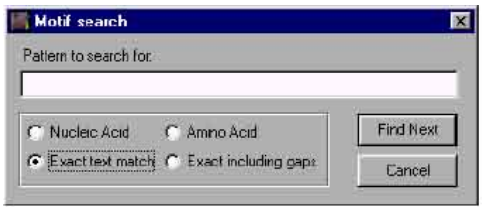
Searching functions（搜索功能）
以下的搜索选项在EDIT菜单中。搜索功能在BioEdit中不是很好使用，需要改进。

Simple search: Find and Find Next（简单搜索：寻找和寻找下一个）
这是一个非常简单的搜索功能，而且需要改良。简单搜索的菜单选项是“Edit->Find”。显示一个标准的搜索对话框，允许对选中的序列进行精确的字符串搜索。搜索通常从文档的开始向下进行，只包括标题被选中的序列（搜索只在序列中进行，不包括标题）。找到文本时，文档窗口中的第一个找到的例子亮度增加。当前的搜索位点被记忆。想要继续搜索去寻找下一个例子，选择“Edit->Find Again”（默认值是F3）。如果再次选择“Edit->Find”，搜索位置重新设定为文档的开始。

Find in Titles and Find in Next Title
（在标题中寻找和在下一个标题中寻找）
想要增加所有包含特殊文本标题的亮度，选择“Edit->Find in Titles”。想要增加下一个包含特殊文本的标题亮度，选择“Edit->Find in Next Title”。搜索开始于最后选中的标题，并指定方向。

Find Next ORF（寻找下一个**ORF**）
当搜索ORF时，只有选中标题的序列才能被搜索，而且搜索开始于最后选中的核苷酸。想要搜索，选择“Edit->Find Next ORF”，或者“Sequence->Nucleic Acid->Find Next ORF”。将根据参数对话框中ORF页指定的参数值，进行搜索。当找到一个ORF时，在文档窗口中的序列亮度将会增加。

Search for user-defined motif（搜索用户定义的基序）
BioEdit 4.7.8和以上版本允许，搜索用户根据单字母指定核苷酸和氨基酸定义的序列模式。想要在选中的序列中搜索序列，选择“Edit->Search for user-defined motif”。将会显示以下对话框：



再输入栏中键入想要搜索的文本，选择搜索类型。
在所有的四个搜索类型中，‘*’ 是通配符，可以指定任何同一性的残基。一个缺口可以指定为‘-’, ‘~’ 或‘.’。

Search type（搜索类型）：
Nucleic Acid（核苷酸）：假定被搜索的序列是所有核苷酸序列。搜索时反应迟钝的，而且只依赖于残基的同一性。缺口被忽略。以下是遵循退化残基的详细说明：
R = A or G
Y = C or T/U
K = G or T/U
S = G or C
M = A or C
W = A or T/U
B = G, C or T/U
V = A, G or C
D = A, G or T/U
H = A, C or T/U
N = A, G, C or T/U

退化匹配只限于一个方向。查询中的“R”和目标中的“R”、“A”或者“G”匹配，但是查询中的“A”和“G”不能和目标中的“R”匹配。

Amino Acid（氨基酸）：假定被搜索的序列是所有氨基酸序列。搜索时反应迟钝的，而且只依赖于残基的同一性。缺口被忽略。以下是遵循退化残基的详细说明：
X = 二十个标准氨基酸中任何一个
B = D或N
Z = E或Q

像核苷酸一样，退化匹配限于一个方向。查询中的“B”和目标中的“B”、“D”或者“N”匹配，但是查询中的“D”和“N”不能和目标中的“B”匹配。

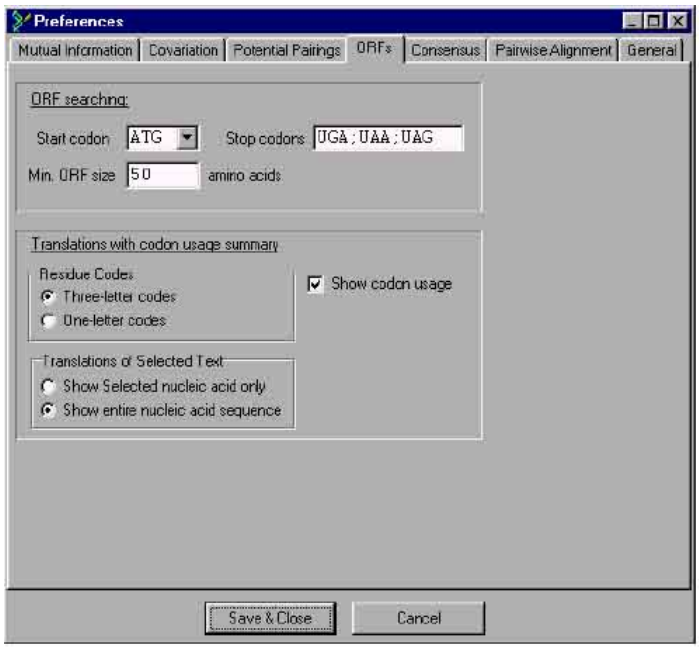
以下是标准单字母氨基酸码：
A = Ala = alanine
C = Cys = cysteine
D = Asp = aspartate
E = Glu = glutamate
F = Phe = phenylalanine
G = Gly = glycine
H = His = histidine

I = Ile = isoleucine
K = Lys = lysine
L = leu = leucine
M = Met = methionine
N = Asn = asparagine
P = Pro = proline
Q = Gln = glutamine
R = Arg = arginine
S = Ser = serine
T = Thr = threonine
V = Val = valine
W = Trp = tryptophan
Y = Tyr = tyrosine

Exact text match（精确文本匹配）：
一个反应迟钝的搜索，缺口（‘-’, ‘~’或‘.’）被忽略，‘*’可以代表任何字符。注意：即使是核苷酸序列，“T”和“U”是不同的，不考虑退化的同一性。

Exact including gaps（精确包含缺口）：
像一个精确的文本匹配，但是不忽略缺口。但是，缺口表示为‘-’，‘~’或‘.’，搜索不需要精确指出缺口字符的位置。‘*’是这个搜索中的通配符，可以用于指定任何形式的缺口。

Preferences for translation output and ORF searching
（翻译输出和**ORF**搜索的参数选择）
想要制定ORF搜索的参数或者核苷酸序列翻译的格式，通过“Sequence->Nucleic Acid->Translate-> ...”菜单选项，选择“Options->Preferences->ORFs”：



ORF搜索：用于ORF搜索的起始密码子通常是ATG，但是，你可能想要搜索选择性的起始

密码子。想要允许一次超过一个起始密码子，用“；”分隔密码子。相同的结构也用于终止密码子。如果你想要允许密码子的通读，把它从列表中移去。将会保存所有子序列搜索的参数。

搜索ORF时，只有标题被选中的序列才能搜索，而且搜索开始与最后被选中的序列。想要搜索，选择“Edit->Find Next ORF”，或者“Sequence->Nucleic Acid->Find Next ORF”。

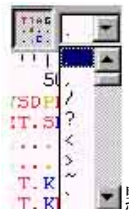
格式化的核苷酸翻译可以通过选择“Sequence->Nucleic Acid-> Translate”显示。如果点击“Show codon usage”，将会显示一个关于核苷酸翻译中密码子选择的一览表。

Conservation plot view（保持标绘视窗）

有时候它可以很方便的通过参考标准序列（通常是顶端的那个）绘制排列图，和标准相同的任何残基都被绘成一个特定的字符（通常是一个圆点）。BioEdit由两种方法执行这个功能：

1.选择“Alignment->Plot identities to first sequence with a dot”，形成一个全新的序列排列文档，而且和第一个序列有同一性的被绘制成圆点。在这个新文档中，修改成圆点的序列残基数据是不保留的，新的排列文档值产生一个排列图。你可以选择“File->Graphic view”，不要点击类似性和同一性描影的选项，绘制成图表。

2.想要根据标准序列同一性标记的排列动态视图，按下工具条上的  按钮，或者选择菜单中的“View->Conservation Plot”。当按下保持标绘视图按钮时，会出现一个用于标绘同一性的指定字符选项：

 默认值是圆点，但是理论上任何字符都可以使用。实际上，句点或空格通常容易发现。

你可以通过右键点击你想要成为参考值的序列标题，无数次的改变参考序列。但是参考序列被它在列表中的数值内部处理，所以如果你要上下移动序列，你将不得不再次右键点击它的标题。

基本分析工具

BioEdit 自带了一些分析工具，以下详述。这些工具分为两类：

1 外部的独立程序。它们由其他作者编写，它们或随 BioEdit 一起发布，或可通过其他的途径获取并在 BioEdit 界面下运行。用户可在 BioEdit 提供的图形界面下通过配置产生命令行指令，从而运行外部的程序并将序列数据输入，这样在一个界面即可获得一个简单的序列分析环境。

2 BioEdit 自带的分析功能。

外部附带的程序

安装 TreeView:

TreeView 是由 Roderic D.M.Page 编写的用来观看系统发育树的程序。以前版本的 BioEdit 在 apps 文件夹中包括了 TreeView 的可执行支持库。应作者要求，全部 TreeView 的安装程序，即 TreeView.zip，现随 BioEdit 发布。安装时，首先把文件解压缩到一个临时目录中，然后执行 setup.exe，程序将自动安装到你的系统中。配置 TreeView 时：

1）从“accessory application”菜单中选中”add/remove/modify an accessory application”.

2）在“name of accessory”栏填写“TreeView”选择“specify”按钮，在打开的框中，选择 TreeView.exe。

3）选择“prompt for input file”.

4）在“general description”栏，填写”TreeView version 1.5.2.Copyright Roderic.D.M.Page,1998.r.page@bio.gla.ac.uk.http://www.taxonomy.zoology.gla.ac.uk/rod/rod.html”

5）点击对话框底部的“add/modify”,关闭窗口。此时出现提示，即将关闭程序再重新打开运行。

关于配置外部程序的详细资料，请参阅“配置及使用外部应用程序“一节。

配置及使用外部应用程序

BioEdit 提供了一个添加和配置外部程序的接口，外部程序将被添加到序列联配文件的“accessory application”菜单中。当正确配置后，在菜单中选择某一应用程序后，通过 BioEdit 提供的一个图形界面，此程序即可调用。虽然任何应用程序经配置后均可使用，但 DOS 及 WIN32 下的程序由于可接受命令行参数而运行，从而更为方便。BioEdit 可自动将序列数据输入到应用程序中并在完成分析后自动输出结果。可同时打开多个输出文件，某个程序的输出结果可被另外程序调用。

添加及配置一个新应用程序

BioEdit 2.0 及以上版本提供了一个配置外部应用程序的图形界面，从而在序列联配文件中调用外部应用程序，但首先必须要在 BioEdit 环境以外会使用此外部应用程序。BioEdit 还自带了一些应用程序，它们在安装 BioEdit 同时已经安装及配置好了。Joe Felsenstein 允许随 BioEdit 一同发布 PHYLIP。Roderic D.M.Page 允许发布 TreeView.但 TreeView 不再预先配置。应作者要求，TreeView 的安装程序随同 BioEdit 一起发布。在安装 BioEdit 时，TreeView 被放置在主安装文件夹中。详细的情况请参阅“安装 TreeView”。

