

Sequin 帮助文件

生物软件网 (<http://www.bio-soft.net>) 提供

翻译: qshchen@sohu.com

Sequin 帮助文件

目录

- 介绍 (Introduction)
 - 帮助文档 (The help documentation)
 - 表格的组织 (Organization of Forms)
 - Sequin总览 (Overview of Sequin)
- “Sequin欢迎表格” (Welcome to Sequin Form)
- 提交者表格 (Submission Authors Page)
 - 提交页 (Submission Page)
 - 我们是否可以在发表前公布此记录? (May we release this record before publication?)
 - 稿件的临时标题 (Tentative title for manuscript)
 - 联系页 (Contact Page)
 - 作者页 (Authors Page)
 - 机构页 (Affiliation Page)
- “序列格式”表格 (Sequence Format Form)
 - 提交类型Submission type
 - 序列数据格式Sequence data format
 - 提交类别Submission Category
- 物种和测序格式Organism and Sequences Form
 - 物种页Organism Page
 - 物种Organism
 - 测序部位Location of Sequence
 - 翻译遗传密码Genetic Code for Translation:
 - 核酸页Nucleotide Page
 - 分子Molecule
 - 拓扑结构Topology
 - 5'末端未完成/3'末端未完成Incomplete at 5' end/Incomplete at 3' end
 - FASTA定义行 (以序列号起始) FASTA def line starts with sequence ID
 - 核酸序列的FASTA格式FASTA format for nucleotide sequences
 - 核酸的定义行 (标题) Nucleotide Definition line (title)
 - 核酸FASTA格式的第一行的例子Example of the first line of a nucleotide FASTA sequence
 - 核酸序列对比的FASTA + GAP格式FASTA+GAP format for aligned nucleotide sequences
 - 以PHYLIP格式输出的核酸序列对比PHYLIP format for aligned nucleotide sequences
 - 以NEXUS格式输出的核酸序列对比NEXUS format for aligned nucleotide sequences
 - 对PHYLIP和NEXUS的一些补充信息Source Modifiers for PHYLIP and NEXUS
 - 导入系列对比序列片段Importing aligned sets of segmented sequences
 - 向Sequin导入核酸序列To import a nucleotide sequence into Sequin
 - 向Sequin导入一系列核酸序列To import a set of nucleotide sequences into Sequin
 - 蛋白质页Protein Page
 - 由蛋白测序确认的概念性翻译Conceptual translation confirmed by peptide sequencing
 - N-端不完整/C-端不完整Incomplete at NH₃ end/Incomplete at CO₂ end
 - FASTA定义行 (以序列号起始) FASTA def line starts with sequence ID
 - 建立有CDS间隔的原始mRNA Create initial mRNA with CDS intervals
 - 蛋白序列的FASTA格式FASTA format for protein sequences
 - 蛋白质的定义行 (标题) Protein Definition line (title)
 - 蛋白质FASTA格式的第一行的例子Example of the first line of a protein FASTA

- sequence:
 - 向Sequin导入蛋白序列To import a protein sequence into Sequin
 - 向Sequin导入一系列蛋白序列To import a set of protein sequences into Sequin
- 序列导入确认窗口Sequence Import Check Window
- 注解页Annotation Page
 - 对一系列序列增加基因、rRNA和CDS属性Adding Gene, rRNA and CDS features to a set of sequences
 - 对一系列序列添加标题Adding a title to a set of sequences
- 来源限定词表格Source Modifiers Form
- 组配跟踪Assembly Tracking
- 编辑记录/档案Editing the Record
 - 概述Overview
 - 向数据库提交最终记录Submitting the finished record to the database