要在“accessory application”菜单中添加新应用程序，从“accessory application”菜单下选择“add/modify/remove an accessory application”,然后需要指明一些设置参数。由于外部应用程序由不同的作者编写，所以每个程序的配置也不同，但许多程序并不需要太多设置。使用者必须首先会在命令行下使用某个程序，然后才能通过正确的配置在 BioEdit 中调用。所以请首先参考程序的使用文档学习使用方法。以下是配置界面中的各个选项，但只有前两项是对所有的程序都必须的。

1 附带程序的名称（name of accessory）:即在“accessory application”菜单选项中出现的名称，任何名称均可，但建议短些。

2 程序（program）:即程序存放的决定或相对途径，包括程序名称（通常是.exe 文件，但也可能是.com 或.bat 文件）。如果要指名程序相对于 BioEdit 安装目录的相对途径，就把主安装目录指明为“<BioEdit>”(不分大小写)。例如一个应用程序”myapp.exe”存放在 BioEdit 的主安装目录下的 apps 目录下，则可指明“<BioEdit>\apps\myapp”(不带引号)。这样当移动 BioEdit 目录时，不会影响查找某个应用程序。另外也可指明决定途径，可点击”specify”浏览磁盘并查找程序。

3 自动向程序中输入序列（automatically feed sequences to app）:如果应用程序分析序列或联配数据，（如 ClustalW.PHYLIP）可选择自动向程序中输入序列，使序列联配

后直接运行程序更为方便。如果某个应用程序只需要一个或几个序列也可使用此功能，因为程序运行时 BioEdit 只自动输入选定的序列。（如果没有选择，序列将自动地全部被选中）

4 特别的文件名（specific file name）（针对自动输入的数据）：某些应用程序需要一个特别的文件名。例如，PHYLIP 处理名为“infile”的文件。如果程序需要这样，那么选中该栏并在”file name”栏填入需要的文件名，不要包括文件路径。文件会自动保存在应用程序的目录下。

5 去除空位的序列(degap sequences):一些应用程序(如 DNAml, Protodist,DNAdist)需要联配数据的空位包括在输入序列中,而另一些应用程序(如 BLAST)只要求简单的序列数据而不是联配后的数据,此时空位必须去除,否则空位将被看作残基.

6 格式(format): BioEdit 可按以下 8 种格式向应用程序中自动输入序列.(Fasta,GenBank,PHYLIP2/3.2,PHYLIP4,NBRF/PIR,MSF,GCG,EMBL)如果你有某个程序要求另外的格式,你需要先选择不自动输入序列,然后把文件转换成程序所须的格式,最后运行程序.或者另外单独运行该程序.如果你觉得在 BioEdit 中运行该程序很方便,请将文件的格式及规格通知我(tahall2@unity.ncsu.edu).我会编写出文件输出过滤程序,并将它加入到新的 BioEdit 程序中，同时通知你下载的地址和时间.

7 输入文件提示(prompt for input):在程序运行时你可能希望 BioEdit 提示你指明输入文件名.选择该栏 BioEdit 出现提示对话框.

8 输出文件提示(prompt for output):在程序运行时你可能希望 BioEdit 提示你指明一个输出文件名.

9 以联配文档形式打开(open as aligment):程序的主输出以联配文档格式打开.缺省情况下,主输出是一个文件,但可在”additional output files”栏中指明其他的输出文件.此栏也可空置或指明单一的输出,在功能上无区别.

10 以文本形式打开(open as text) 如果输出是文本数据,可选择在 BioEdit 的文本编辑器中打开.

11 在外部程序中打开(open with external program) 如果你需要在象 Microsoft Exel 这样的制表程序中打开表格或矩阵数据,你可指明任何外部程序并自动在其中打开输出数据.你可点击”specify”浏览磁盘选择程序,你也可指明相对于 BioEdit 安装目录(以<BioEdit>表示)的路径.

输出也可以同时选择上面三种形式.

12 使用输入前缀(use input prefix):某些程序需要在命令行中以前缀的形式指明输入文件,输出文件,参数的输入等,而另有些程序仅需要以规定的顺序写入输入,输出及参数.如果你的程序需要使用前缀,选择该栏,并在编写框中准确写入前缀.注意:如果前缀和文件名之间有空格,在前缀后要输入空格.例如,某个程序可能是”-i inputFile”而另一个可能是”input=inputFile”.当前者少一个空格或后者多一个空格，程序都可能无法运行.

1 3 使用输出前缀（use output prefix）：同 1 2

1 4 指明输入文件名（input name required）:某些程序在命令行中要指明输入文件名，（如 ClustalW）而另有些并不需要（如 PHYLIP）.如果你的程序需要就选中此栏.

1 5 指明输出文件名（out name required）:同 1 4

1 6 任意指明输入输出文件名（input and output name arbitraty）：有时在命令行下随需要指明输入输出文件名，但可以是任何文件名．此时可选择此框，BioEdit 将自动为输入输出分配任意的文件名．例如，当配置 ClustalW 时，选择自动向程序输入序列，自动以联配形式打开并选中了 " input name requied”及 " arbitrary”后，输入及输出文件就被分别分配了 " ～in Temp.tmp”和 " ～out Temp.tmp”的名称.

1 7　从标准系统输入到标准系统输出（from stdin and to stdout）:stdin 和 stdout 分别指标准系统输入和标准系统输出，缺省指从键盘输入和从显示器输出．某些程序需要从 stdin 输入，向 stdout 输出（如 FastDNAml）,为了使程序自动运行，需要使输入输出改向．即当需要从某个文件向程序输入时，选中＂ redirect general stdin from file”框，在栏中写入输入的文件名及路径．当需要保存输出到某个文件或在其他程序中使用时，选中＂ redirect general stdout to file”框，在栏中写入输出的文件名及路径．

1 8 选择框（checkboxes）:BioEdit 界面下运行外部应用程序可包含多达 5 0 个选择框，这样当运行程序时，可适当设置．添加选择框时，在＂checkboxes”框中写入需要的选项，点击＂add/modify”按钮，在出现的对话框中可指明选择框（选择或不选择）及命令行的缺省值．如果没有命令行参数，就空置．点击ＯＫ确认．要修改某个已写入的选择框，在下拉列表中选中后，点击＂add/modify”. 要删除某个已写入的选择框，在下拉列表中选中后，点击＂delete”按钮．

1 9 输入框（inputs）:某些程序需要输入一些能影响程序执行的数据．（如ＣＡＰassembly 需要用户设置最小碱基重叠值 base overlap 及最小匹配百分比 percent match）象选择框一样添加，修改及删除输入框．输入框可与一个选择框联系以供在程序中选择是否使用该输入值.同时还可在命令行中指明参数的命令前缀(可能需要,如果不需要就空置)．理论上，最大可有 5 0 个输入框，每个均可选择是否与一个选择框联系，如果选择则必须输入该框的名称及缺省值．

2 0 其他输出文件（additional output files）:除了主输出文件外，还可指明 10 个其他输出文件并分别处理。注意：此处假设外部应用程序自动产生其他的输出且文件名不用在命令行中指明。如果实际的程序并非如此，可以在“default command line”中添加适当的参数或如下配置。象对待选择框和输入框一样添加，修改和删除此项。为每一个其他输出文件命名，一般不需要包含路径（大多数程序会自动将输出保存在本身所在的目录下），但如果指明路径是必须的可指明。某些程序（如 ClustalW）可产生一个和输入文件名相同但文件扩展名不同的输出文件。在这种情况下，指明文件名为”<infile>.ext”(例如，如果输入文件名是”temp.tmp”并且输出指明为“<infile>.out”则输出文件名是“temp.out”

输出可按联配文档，文本文件或由外部程序三种形式以及三者的任意组合形式打开。如果输出没有由外部程序打开，包含输出的临时文件会自动删除。

2 1　缺省命令行（default command line ）:在程序运行时一些参数需要指明，填入此栏中。如 ClustalW 可按 GCG,GDE,PHYLIP,PIR(NBRF)格式输出。由于 BioEdit 可从内部读出ＮＢＲＦ／ＰＩＲ文件，把缺省命令行设置为“/output=PIR” 可以使 BioEdit 很快地在联配文档中读出输出。

2 2 在命令行中加入输入文件(名)(add input file to command line):如果选中此栏，BioEdit 将根据对输入文件的配置自动产生包含输入文件及命令前缀的命令行。当输入文件在命令行中的位置很重要且不在头尾时，可不选择此栏并直接在“default command line”中写入命令。否则要在”at the beginning”和“at the end”中选择以指明输入文件名是在命令行的何处。

2 3　在命令行中添加输出文件（add output in command line ）：同 2 2

2 4　查看文档（view the documentation option）:如果应用程序带有文档或你是在为其他对此程序不熟悉的人配置，你可能希望有一个直接和文档的链接。如果文档是文本格式,你可选中该栏，点击“specify”指明要查看的文档或在“documation file “中写入文档路径。同样，“<BioEdit>”指 BioEdit 的安装目录。如果选择此栏，将在界面上出现”view documentation”按钮。

２５　包括选项框（include an option box）(在其中输入命令行参数)：如果你希望在程序运行时在界面上出现一个输入框，以便可以随时输入命令行参数。就选择此栏。如果你的程序命令行选项需要很独特的安排，但通过 BioEdit 运行也十分方便，你可创建一个仅包含此选项框的界面，在框内写入命令以运行程序。

２６　标准系统输出改向（redirect stdout）:某些程序可能是打印输出或屏幕输出，如果你希望把输出保存并且以后在 BioEdit 中调用，选择此项并指明输出文件名以及在 BioEdit 中以何种形式调用。

２７　标准系统输入改向（redirect stdin）:某些程序（如ＰＨＹＬＩＰ）提供了一个可进行参数设置的菜单。如果某些设置任何时候总是需要的，那么可以创建一个包含这一系列设置的文件，同时把 stdin 改为从此文件中读取。这样就可替代从键盘上输入的菜单选项。目前为止，此功能还没有在 BioEdit 中实现，但并不表明它不可能实现。我之所以没有删除此选项就是希望今后能解决此问题。但现在我不能确定该功能是否在某些程序中可用。最好不要使用该选项。

２８　程序的一般描述（general description）:此描述长短随意，但必须以单行文本格式输入。

如果输入返回值，描述将在第一个字符返回时被打断。描述一般包括程序及作者的简单情况。

２９　添加及修改键（add/modify）:点击此按钮保存输入信息并在“current configuration “框中列出现在的配置信息。

点击“close”按钮将不更新配置信息并关闭对话框。点击“print configuration”按钮将打印现有配置信息。

修改配置信息

从“accessory application”菜单中选择“add/remove/modify an accessory application”,在打开的对话框中点击”name of accessory”栏的箭头，在下拉列表中选择要修改的程序，点击“open”即可进行修改。同样可使用“add/modify”及”delete”按钮对选择框及输入项进行添加，修改和删除。

移去应用程序

要移去某个应用程序，在“name of accessory”的下拉菜单中，选中该程序，点击”delete”.

应用程序配置信息的存放：

应用程序的配置信息保存在一个名为“accApp.ini”的文件中，该文件位于 BioEdit 的安装目录下”apps”文件夹内。该文件的结构和在 windows 目录”BioEdit.ini”文件相同。以下是 ClustalW 的配置信息。可能乍看起来有些迷惑，但只要稍加解释即可明白。该文件可直接在编辑器内编辑，但通过图形更容易且好理解。该文件的格式如下：选择框（Checkboxes）以“c<checkbox#>”表示，输入项（inputs）以”i<input#>”表示，数字都从 0 开始。参数不使用或无值时，以空白表示。用 1 或 0 代表ＹＥＳ或ＮＯ选择。

（译者注：ClustalW 的该文件配置参数可通过上述方法打开查看或参阅英文文档。）

应用程序示例：配置 **ClustalW**

以下是配置 ClustalW 的步骤。事实上，在 BioEdit 中已经编写了运行 ClustalW 的界面，但你会发现在正确配置后，新的界面和 BioEdit 内置的老界面有些不同，但功能完全一致。你还可以发现，从新的界面下运行 ClustalW 是在不同于 BioEdit 的”线程”中运行的。这说明你可以继续做其他事情而让应用程序在后台工作。

要配置应用程序，你首先必须知道需要哪些命令行选项以及他们的详细的名称。对于新程序请参阅该程序的说明文档。

以下是配置 ClustalW 的步骤：

1 从“accessory application”菜单中选择“add/remove/modify an accessory application”,打开配置对话框。

2 在“name of accessory”栏键入“ClustalW example application”

3 点击“Specify”,选择“clustalw.exe”.（此时从“app”文件夹开始浏览，clustalw.exe 就包含在此文件夹内。以后每当出现”BioEdit“时，就改为<BioEdit>）例如，如果在”program”文本框中出现“C:\BioEdit\apps\clustalw.exe”则改为”<BioEdit>\apps\clustalw.exe”

4 选中“automatically feed sequences to App”框。

5 选中“Fasta”框，以 fasta 格式作为输出格式。

6 如果希望在运行 ClustalW 前去除序列中的空位，则选择此框，否则不选。但在本例中选和不选都可。

7 选中“open output as new alignment”.此区域内其他选项空置。

8 选择“use input prefix “和“use output prefix”框。

9 在“input file command prefix”栏键入“/INFILE=”

10 在“output file command prefix”栏键入“/OUTFILE=”

11 选中“input name required”和”output name required”两栏及“arbitrary”栏。

12 空置“Space between input prefix and command “和“Space between output prefix and command “.

13 选中“add input file to command line”和“at the beginning “两栏。

14 在“CheckBoxes “栏键入“Full Multiple Aligment”,点击”add/modify”.在出现的对话框中，选中“command if checked”并键入”/ALGLN “。然后选中”Default checked “框。空置“command if not checked”

15 在“CheckBoxes “栏键入”Calculae NJ Tree”点击”add/modify”在出现的对话框中，选中“command if checked”并键入”/TREE”。空置“command if not checked”.。不要选中”Default checked “框。

16 在“CheckBoxes “栏键入”FASTA Algorithm for Guide Tree”，点击”add/modify”在出现的对话框中，选中“command if checked”并键入”/QUICKTREE”。空置“command if not checked”。不要选中”Default checked “框.

17 在“inputs”框中键入”Number of Bootstraps”，点击”add/modify”在出现的对话框中，选中“command prefix”键入”/BOOTSTRAP= “。在”default value input “框中键入“1000”。选中“associate a checkbox”并键入”Bootstrap NJ Tree”.选中“default checked”。

18 在“default command line”栏键入“/output=PIR”。

19 选中“view documentation option”。点击“specify doc file”，然后选择”clustalw.txt”.最后把文档路径修改为以<BioEdit>/开始。

20 选中“include an options box “

21 在“general description”栏键入关于 CluastalW 和作者的一些信息。（请参阅英文说明）

22 如果已经配置了 TreeView,在“additional output file “栏键入“<infile>.dnd”,点击”add/modify”在出现的对话框中，选中”open with external program”，然后点击“specify”,选中”treev32.exe”,（该文件一般在 C:\Program File\Rod Page\Tree View 目录下。）

23 在“additional output files”框中键入”clustalw.sto”并点击“add/modify”按钮。在出现的对话框中，选中”open as new text document”。点击”OK “。

24 选中“redirect general stdout to file “,键入“clustalw.sto”。

25 点击“add/modify”按钮，在“current configuration”栏将会出现配置信息的总结。

26 点击“close”关闭对话框。将提示是否重新打开 BioEdit 以使配置更改生效。点击 YES。在“accessory application”菜单中出现”ClustalW Exmple Application”.

27 选择一个序列文件，在“accessory application”菜单中选择”ClustalW Exmple Application”.

将会看到以下界面（请参阅英文），它和 BioEdit 内置的 cluatalw 功能相同。

以下部分标有 REFERENCE 表示作者给出的参考文献，请参阅英文。（译者注）

BLAST

BLAST（Basic Local Alignment Search Tool 基本局域联配搜寻工具）是搜寻查询序列和已知序列之间同源性的最方便的方法。它搜寻查询序列和数据库序列之间局部联配区域的相同或相似性。其算法的基本思路如下：

1. 查询序列被分成短的序列片段，称为“字”（word）。（按蛋白序列和核酸序列不同，“字”长为 3—8 个残基）（译者注：原文中“残基“（residue）一词有时指代蛋白序列中的氨基酸和核酸序列中的核苷酸；有时仅指代核酸序列中的核苷酸，如在 RNA 结构分析中。）
2. 构造一个“索引”表（a lookup table）。表中的序列具有相同的“字”长，且和来自查询序列的“字”的配对高于一个预定的阈值。
3. 在数据库序列中搜寻是否出现表中的序列。
4. 如果在库中找到某个“字”和查询序列的“字”的联配高于临界阈值，联配向序列两端延伸。联配一直延伸到某一设定的延伸长度，如果联配值不再增加，延伸终止。
5. 延伸终止后，报告那些联配值高于另一设置的阈值的联配。
6. 当产生一个阈值联配后，再次对库搜寻其他非冗余高分片段配对（high-scoring segment pairs HSPs）,这些 HSP 联配值高于另一个设定的阈值。（两个序列之间的几个不显著的联配合在一起可能有显著的相似性，这种相似也表明了同源性。）
7. 计算出一个概率，表示一个和查询序列长度相同的随机序列搜寻同一个数据库得到相似分值的 HSP 的概率。

REFERENCE

BLAST 程序

Blastn:核酸序列对核酸库的搜索。

Blastp:蛋白序列对蛋白库的搜索。

Tblastn:蛋白序列对核酸库序列六个读框的搜索。

Blastx:核酸序列的六个读框对蛋白库序列的搜索。

Tblastx:核酸序列的六个读框对核酸库序列的六个读框的搜索。（很慢）

本地使用 **BLAST**

NCBI BLAST2.0 版本包含在 BioEdit 中，存放在安装目录下\apps 文件夹中。如果你对在一个仅经过部分测序，序列还未整理的基因组中发现特殊的基因感兴趣，此程序会很有用。使用本程序时，你首先要构建一个本地数据库，数据库的数目自定，但必须遵照如下要求：

1 核酸和蛋白库不能混淆；

2 数据库存放在\database 目录下，以便 BioEdit 中的 BLAST 程序查找。（但 NCBI blastall.exe 是个完全独立的程序可在 BioEdit 以外运行。）当在 BioEdit 下创建数据库时，它自动存放在\database 目录下。

创建本地数据库

从“accessory application”菜单中选中“BLAST”，再选中“Creat a local …database file”. 提示输入 FASTA 格式的文件。剩下的步骤自动完成。新创建的数据库名称将出现在本地

BLAST 界面的数据库列表内。注：如果在本地 BLAST 界面的数据库列表内没有你新创建的数据库，那么退出 BioEdit,再重新进入。如果仍然没有，就重命名*.pin(蛋白)和*.nin(核酸)文件，你可以仍然按原来的名称命名。

本地 **BLAST** 搜寻

选中要查询的序列标题（title），从“accessory application”菜单中选中“BLAST”，选中“Local Blast”。序列中的空位会自动剔除。如果你想成批的处理序列，可选择多个序列。然后在打开的界面上选择你想运行的程序（如 blastp）,在右上方的下拉条中选择要搜寻的数据库。再选择是将结果保存在用户命名的文件中还是由 BioEdit 自动创建一个临时文件保存。

BLAST INTERNET 客户端程序

BioEdit 包含了 NCBI BLAST 客户端程序的第三版。（存放在\apps\blastcl3.exe）。在联配窗口中选择一个或多个序列，接着选择“accessory application”->BLAST->NCBI BLAST over the internet”出现以下窗口（请参阅英文示例）

你可同时 BLAST 多个序列，也可选择以 HTML 格式输出，。如果选择了 HTML 则输出自动在你的浏览器中打开。否则是在 BioEdit 内的文本格式。你可选择任何标准的 BLAST 的格式选项（成对配对（pairwise）是缺省值）（请参阅英文示.）

REFERENCE

ClustalW

ClustalW 是由 Thompson,J.D,等编写的多序列联配程序。它能基于轮廓的逐步联配（profile-based progressive aligment procedure）多个序列。在 BioEdit 中此程序未做更改，并提供了与帮助文档的链接。此帮助文档即**\apps** 目录下的”clustalw.txt”。BioEdit 内的 ClustalW 界面简单明了,选项的描述参见帮助文档。

当 BioEdit 内的 ClustalW 运行完毕后,将产生新的联配文档，此文档内的序列按你提交时的顺序排列,并包括原始的标题,GenBank 及其他的图形注释，用户定义的分组信息。这样，序列原有的信息不会丢失而且不易混淆。

运行 ClustalW 时,先选择要联配的序列标题,如不选则默认为全部序列。接着”accessory application”->”ClustW Multiple aligment”。

使用互联网工具

用 Webcutter 产生限制性内切酶图谱

在编辑框中选择序列标题,选中”World Wide Web”->”Auto-fed Restriction Mapping”接着选定一些选项,最后点击”Analyze Sequence “，不久就可返回限制性内切酶图谱。你也可在外部的浏览器内使用该项。

BioEdit 现在也自带了限制性内切酶图谱功能。

HTML BLAST 网络浏览器

使用它和 IE 及 Netscape 的唯一区别是，选中的序列会自动剔除空位并被输入 BLAST 窗体。在 BioEdit 中使用其它外部浏览器同样如此。它最明显的优点是 BLAST 的结果带有通往 ENTREZ 和 Medline 的链接。由于 BLAST 客户端的程序超过了 1M,下一版的 BioEdit 可能不再包括。

使用 WWW BLAST,在选择序列后,选中”World Wide Web”->”Auto-fed NCBI Standard BLAST”。

PSI-BLAST 位点特性反复 **BLAST (Position-Specific Iterated BLAST)**

PSI-BLAST 位点特性反复 BLAST (Position-Specific Iterated BLAST)是 NCBI 提供的最新的搜寻算法，它是原始 BLAST 算法的一个变种。它提供了类似于在一个由一组同源序列构成的一致性矩阵中搜寻的方法，这样对于远缘同源性同源性更敏感。

PSI-BLAST 是 BLAST 的进一步发展,它将标准的 BLAST 产生的一批联配高于设定阈值的高分片段配对(HSP)与查询序列联配，产生一个位点权重的公共矩阵。然后此矩阵代替查询序列对库搜索。每次叠代都用新产生的矩阵代替查询序列再对库搜索。如此反复，矩阵也随之不断更新。多数情况下,矩阵最终收敛于一点，此处矩阵不再改变。得到的联配可揭示常规 BLAST 完全忽略的远缘同源性.

REFERENCE

PHI-BLAST

PHI-BLAST(Pattern-Hit Initiated BLAST)模式匹配启始 BLAST。搜寻用户指明的模式（pattern）或 motif，给出在模式匹配周围的类 BLAST 的局域联配。

REFERENCE

Prosite profile and pattern scans(Prosite 轮廓和模式扫描)

自动连接 Prosite 轮廓(profile)和模式(pattern)扫描的网页。把蛋白或核酸序列和与轮廓库(profile library)比较；把蛋白序列和 prosite database 中的模式（pattern）比较。从"sequence"->"world wide web"进入。

REFERENCE

nnPredict 蛋白质二级结构预测

nnPredict 是由 Donald.Kmeller 编写的基于神经网络算法的预测氨基酸序列中残基的二级结构的程序。BioEdit 提供了 www 上的链接。

REFERENCE

其他的链接

ENTREZ &PubMed

由 NCBI 维护，PubMed 包括免费的所有 Medline 目录。

Pedros BioMolecular Research Tools

包括丰富的生物技术和分子生物学的链接,尤其是通向 www 上的服务。

构建 **BioEdit** 的 **www** 书签

BioEdit 的"world wide web"菜单内可保存多达 500 个 www 书签。也可在序列编辑器中使用你的外部浏览器，以到达需要的序列分析网址。这些书签保存在"apps"文件夹中的"bookmark.txt".如果名字改变，BioEdit 将不认识。如果文件损坏，你可以编辑缺省的书签文件。要求的格式是每个输入包括两行文本,一行描述,另一行给出 URL。例如：name=description,address=exactURL。如果你的输入在"world wide web"菜单上没有出现,请检查格式是否正确,不正确的格式将被忽略。你可在文本编辑器内对此文件修改,也可以从"world wide web"->"View bookmarks"修改，它将在文本窗口打开。

BioEdit 内置的分析工具

氨基酸和核苷酸的组成**(**分析**)**

从 "Sequence" 菜单下进入 "Protein", 再进入 "amina acid composition"; 或者从"Sequence"->"nucleotide acid"->"nucleotide composition",分别可对序列的氨基酸和核苷酸组成分析,结果以摘要和图例的形式给出。图例中的柱形条表示每种残基在序列中的摩尔比。如果核酸序列内有简并符号也会在图中列出。例如,某序列仅有 A,G,C,T,则图中只有 4 个柱形条,但若有 R,Y,M,等它们也会被计算在内。以下是个例子.(请参见英文)

柱形条的颜色是在颜色列表(color table)中选定的。

氨基酸的情况类似，但是仅计算标准的 20 中氨基酸,其他的将被忽略。

分子量：

蛋白质的分子量是每个氨基酸分子量的总和。即 HN-C(R)-C=O,其中 R 是侧链基团，加上氨基端和羧基端的氢和氧。核苷酸的分子量是先将单磷酸核糖核苷酸或脱氧核糖核苷酸求和

再每个核苷酸扣除一个水分子，链的 5'端多一个氢，3'端多一个水以形成磷酸基团和羟基。以下是核苷酸的分子量：在 RNA 中,A(328.2)；G(344.2)；C(304.2)；U(305.2)；平均(320.5)。在 DNA 中：A(312.2)G(328.2)C(288.2)T(287.2)平均(304.0)。数据以单磷酸核苷酸中的所有原子之和扣除一个氧原子和两个氢原子。（具体的 C，O，N，P，H 的原子量请参见英文）

熵图（**Entropy plot**）

在联配文件中有专栏用熵图来衡量可变性。它衡量的是在联配中每个位置的"信息量"的缺乏。准确地说,是每个位置的可预测性的缺乏。假如在一个联配中有 x 条序列(x=40),在第 5 个位点(y=5)上,所有的序列都是 A,我们可以认为我们对第 5 个位点的"信息"认识已经很多,如果让我们预测另一个同源的序列的第 5 个位置,我们会说是 A,即我们对第 5 个位置有最大的信息量,熵是 0。但如果在位置 5 上,有 4 种可能(A,G,C,T),每种出现的频率均为 0.25，则我们的信息量降至 0,而熵具有最大的可变性。

信息量的单位是"比特"(bit),表示"或者…或者","YES or NO","开/关"等有两种状态的情况或以 2 为底的记数系统。在核酸序列中,每个位置有 4 种残基的可能性，则需要有 2bits 的信息量才能决定此位置(如嘌呤或嘧啶,1bit;如果是嘌呤,则是 A or G,另 1 bit)。对于包含 20 个标准氨基酸的蛋白序列则需要 5 bits(No1-No10;No1-No5;No1-No3;No1-No2;No1? 共 5 个 yes or no 的回答.)如果在联配序列的某一位置,总是出现 A,我们对此位置有最大信息量(2bits),十分肯定。

在数学上,Claude Shannon 建立了信息论的基础,H(l)=-∑f(b,l)logf(b,l) 以 bit 为单位 H(l)指 l 位置的不确定性,即熵值。b 代表某个残基,f(b,l)指残基 b 在 l 位置出现的频率。信息量定义为某个位置 l 的不确定性(或熵)的减少。联配质量的提高,某位置(尤其是保守区)的熵值应减少。

BioEdit 以每个位置熵值，而不是信息量作图，因为为了决定位置的信息量，全部可能的残基数必须知道,但这取决于分析时是否使用空格或简并核苷酸残基(如 S,M,K,W)，但使用熵值时,序列以字符矩阵处理。矩阵的每列某个位置的熵值和序列上某位置所有可能的信息无关,而只取决于某个字符在此列出现的频率。为方便起见,BioEdit 使用自然对数,而不是以 2 为底数。单位是"尼特"(nit)而非"比特",但数据之间的关系是相同的。熵值 H(l)=-∑f(b,l)ln(f(b,l),给出每个位置相对于其它位置的不确定性的度量。最大总不确定性定义为,在一列中出现不相同字符的最大个数.例如,有联配中有 42 条序列,每个序列可由 20 个氨基酸和空位组成,不包括用户自定义的字符,在矩阵中每个字符恰好出现 2 次,则最大不确定性是 21*(1/21)ln(1/21)=3.04,这不是以 bit 为单位的,如要转换只须转换为取 2 为底数的对数 H(l)=-∑f(b,l)*(lnf(b,l)/ln2).但这并不需要。

要做熵图时,先选择要分析的序列,再从"Aligment"进入"Entropy(Hx)",将会出现一个图形窗口,在文本中出现熵值。如果使用屏蔽(mask)（译者注：请参见后面" 使用屏蔽（masking）"专门的一节说明 mask 的含义，一般情况下译文中用"屏蔽"一词，但有时也直接写为 mask）,则只有屏蔽的位置被分析,如果使用编号屏蔽（numbering mask），则号码表示的是编号屏蔽中的位置。

疏水性轮廓**(profile)**

平均疏水性轮廓采用 Kyte &Doolittle 的方法,他们从文献的实验数据中为每个氨基酸编辑了一个"hydropathy score"(hydropathy 分值)。一个设定大小的窗口沿序列移动,随之计算 hydropathy 分的总和。平均分值(总和/窗口大小)作为序列中各个位置的疏水性值,并以窗口中中间残基的疏水性值作图。

瞬间疏水性轮廓(hydrophobic moment profile)作图绘出的是设定长度的片段的瞬间疏水性轮廓。例如,设定窗口值为 21 残基,给某位置绘出的值是当窗口移动到此残基左右两边各有

10 个残基时的瞬间疏水性值。此值由 Eisenberg 等给出的公式计算:μH={[∑Hnsin(δn)]^2+[Hncos(δn)]}^(1/2),其中μH 指瞬间疏水性,Hn 指残基 H 在位置 n 的疏水性分值，δ＝１００度，n 指片段中位置，每个设定窗口长度的片段都计算瞬间疏水性.

平均瞬间疏水性作图使用同样长度的窗口绘出设定窗口长度片段的平均瞬间疏水性．例如，设定窗口长度２１，以第一个残基所在的片段为窗口长度也是２１，计算其值，以第二个．．．．依次类推，共有２１个窗口长度为２１的片段，计算总和，求平均值即代表原来片段的中间一个残基的平均瞬间疏水性.

以前版本的 BioEdit 仅绘出某个片段第一个位置的平均疏水性，结果图在通过了 L_W 点，（L 指序列长度，W 指窗口长度）后，平均疏水性值不真实地趋近于０．现在的方法与 Kyte&Doolittle 的方法更类似.

注：我不具备对疏水性分析的预测能力评论的资格，BioEdit 也不对疏水蛋白或跨膜蛋白做任何推论，此图由用户自己判断解释.

ＲＥＦＥＲＥＮＣＥ

以下是对＂bacterio.gb”样本文件绘图（参见英文）．是由古细菌紫红质蛋白（Archaeal bacteriorhodopsin protein）的联配产生．细菌紫红质与视紫红质类似，是和膜结合的光能传递质子泵，具有几个跨膜区.

1 细菌紫红质蛋白的 Kyte&Doolittle 平均疏水性轮廓 ，窗口值＝９
2 ８个不联配的细菌紫红质蛋白的 Kyte&Doolittle 平均疏水性轮廓 ，窗口值＝９
3 ８个联配的细菌紫红质蛋白的 Kyte&Doolittle 平均疏水性轮廓 ，窗口值＝９
4 ８个联配的细菌紫红质蛋白的 Kyte&Doolittle 瞬间疏水性轮廓 ，窗口值＝９
5 ８个联配的细菌紫红质蛋白的 Kyte&Doolittle 平均瞬间疏水性轮廓 ，窗口值＝９

相同性矩阵（**identity matrix**）

相同性矩阵表示在目前联配情况下，矩阵所有序列中两条序列之间相同碱基的比例。输出是一个二维矩阵表，它可以”tab-delimited”或“comma-delimited”(*.csv)作为保存类型。输出完全取决于矩阵的质量。序列在分析时不会自动联配，所以要先联配，BioEdit 以 ClustalW 作为联配工具。

要进行相同性分析，首先选定序列（2 个序列以上，不一定将联配中所有序列都包括）若不选定则自动分析整个联配，然后从“Aligment”->”Sequence Identity Matrix”。

注：序列以前五个字符作为标题。以下是细菌 Rnase P 蛋白的部分结果。

每对序列分值从下列方法产生：

1 同时在各个位置比较各对序列;
2 空格及占位的字符”-“;”~”,”.”,”*”都被当做空格;
3 如果每对序列中相同位置都有空格,则不会对分值有影响(它们不是相同性，只不过是不存在);
4 一条序列是残基,一条是空格的位置被当做一个错配;
5 把两条序列中的共有空格扣除后,分值指相同的残基和两序列中最长的残基的比值。

以上的方法只要联配准确,结果是可以保证的。联配后,其它的序列将以序列中最长的一条为标准,在末尾补上空格。

根据密码子的使用翻译核苷酸

核苷酸序列可根据三联体密码翻译预测的蛋白序列。从”Sequence”->”Protein”->”Translation”,

选择要按何种读框翻译，例如，以下是一个假设的 Methanobacterium(甲烷细菌)的 ORF(开放阅读框架)。密码子使用表（参见英文）：每个密码子按左边->顶部->右边读出核苷酸,每栏包括(1)此密码子在序列中出现的次数,(2)某生物体内此密码子的使用偏爱 (在编码相同

氨基酸的密码子中的比例) (3) 以 3 个字母表示的氨基酸。此表迟早会有用的，如在表达重组蛋白时。

你可以翻译部分序列,也可以使用单个字母表示的氨基酸。

核苷酸位置记数

此项列出每个位置各种核苷酸出现的次数，如果使用屏蔽(Mask), 只有屏蔽的位置列出. 如使用屏蔽位置编号,则根据屏蔽的位置为每个位置编号。以下是输出示例（请参见英文）

在联配中搜寻保守区

有时，即使序列之间的变化很大时，在几个序列中搜寻保守区是有用的。例如,根据一系列同源序列发现通用的 PCR 引物。BioEdiot 查找的是低平均"熵"的区域.

首先选择你的序列,从"Aligment"->"Find Conserved Region".以下是一个例子。（请参见英文）

对话框中各选项的内容：

"Donnot allow gaps"：保守序列中任一区域都不带空位

"Limit gaps in any segment to x"： 保守区中允许有不超过 x 个空位

"Limit max contiguous gaps to x"：保守区中不论总空位多少,每行中不能超过 x 个

"Minimum length"：保守区中不论空位多少,它所包含的实际残基数最小值(不包含空位)

"Max entropy average entropy"： 最大平均熵值(Hx/n,n 指片段的残基数)

"Max entropy per position"： 每个位置的最大熵值,可大于或小于最大平均熵值.

"Allow x exception"： 如果选择,在每个片段中每个位置的最大熵值允许 x 个例外

报告单：

可选文本（text）或一系列联配（Fasta 格式）文件两种．（或全选）

由于 BioEdit 只允许同时打开２０个联配文件，所以最好先使用文本搜索，确定适当严紧参数以使得搜索出很少的区域．例如，以下是７５条 methanogenic Archaea 16S 核糖体序列。（请参见英文）请将第一个区域与在联配窗口下，以信息为基础的示例比较．

最下面是在较宽松的参数下同一个例子的搜索的结果，它发现了更多的区域．

两条序列的打点图：

打点图在每个位置比较两条序列。最简单的形式是一条序列沿 x 轴，另一条沿 y 轴。在行和列的序列中具有相同残基的交叉处打点。

BioEdit 使用用户定义的窗口，匹配沿对角线列出（表示从位置 x,y 开始有不间断的匹配，长度和设定的窗口长度相同）。首先选定序列再从 "Sequence"->"Dot Plot"进入选项菜单将出现以下的对话框。（参见英文）BioEdit 在每个扫描窗口的 x,y 中心点上打点。（例如，两条序列完全匹配，每条有 100 个残基，窗口值 20，则在对角线上出现连续的打点，，但启始于 x,y=10，10，结束于 x,y=90,90. 因为每个窗口的中心点要被打点。）

选项 "Do full shaded alignment"表示将列出窗口值以下的匹配数目并在窗口中心打点，点的明暗度与匹配数和窗口长度的比值成正比。而传统的做法只是简单地根据错配的阈值打点或不打点。

如果仅需要传统的作图，那么就不选择 "Do full shaded aligment",并指明错配的阈值。

选择 "Count similar amino acids as well as identities",可以把 "相似" 的残基当作匹配。这项只有在比较蛋白序列时可行。

选择 "Save matrix output as.." 并指明文件名可以保存矩阵的数据。

BioEdit 产生文本格式的矩阵文件，它来自矩阵作图。在作图中由于每个数据点至少 1 个像素，所以此项打点功能不适合长的序列(<1500-2000 残基)。

最佳比对序列联配（**Optimal Pairwise Sequence Alignment**）

在 BioEdit 的联配文件中可直接选择简单的最佳比对序列联配。使用的是 Smith and Waterman 算法，实际是由 Gotoh 修正的，和 Meyers and Miller 版本类似的 Smith and Waterman 算法，（它本身也是原始的 Needlman-Wunsch 算法的发展）。此算法中，在所有的联配矩阵中保存一个指针，以对最佳联配回溯。其基本的算法是：

Si,j=MAX{Pi,j　Si-1,j-1+sub(ai,bj)　Qi,j}

Pi,j=MAX{Si-1+wl　Pi-1,j+}

Qi,j+MAX{Si,j-1+wl　Qi,j-1+v}

其中，i,j 指矩阵 a*b 中的行和列；序列 a 垂直书写，序列 b 水平书写。Sub(ai,bj)是由打分矩阵而来的，a 中的 i 残基和 b 中的 j 残基的匹配值。Wl: 是启始一个空位（gap）的罚分。V 是延伸一个 gap 的罚分。Si,j 指在 S 矩阵中位置（i,j）的总联配分，矩阵 S 包括了所有可能的联配排列的总分，Pi,j 表示在 P 矩阵中的 i,j 的分值。矩阵 P 包括了所有可能的联配位置的分值，此分值取下面两值中的较大的，即上一行位置 j 中的总分值加上启始空位罚分或上一行位置 j 的 P 矩阵分值加上空位延伸罚分。Qi,j 表示在 Q 矩阵中的 i,j 的分值。矩阵 Q 包括了所有可能的联配排列的总分，此分值取下面两值中的较大的，即上一列位置 i 中的总分值加上启始空位罚分或上一列位置 i 的 Q 矩阵分值加上空位延伸罚分。矩阵 P，Q 使延伸空位或启始空位的罚分不同。（可以想象某序列出现删除或插入整个一段区域（可能是蛋白编码区），但代表的进化事件只有一个。在此种情况下，**a constant gap penalty is not likely to perform as well as an affine gap system.**大概是指此时连续的罚分不大可能发生。译者注。）P，Q 矩阵检查在每个通过联配的路径上，是启始一个空位或延伸一系列空位还是继续下一对残基的联配，可以提高总分值。

一个实际的联配通过以下步骤产生：

（1） 为序列 a,b,计算 S，P，Q 矩阵；

（2） 每次向矩阵 S 填入一个值，就保留此值在矩阵中对应的“格子”（cell）的指针，如果下一个 S 来自配对的残基，指针指向 i-1,j-1,如果来自 Pi,j 指针指向 i-1,j;来自 Qi,j 指针指向 i,j-1；

（3） 在所有可能的路径结束后，最佳联配可由指针回溯而构建。

选取不同的矩阵对联配影响很大。对于两个进化距离相近的序列，最好使用反映相近序列的矩阵（如 PAMn 矩阵系列中，n 的值较低的矩阵，PAM120，PAM80）如果距离较远，使用反映趋异大的矩阵，PAM250。例如，以下序列 A=TETSEFLY；B=TESTSEQ. 选择启始空位罚分-8，延伸空位-2，使用 BLOSUM62 联配，结果和使用 PAM80 是不同的。以下是得到的矩阵（参见英文），回溯后得到联配 TETSEFLY/TESTSE-Q。此结果和手工联配结果不同。由于空位延伸和启始的罚分设置，此联配在比较启始 1 个空位和延伸 5 个空位时，更采用后者。但如果使用 PAM80 则结果是 TE-TSEFLY/TESTSE—Q。这和你期望的相近。以下是产生的联配矩阵（参见英文），回溯指针即可得到联配。

可能需要在初始设置 i=0 对所有的 j 或 j=0 对所有的 i,计算分值时，启始或延伸空位都不罚分，这样一条序列可沿另一条序列滑动，直至找到一个确切的配对，从而真正开始联配。选择“Sequence->Pairwise alignment ->alignment two sequences（allow ends to slide）”即可达到此目的。而与之相对，也可选择”alignment two sequences(optimal GLOBAL alignment)”，此时在序列末端的空位将象内部空位一样罚分。

以下列出 BLOSUM62 和 PAM80 加以比较，注意，错配的残基在 PAM 中将比在 BLOSUM62 中更趋向负值。

REFERENCE

最佳比对联配的优先选项

从“Option”->”Preference”->”Pairwise Alignment” 为配对设置空位罚分参数并为核酸序列设置匹配和错配的参数。参见”Optimal Pairwise Sequence Alignment”

成对联配中的取代矩阵和在联配中设置阴影（**substitution matrices used for pairwise alignment and alignment shading**）

当试图在两条或多条假设同源的（即来自一个共同的祖先序列）序列之间构造最可能的联配时，需要设置 2 个联配残基的“相似性”标准，以评估它们对于总联配的贡献。虽然相似的水平在字面上是没有意义的，（它们或完全相同或完全不同；或者在各自序列中的位置与祖先序列的关系完全相同或完全不同）我们既没有祖先序列，也没有导致现在序列的步骤。我们需要评价体系以确定一个残基通过自然选择被另一残基取代的可能性（残基 a 被残基 b 取代的频率）一个所有残基两两比较的相似性水平称为取代矩阵。此打分矩阵的产生及使用方法请参见 REFERENCE

BIOEDIT 提供了一些常用的取代矩阵，它们可用于最佳配对联配或为相似氨基酸设置阴影，均可以图示或文本的格式表示。

以下是在 BIOEDIT 的 apps 文件夹中以文本格式存在的矩阵

BLOSUM62：由 Henikoff 建立，一般认为在对库搜索中较 PAM 更敏感，是 NCBI BLAST 默认的矩阵。

PAM40，PAM80，PAM120，PAM250：由 PAM 程序产生的矩阵。

PAM（point accepted mutation）”默认点突变“：指平均 100 个氨基酸中产生 1 个取代的变异单位或进化“时间”。大的 PAM 表示大的进化趋异性。PAM 由密切联系的氨基酸序列联配后将数据外推到 nPAM 以反映 PAMn.(如 PAM120 矩阵表示在 120PAM 单位的进化距离下，所有残基之间组合预期的取代频率)。矩阵中数字是按 log odds 处理，即每个取代频率和随机条件下出现此残基的概率的比值，然后取自然对数。表示为 $\log q^n_{a,b}/p_a,p_b$; $q^n_{a,b}$ 指在 n 个进化单位下，残基 a 取代 b 的频率，p_a,p_b 指找到残基 a,b 的各自概率。PAM250 表示当每 100 个残基发生 250 个取代时，一个氨基酸发生取代的频率与找到这两个氨基酸的随机概率的关系。

DAYHOFF：BIOEDIT 中的此矩阵是最初的 PAM250 矩阵数值上四舍五入取整后产生的。对于最初的 PAM250 可参考网址及 REFERENCE（请见英文）

此矩阵仅适用某人对使用它做 BLAST 有兴趣时。它已经基本上由从更新库中产生数据的 PAM 矩阵替代。但是对库搜索 BLOSUM 似更好。

IDENTITY 和 **MATCH**：这些矩阵适用于建立在相同性基础上的联配或设置阴影。对于设置

阴影它们是相同的。对于联配，IDENTITY 在所有情况下，错配分值是-1000，匹配分值是1，这可专门用于搜寻一串完全匹配残基。对于最佳联配，它允许序列的末端“滑动”，这样可以发现两条虽不完整（一条或两条序列在一端缺失了部分）但完全相同的序列。如果在局域 BLAST 中采用，将只选择完全配对而不包括内部的错配。MATCH 中所有情况下配对分是 1，错配分是-1，在仅依赖相同性的基础上，可在 BLAST 中用来搜索氨基酸序列。但不一定决对不出现内部的错配。

GONNET：另一个 PAM250 矩阵。

REFERENCE

产生一致的序列（**consensus sequences**）

从”Options->””Preference”->”Consensus”下设置参数产生一致的序列。可选择在一致序列中是否存在空位，如选择那么就会在一致序列中出现，如不选，不会出现。但仍计入分值中。但如果不将空位当作有效字符（参见“valid character vs non-residue characters”一节），它们也不会被计入。所有的字符分为有效字符和非残基的字符，所以如果要所有的类似“-,~,.”被识别，它们必须包括在氨基酸和核苷酸的有效字符集中。

RNA 的比较分析：

系统发育比较分析的基础：

RNA 的结构定义为核苷酸的碱基的相互作用。最简单情况下，即螺旋中的碱基对之间的 Waltson-Crick 碱基配对。RNA 结构的系统发育比较分析方法建立在如下假定上，即在进化中核苷酸改变，但重要的 RNA 二级和三级结构保持不变。一个可能破坏结构的碱基变化可以由序列中另一处的变化补偿以保持结构稳定。所以不同物种的同源 RNA 中将包含“补偿碱基变化“或”共变化，协变（covariation）“。所以通过检查来自各个不同生物的同源 RNA，确定这些“补偿碱基变化“，从而阐明结构。例如，一给定的序列，GAAGA 将可能与序列中任一 UCUUC 配对，而后者可能在序列中出现数次。如何确定到底是和哪一个配对呢？可以检查不同生物的同源 RNA 序列，找出”补偿碱基变化“。例子（参见英文）

在此例中，只有最后一个 UCUUC 才可和 GAAGA 配对。象这样在序列中 2 个位置出现“补偿碱基变化“，被认为是螺旋存在的证据。两条序列不能形成互补表明不存在配对。在”系统发育比较分析“中关键是序列联配，同源序列必须适当联配。此处同源性是严格意义的：同源的核苷酸来自一个共同的祖先。所以开始时，先使用关系紧密的序列进行联配，这样在序列相似性基础上联配，不需要加入许多联配的空位。联配后互补序列的”协变“可被立即发现，从而开始构建二级结构。然后差异大的序列可以添进联配中。这样持续添加新序列，进行”协变“分析，直到联配和二级结构模型出现。此过程的完全描述可参见”More Information”.一旦一个完整的二级结构模型形成，“协变”分析可以鉴定非螺旋区的核苷酸之间的相互作用以及不规则的相互作用。之所以可以被鉴定是因为涉及的核苷酸即使不形成规则的碱基配对或是一个螺旋的一部分，也仍一致的变化。

使用屏蔽（**Masking**）

屏蔽指仅采用联配中部分区域进行分析而排除其他。例如，如果你有一个长的 RNA 序列，你希望仅比较分析一个小的区域，从而发现局域二级结构，这需要排除其他区域的数据。如

果在使用“协变分析”（covariation）,”潜在配对“(potential pairwise),”相互信息分析”（mutual information）之前使用 masking，这样程序结果将只分析指明的区域。有时当针对某一 RNA 二级结构成分分析时，一个来自其他物种的标准的编号系统已经存在了。（如，E.coli 的 Rnase P 的 RNA 常用来为所有细菌的 Rnase P 的 RNA 编号）此时序列可设为编号屏蔽（numbering mask）根据它为碱基位置编号。通常，编号屏蔽和序列屏蔽相同。

对于任何屏蔽，“-”，“~”，“.”等空位指示符代表在分析时位置不计入在内，而其他的字符等指明计入在内。

在 BioEdit 产生屏蔽后，“*“表示此位置计入，”—“表示此位置排除。如———***——，表明前三个不计入，后三个计入。

序列屏蔽和编号屏蔽可同时采用，但每一个都不是必须的。

如果选择了编号屏蔽，则首先要指明被编号的屏蔽区域，从“ Sequence”->”Set as Sequence Mask”或” Set as Numbering Mask”分别设置序列或编号屏蔽。

从“Sequence”->”Creat New Mask”下，出现一串星号，用鼠标固某区域，从” Sequence”下“Toggle Mask”对位置属性更改。（计入或排除）

共变化（**Covariation**）

共变化指序列中两个残基步调一致地变化。严格地讲，即每当联配序列中 x 变化时，y 也变化，两者是一致的。（例如，当 x 变为 A，y 变为 T。.每次 x 变为 A，y 一定变为 T）。残基间的共变化表明，它们之间一定有重要的相互作用，当重要结构残基突变时，自然选择保留了那些有补偿突变的序列。

共变化的例子

假设我们现有一个联配序列，它表示了几种物种共有的一个特定的 RNA 的保守的结构。我们希望从联配中包含的信息推测出 RNA 二级结构。下面是一个联配的例子（见英文）

在上述联配中共有 3 对“共变化”的位置点：2/28，5/25，9/21。两个碱基共变表明它们很可能相互作用，如果一个突变发生在与其他碱基有重要作用的碱基上（常是碱基对），选择压力可能会只保留在另一处碱基上发生补偿突变的碱基。事实上，上述的碱基共变化都发生在规则的碱基对（Watson-Crick 碱基对或在 RNA 中 G-U）表明它们可能是碱基配对。共变化碱基对 2/5 分别和 5/25 的距离相同，而 5/25 分别和 9/21 的距离也相同，而且界于它们之间的碱基也可形成碱基互补，这都表明联配序列的两端可能闭合形成螺旋，如下是“Sample1”形成的结构：

螺旋中其他的碱基是不变的，虽然 RNA 比较分析不能提供碱基相互作用的直接证据，但它和潜在配对分析联合起来进行分析，表明残基可能存在于一个螺旋配对中。

在 **BioEdit** 中使用共变化分析

BioEdit 提供两个基本的输出格式：清单格式（list format）和列表格式（ table format）均可点击查看描述文件及示例。两者都是纯文本。列表格式可按“tab delimited”和“comma

delimited “（*.csv）两种保存类型保存。前者在文本编辑器（text editot）中观看最好，后者可方便的被输入制表软件中 MicroExcel(一般也可读出 tab delimited 文件)。文件可按 PC 或 Macintosh 格式书写。

进行共变化分析时：

1 设置选项（文件格式，输出格式，你也可同时选择清单和列表两种）

2 如果只分析序列中一部分就创建一个屏蔽（mask）,(或者把一个已存在的序列设为 mask), 如果你只分析序列中一部分并且希望联配中位置编号和一个标准序列中的编号相应，（此标准序列中的编号必须包含在联配中）就把序列设置为编号屏蔽（numbering mask）

3 选择所有你要分析的序列。只有选择的序列才可被分析。如果你指明的 mask 并不是一个确切序列，你需要从序列中将其排除。如果不选定序列，则默认文档中所有的序列都将自动被选中。

4 运行共分析，从“RNA“菜单下，选择”Covariation”。将提示你添入文件名称。如果你选择清单格式，BioEdit 将在文本编辑器中打开文件。

注意：1 清单格式的文件可能会很长。每列以字符串的形式打印出来。如果两列共变化，则打印出来时，两列上下并排。你还可以在选项中指明显示“位置”（position）。. 同时可选择是否显示不变化的位置。

2 列表格式的文件在屏幕上观看很方便，但打印出来很麻烦，尤其是序列较大时。

以列表格式输出

共变化表是一个二维的各联配位置的矩阵（每个位置和所有其他的位置比较）。当 2 个位置共变化，“5”标在两个位置的交叉处；如都未改变，标“1”；如既不是共变，也不是未改变，标“0”。以下是一个联配的例子。（参见英文）

以下是共变化数据的输出格式（更大的列表可在字处理程序中使用，也可在编辑器中如 wordpad 中观看）（参见英文）

有时查看一下和共变化位置同行的未改变位置是否可形成配对也是有用的。在潜在配对（potential pairing）分析中也可获取此类信息。也可产生以“comma delimited “(*.cvs)保存类型的文件，它在制表软件 Excel /Quattro Pro 中较容易打开。

以清单格式输出

在本节第一个例子中的格式即是清单格式的输出，有以下选项：

Show nucleotides：以一串核苷酸碱基形式报告联配中的每列。

Show position only：只显示共变化的碱基位置。当你希望产生的文件小些或仅在屏幕上观看结果时，此选项很有用。以下是上一个例子选择此项的输出：（参见英文）

共变化分析的优先选项（**Convariation analysis preferences**）

从 “Option”->”Preference”进入选择菜单。可同时选择列表格式和清单格式。如使用”mask numbering”,结果报告中的位置对应的是选定的 numbering mask(编号屏蔽)中的实际编号。只能在清单格式中（list）显示核苷酸和位置的比较。

共变化的算法

共变化的算法很简单。在联配序列中，只要两个位置至少发生了改变（保证它们不是未改变的）而且改变遵循相同的变化格式（保证是一致变化的），它们即称为共变化。算法如下：

1 联配序列被分为垂直的列（联配序列实际上是一个二维的（行和列）字符矩阵）

2 把每列用下面方法转换为一串数字：

a 每列的第一个残基定为 1

b 如果每列的第二个残基与第一个相同，也定为 1，否则为 2

c 同样依此规则检查每个残基，如果它是在此列中第一个出现的，就分配一个新的数字，否则就采用之前此相同残基分配的数字

d 上述完成后，每列中未改变的位置就是一串 1， 共变化的位置就会有相同的数字模式。

以下是例子及转换后的结果。（参见英文）

可见位置 1 是位改变的，所以是一串 1 代表；位置 2，3 是共变的，所以代表的数字相同；位置 4 和 1，2，3 的变化都不同。只要比较转换后的数字是否完全匹配就可以很容易的完成此算法。但它不允许出现例外（如，原来的 A-T 配对可能在一序列中转换为 G-C,而另一序列中为 G-U,这在共变化分析中将被忽略）另一方面，在大的序列中它并不能推测什么样相互作用可能发生并挑出必需的三级结构中的相互作用。

REFERENCE

潜在配对分析（**potential pairing**）

当 RNA 分子中两个核苷酸之间存在配对碱基的相互作用力。一个碱基发生突变，另一个碱基为了补偿这一突变可能不仅仅是某一特定核苷酸突变（例如，原来的 A-T 配对可能在一序列中转换为 G-C,而另一序列中为 G-U,） 这在共变化分析中将被忽略，因为此种改变并不遵循完全相同的模式。要鉴定这种情况，可以在潜在配对中选定碱基配对的规则（在选项中设定）。

BioEdit 中并不要求有位置变化，所以未改变的位置上只要可以形成碱基对，也能被发现。同时也可在 “preference”中设置以滤出未改变的位置之间的碱基配对。以下是一个联配序列，它和在共变化分析中使用的相同。设置允许 A-U/G-C/G-U 碱基配对规则以及 1 个错配，产生下列的结果（以清单格式，滤除了未变化位置的潜在配对）比较这一结果和共变化的结果，发现位置 3/27 有一潜在的配对，而共变化的结果未检出。潜在配对的数据也可以按允许的配对出现的频率或原始允许配对的数目列出一个（二维矩阵）表。

使用潜在配对（**Using Potential Pairings in BioEdit**）

BioEdit 提供两个基本的输出格式：清单格式（list format）和列表格式（ table format）均可点击查看描述文件及示例。两者都是纯文本。列表格式可以“tab delimited”和“comma delimited “（*.csv）两种保存类型保存。前者在文本编辑器（text editot）中观看最好，后者可方便的被输入进制表软件中 MicroExcel(一般也可读出 tab delimited 文件)。文件可以 PC 或 Macintosh 格式书写。

1 设置选项（文件格式，输出格式，你也可同时选择清单和列表两种）

2 设置允许的配对类型。A-T/G-C/G-U 是缺省值。但如果更改选项并保留后，下一次自动以上一次更改后的情况为缺省值。

3 如果只分析序列中一部分就创建一个屏蔽（mask）,(或者把一个已存在的序列设为 mask), 如果你只分析序列中一部分并且希望联配中位置编号和一个标准序列中的编号相应，（此准序列中的编号必须包含在联配中）就把序列设置为编号屏蔽（numbering mask）

4 选择所有你要分析的序列。只有选择的序列才可被分析。如果你指明的 mask 并不是一个确切序列，你需要从序列中将其排除。如果不选定序列，则默认文档中所有的序列都将自动被选中。

5 运行共分析，从“RNA“菜单下，选择”Potential Pairings”。将提示你添入文件名称。如果你选择清单格式，BioEdit 将在文本编辑器中打开文件。

注意：1 清单格式的文件可能会很长。每列以字符串的形式打印出来。如果两列共变化，则打印出来时，两列上下并排。你还可以在选项中指明显示“位置”（position）.同时可选择是否显示不变化的位置。

2 列表格式的文件在屏幕上观看很方便，但打印出来很麻烦，尤其是序列较大时。

以清单格式输出（**List Output**）

在上一节中的例子即是清单格式的输出。如果你不想列出核苷酸，也不想列出未配对的位置，选择“Position only”。下面的结果（参见英文）就是选择了此项的输出，同时滤除了未改变位置的配对。

以列表格式输出（**Table Output**）

潜在配对的数据可以按二维矩阵列表格式输出。在两个位置的交叉处列出允许配对的数目或允许配对的频率。格式同共变化分析中的列表格式一样。

潜在配对分析的优先选项（**Potential pairings analysis preferences**）

选中选项框以确认碱基配对规则，一般缺省值是在 RNA 螺旋中常见的 A-T/G-C/G-U，建议也选择“gap-gap”配对，因为 gap 代表了这个和其他序列同源位置的丢失，但这不意味着在其他存在此位置的序列中不会存在确定的结构。如果不允许“gap-gap”这些位置将被当作错配。象共变化分析一样，位置编号既可用整个联配的位置编号，也可指明一专门的编号屏蔽（numbering mask）,设置好选项后，按”Save Preference”后，设置值将成为缺省值。选中“Numberical Table”后，选择”Integer”将输出联配位置潜在配对的原始数目；选择

“Frequency”将输出配对位置的潜在配对频率（配对数/总序列数）。

潜在配对的算法

潜在配对的算法直接明了。关键是按照设定的允许配对一一搜寻位置之间的配对。BioEdit 给每个核苷酸分配一个整数值，这些整数任何两个的和都是独一无二的(包括每个整数与自身的和)（参见英文例子）。

再建立一个有 28 个数据点的排列，每个数据点可取值为 1 或 0，取值为 1 时，配对允许；取值为 0 时，配对不允许。这样上面的数值可看作此排列的一个“目录”（index）。当在联配中扫描每对位置时，此排列“目录”中的数据（代表两个残基数值的总和）求和，得到一个每对位置的总和。（如果配对则加 1，不配对则加 0）如果此总和大于或等于某一要求的值，（取决于允许错配的情况），输出中即会报告存在潜在配对。（对于列表格式的输出，将会报告所有序列中各个位置的情况）

交互信息分析（**Mutual Information Analysis**）

在使用交互信息之前，请参阅相关文献。（见英文）

概述

交互信息，象在系统发育比较分析中的应用一样，主要是衡量在一个适当联配中，两个位置共有信息的信息量。符号是 M(x,y)(位置 x,y 的相互信息)。M(x,y)表明两个位置相关的紧密程度。此相关程度显示了两位置的直接相互作用，如碱基配对。BioEdit 另外计算 R1 和 R2 两个参数，它们分别表示位置 x,y 对 M(x,y)的贡献。后面将详细谈到。

什么是交互信息

交互信息分析是以下思想的拓展，即对某个特定位置的不确定性表示是信息含量的下降。在预先对某位置一无所知的情况下（如 RNA 中核苷酸），不确定性最大。但当你一旦确定了某位置是什么核苷酸时，不确定性消除了，此位置的信息量达到最大。现在考虑有多条序列，在某位置均含有一个同源核苷酸。知道第一条序列上此位置上的核苷酸并不能为确定第二条或随机的一条序列中此序列的核苷酸提供多少信息。但是如果已知此位置在许多乃至几乎所有序列中均为某一特定碱基（如 C），而不是其它的碱基（如 G），则我们积累了相当多的“信息”，可预测另一个未检测的序列中，在此位置某核苷酸出现的可能性。即在另一未检测的序列中，此位置核苷酸的不确定性下降了。这可作为序列的标识（在 BioEdit 中做的是熵图）。

交互信息进一步拓展了这一思想，对配对位置的信息量进行检查，此信息量依赖于并联系每个位置单独的信息量，但不能将两者混淆。总的讲，它衡量不确定性的下降，此不确定性指两种事物相互影响相互作用的程度。Robin Gutell 发展了用交互信息预测 RNA 结构的方法，也很适合系统发育比较分析，因为两个位置交互信息高也提示这 2 个残基直接相互作用。例如，（见英文）

总共 8 个序列，其中位置 1，4 是不改变的。信息量最大。位置 2，3 中 C/G/U/A 各出现了 2 次，信息量为 0，我们无法预测下一个序列中这 2 个位置的核苷酸。但序列中，“共享”（shared）信息却是不同的。交互信息指我们对于某位置某核苷酸的出现是如何影响另一

位置核苷酸的不确定性的下降。虽然对于位置 1，4，我们对出现什么核苷酸有很大把握，但我们却不了解它们是如何相互影响的。（因为它们永不改变，也不能用变化来检测）它们之间的交互信息是 0。而另一方面，虽然对于位置 2，3，我们不知道各自独立的信息，但它们都含有它们之间是如何影响彼此的共有信息。我并不能猜出新序列中位置 2 的核苷酸，但如果告诉我位置 3 是 C，我会强烈感到位置 2 是 G，这即建立在“交互信息分析”（它们遵循共同的配对模式）交互信息也表明这些碱基可能相互作用。（它们特殊的核苷酸类型进一步显示它们可能碱基配对）。

交互信息分析用数学形式衡量每对位置的共变化，但和分析共变化不同，它是定量地衡量共变化的程度。在交互信息的数学概述中将进一步介绍。

交互信息的数学概述

交互信息指两个相互作用位置共有的信息量或者根据一位置的信息量可降低另一位置的不确定性。在系统发育比较分析中，指联配中两个位置相互依赖的程度，1992 年，Robin Gotell 把交互信息引入 RNA 结构预测中。如果两个位置（x,y）有强烈的相互作用，交互信息 M(x,y)相对也较高；如果 x,y 不相关，M(x,y)也较低，为 0。M(x,y)定义为（在 nits(尼特，与比特相应)中）M(x,y)=∑(fbxby)ln(fbxby/fbxfby),其中 bx,by 指 x,y 位置的碱基（A/G/C/U/GAP，歧义的碱基被忽略）；fbx,fby 指 x,y 位置的出现频率；fbxby 指在 x,y 位置出现的碱基对频率。当碱基未改变时，交互信息是 0，无法发现相互作用（但这并不表明不存在）例如，如果 x 恒为 A 时，除了 bx=A 时，对于所有的组合 fbxby=0。当 bx=A,fbxfby=fby,fbxby=fby,所以 fbxby/fbxfby=1,ln(fbxby/fbxfby)=0.M(x,y)=0。当两个位置变化最大时，对于所有的 b,fbx=fby=1/n,n 是可选的碱基（在 BioEdit 中为 5 个，gap 当作 1 个）对于所有的 bx，by,有 0<=fbxby<=1/n.（因为 x,y 位置上的碱基组合出现的频率不可能超过单个碱基出现的频率）。fbxby=(1/n)^2, 所以（fbxby）log2(fbxby/fbxfby)=1/n^2(ln((1/n^2)/1/n^2))=1/n^2(ln(1))=0。即当 x,y 位置上碱基改变是相互独立时，M(x,y)=0。当 x,y 位置上碱基改变是相互联系时，fbxby=1/n(可能配对时)，其它情况下为 0。所以 fbxbyln(fbxby/fbxfby)=0, 只有当碱基完全联系时即 fbxby=1/n,(fbxby)ln(fbxby/fbxfby)=1/n(ln(n))=(ln(n))/n.所以 M(x,y)max=n(ln(n))/n=ln(n).如果 n=5 时，M(x,y)max=ln(5)=1.609.

BioEdit 使用 Gutell 的方法，M(x,y) =∑(fbxby)ln(fbxby/fbxfby)=H(x)+H(y)-H(x,y),H(x),H(y)分别指 x,y 的位置的熵。H(x)=-∑fbxln(fbx); H(x,y)=-∑fbxbyln(fbxby).代入计算可得到同样的结果。BioEdit 为方便起见使用自然对数，若采用以 2 为底数，信息量以“比特”（bit）计算，但数据间的关系是同样的。

由于 M(x,y)衡量两个位置的相互依赖性且同样的依赖于每个位置上碱基出现的频率，所以它是个对称等式，即 M(x,y)=M(y,x)。在有些情况下，两个位置虽有相互依赖性，但一些其它因素限制了某个位置（改变），某位置上改变过小，造成 M(x,y)太小以至共变化丢失。此时，另两个参数 R1（x）,R2(x)可解决此问题。R1/R2 指 x,y 位置上交互信息和熵的比值。R1（x）=M(x,y)/H(x), R1（x）=M(x,y)/H(y),(偶尔 R2（x）=R1(y).如果 x 位置的变化很小，但和位置 y 的变化仍然是相关的。R1(x)将相对较大。R1(x)和 R2(x)一般不相等。

REFERENCE

在 **BioEdit** 中使用交互信息

BioEdit 通过 M(x,y) =H(x)+H(y)-H(x,y)计算交互信息，它可很好的衡量在进化保守分子中两个碱基的相互依赖性。而此种相互依赖性可通过系统发育比较分析而帮助建立及精细化 RNA 分子的二级和三级结构。

在进行交互信息分析之前，象所有的序列比较分析一样，一个高质量的联配是决对必须的。因为如果碱基不是排列在它们的同源位置上，由此产生的数据和结构是不可能正确的。

同样要先选择输出的类型（参阅设置交互信息选项）

BioEdit 提供以下输出格式：

1 列表格式（tabular output）(矩阵格式)

a M(x,y)：全部列表格式（full table）或者仅列出对角线以上的部分（因为 M(x,y)对称）

b R1(x): 仅全部列表

c R2(x): 仅全部列表

各项可单独或全部选择，如果选择超过一个，将只产生一个列表，但是将同时列出各项的计算值。例如，如果同时选择了 M(x,y)和 R1(x)，R1(x) 将出现在 M(x,y)的计算值之下。

当 x=y 时，可选择 M(x,y)=0(实际值是当 x=y 时，M(x,y)=H(x))。此时在 M(x,y)矩阵的打点图中，x=y 处对角线将被隐藏。输出可是“comma delimited”或” tab delimited”，如果为了观看文本文件，推荐使用”tab delimited”，对于较小的列表，可直接使用 BioEdit Rich Text Editor(文本文档编辑器)。外部程序如 Excel 也可调用，只要此外部程序可识别文件格式且接受命令行的参数设置，即可在其中打开列表格式。

2 清单格式（list outputs）

a Pbest:：Pbest 列出所有用户指明的范围内最高的 M(x,y),R1(x),R2(x)的值，报告中含有所有这三个数据，但（各数据）报告和排列的阈值将会根据选项中的声明。Pbest 以一定百分比输出各个位置最高分值，或一定百分比的整个分析的最高分值。Pbest 可声明 0—50%。

B Nbest： Nbest 列出 N 个各个位置的或整个序列的最高分值（用户选择），象 Pbest, 报告中含有所有三个数据，但（各数据）报告和排列的阈值将会根据选项中的声明。

对于各种类型的分析，位置的编号可根据联配编号，即联配窗口中位置编号，或屏蔽序列（mask）的真实位置序列。（但只有此序列包含的残基才计算）。

文件可按纯文本在 PC/Macintosh 格式输出（两者只有返回值不同）

例如，你希望在 Macintosh 下使用 SpyGlass 程序观看数据，M(x,y)矩阵文件即可在 Macintosh 下输出，否则要用字处理程序转换。

按以下操作，在 BioEdit 联配稳定窗口下，进行交互信息分析：

1. 在联配文档窗口打开联配文件。

2. 如果使用 mask,可以新建一个或指明已存在的一个序列作为 mask.

3. 选择输出优先选项，Option->Aligment->Mutual Information.

4. 选择要分析的序列（通过选择序列的标题（title）），只有当序列的标题被选中时，才能被包含在分析中。如果不选择，则 BioEdit 自动选中所有的序列。如果你使用的 mask 并不是序列，而只是用来指明位置，那么一定要在分析时将它排除在外。从“Edit”菜单下，进入” Select all Sequence”,然后 Ctrl+鼠标点击不选择的序列，可以将不需要的序列排除在外。

5. 从“RNA”->”Mutual Information”，提示你为每个输出文件命名。清单格式的输出将在文本编辑器中列出，但矩阵（列表）格式的输出不会。需要在优先选项中设置在制表软件或其他外部程序中打开矩阵列表。

   你还可以在 BioEdit 矩阵打点图器（matrix plotter）中观察你的数据。

交互信息示例

以下是分析细菌 RNase P RNA 的部分序列的一个例子。（点击 Aligment 可以观察此联配。设置输出是全部列表（full table）显示 M(x,y)的数值。 Nbest 列出各个位置 5 个高分值。（参阅交互信息优先选项）。序列和编号 mask 都是根据 E.coli.。序列的编号是根据 E.coli 的 mask 序列。此序列中包含了一个 RNase P RNA 结构区域的“cruciform region”(十字型区域)。（见 RNase P RNA 结构）。由于矩阵文件太大，不能在此说明文件中打开观察。但可通过打点作图方便地的观察。在 BioEdit 矩阵作图程序中，数据既可以数字也可以图形的方式被动态的检查。

RNA 的结构示例

以下是 E.coli RNase P RNA 的二级结构模型。（参见英文）

其中交互信息分析的“cruciform region” (十字型区域)在此输出中是环型的。此图象及全部最新的细菌和古细菌的 RNase P RNA 结构和序列均可在 RNase P 数据库中找到。http://jwbrown.mbio.ncsu.edu/RNaseP

联配示例

以下是细菌 RNase P RNA 联配的一部分，共有包含极丰富信息的 138 条序列。序列包含有“cruciform region”(十字型区域)。其中”~~“位置表示保守性较差的区域，它们已被从联配中删除。（参见英文）

N-best 输出示例

以下是使用 E.coli 作为 mask,上面联配示例的 Nbest 输出示例。（参见英文）

对于所有的配对，交互信息是 0 的位置将不再报告，因为最高分值也是 0（例如，位置 1 是不变的）

交互信息的作图示例

以下是上面联配示例的交互信息分析矩阵作图结果。（参见英文）

信息含量高的配对用图解的方式表示。目前 BioEdit 仅在坐标轴中产生矩阵作图，没有注释信息。今后更高的版本可能会包括自动的高分区域注释。在图中的左下方显示出 E.coli 中此高分区域的结构。如图所示，正交于对角线的区域信息含量高，代表两个对比的位置关系密切（它们相互影响），暗示它们可能存在碱基配对。对角线上出现高信息区暗示着可能出现碱基配对的螺旋。如果你观察 E.coli 的 RNase P RNA 结构，你可发现对角线上局部的高信息区和结构中的螺旋明显相对应（在数据和结构中均以 H1~H5 标记）。

参阅“使用矩阵打点作图器分析交互信息数据“以及”一维矩阵中行和列数据的作图“，可以方便地在分析数据，矩阵数据及线形绘图之间来回观看。

设置交互信息优先选项

从“Option”菜单下进入“Preference”—>” Mutual Information”,可选择的项有：

1 Save Table：点击可选择 M(x,y),R1(x,y)和 R2（x,y）三项数据之一或全部选择。选中后另有以下选项。

A link to external programm（连接外部程序）：在分析完成后指明在制表软件，如 Excel,Quatto Pro 等中自动打开你的列表格式输出。它只对列表格式输出有用。清单格式的输出会在 BioEdit 文本编辑器中打开。

B M(x,y),R1(x)和 R2（x）选项框：根据你的需要可选择三项中之一或全部。但是当选择超过一个项时，各项数据并不分列在不同的列表中。所以如果你使用矩阵作图程序或 SpyGlass Transform 等外部程序观看列表时，你必须对每项数据分列出一个列表。如果你选择一个列表中输出多项数据时，各项数据将象在选项框中的顺序从上到下排列。如果你仅需要列出一半的矩阵（因为 M(x,y)是对称的），可选中“calculate above the diagonal only”(只计算对角线以上部分)框。如果你需要用矩阵作图，你必须选中“caculate full table”(计算全部列表)框。选中” calculate R1”和“caculate R2”将自动产生全部列表。

C When x=y： x=y 时，实际取值是 M(x,y)=H(x).如果你想在打点作图程序中观察此数据并在图中禁止出现对角线，可将沿对角线的各项数据设为 0（M(x,y)=0）.

2 Output file for：文件输出可选择 PC 或 Macintosh 格式。这将影响交互信息分析的输出。

3 Aligment Nubering 或 Mask Nubering ：你可能希望在输出中的位置编号对应联配中的位置编号，此时可选择 Aligment Nubering；你也可能希望在输出中的位置编号对应设置的 mask 序列中的编号，例如当你根据一个标准序列的位置分析新序列的结构时，此时你更希望对应的是标准序列位置编号而不是在联配中已剔除了 gap 的位置编号，你可以选择 Mask Nubering。

4 Pbest 和 Nbest 选项：见“在 BioEdit 中使用交互信息“一节。

在确定以上选项后，你可以选择保存后再关闭对话框，此时各选项框将成为新的缺省值；你也可以不保存而直接关闭对话框，此时各选项将仅针对当前执行的程序。

使用矩阵打点作图器分析交互信息数据

从一个联配窗口或主应用窗口下选择“RNA“，再选择”matrix plotter”,再选择“plot a matrix”,可使用矩阵打点作图器分析交互信息数据。BioEdit 将首先检查是否可识别该文件并作图（任何 tab-delimited 类型的文件或对称的矩阵文件都应可以）接着计算矩阵的行和列的数目并出现一个对话框，提示你输入需要作图的行和列的位置（缺省值是全部的行和列）。例如使用 E.coli 作为序列和编号 masks，对 146 条细菌 RNase P RNA 序列进行 M(x,y)分析得到的矩阵进行上述操作，可出现下面的对话框（见英文）：

注：每次作图的行和列的极限是 2000*2000，如果大矩阵超出此范围时，要分成各个部分作图。

以下是对 RNase P RNA 序列各个位置的输出。（见英文）

作图完成后，可从“View”菜单下打开“data examiner”(数据检查器)，用鼠标在图中各点移动观察各数据点。你也可直接用点击某点在顶部的工具条中将出现数据值。目前此窗口下尚不能直接打印，但今后会增加此功能。但图象可直接拷贝到剪贴板上，再粘贴到其他应用程序中，也可直接保存为位图文件（*.bmp）

在 zoom 选项框中选择图象扩大或缩进的比例（25%——800%）

当前选定的数据点可直接使用“一维矩阵中行和列数据的作图“（见下文）

阈值控制（threshold controls）可通过设置数据点的阈值在矩阵作图中遮蔽（shading）某些数据点，当仅需要显示出高分区域时，此选择可能有用。如下面示图的数据和前面的一样，只是把低阈值设为 0.3881，高阈值设为最高值。设置高阈值可使任何超过此分值的数据点涂上淡蓝色。（参见英文）

矩阵中行和列数据的一维作图

在二维矩阵中，如由交互信息产生的矩阵，观察数据并挑出高分值可能会十分乏味且相当困难。所以才产生了打点作图来观察数据的方法，但由于打点数据是通过暗度的强弱（0~255，一个字节的范围）来反映数据差异的，所以精细的区分很难被发现，例如在分析配对的两残基和第三个残基组成的三联体之间的交互信息时。为解决此问题，BioEdit 提供了沿着矩阵的行向或列向作图的选项。如下面的细菌 RNase P RNA 联配的交互信息数据作图（E.coli 作为 mask）。在图中很难挑出某一核苷酸三联体，虽然碱基配对的位置 94 和 104 是明显可见的，但很难挑出配对核苷酸 94-104 及第三个核苷酸 316 组成的三联体。此图是联配的全部 M(x,y)列表格式的部分图示，设置固定数据点大小为 3*3 象素。鼠标箭头指向位置 94-104，右边的小红色框中心是位置 94-316。（图见英文稿）

此图中位置 316 和位置 94，104 的相互作用并不明显。从作图窗口下“Plot”—>”Line Graph of Rows”进入一维行向作图。

注：此选项仅可在打开的矩阵作图中使用。

下图（参见英文）显示第 316 行的作图，用“Row”旁边的上下箭头选择要观察的行或直接在框中输入要 观察的行。可移动图中的蓝十字点击任何位置，将在顶部工具条的左上方列出位置 x,y 及数据值。其中数据值是指图中位置的高峰对应的数值。

从图中可见，位置 316 和 104，316 和 94 之间相对较高的交互信息。图中选择了位置 104，左上方显示出 M(x,y)=0.306.也可选择以列为 X 轴。

目前版本的 BioEdit 在使用比较分析时，允许使用编号 mask,它和序列 mask 是不同的。这样一方面，当仅对分子的一部分分析时，可以参考参照序列或结构的编号（此序列或结构代表全部分子序列或结构），从而方便数据的观察。另一方面，即使使用了序列 mask 时，也可采用实际联配时的编号，这样分析位置数据时可参考实际的联配。正因为如此，最新版本的 BioEdit(v.5.0.0)中，报告的行向和列向编号是在实际的数据文件中的编号。（例如，矩阵中出现的第一行可能是实际的数据文件的第 234 行，而不是实际的第一行；第二行是 300，二不是 2）。详细的信息，参阅“交互信息“和”系统发育比较分析基础“

交互信息检查器（**mutual information examiner**）

如果希望在联配窗口中观察任两个位置的交互信息，从“View”菜单下进入“mutual information examiner”，此做法借鉴了 Dr.J.Brown 的”Covariation”程序。下面是控制条的格式（见英文示例）
在“x”,”y”旁边输入要分析的位置。如果希望某一特定序列上此位置的信息（序列要无 gap），就把此序列设为”numbering mask”,再输入 x.y 的位置。图中的 x,y 位置（x=261,y=289）对应的是前面在矩阵作图时选择的数据点。一定要将需要分析的序列全部选中。如果你希望排除某些序列不分析，可以不选它们或者先全部选定序列，再 Ctrl+鼠标点击不需要的序列而排除它们。最后点击“calculate”,将出现以下窗口。（见英文）。在窗口中点击”Text”,将出现如下文本编辑窗口，内容可复制和粘贴。

如果要同时分析几个位置，可使用逗号“，”和“-”。如下是一个例子。如果采用此位置输入是 X=a-b,Y=c-d, 同时计算几对位置，BioEdit 假设你要分析是螺旋区域，即位置是反向平行的，所以 c-d 和 d-c 是不区分的。如果你要特别地指明某对的顺序，要采用 X=a,b,c,d Y=e,f,g,h.的形式的输出。表中各项可从“Option”—>”Preference”—>”Mutual Information”选择。以下是各项的选项。（参见英文）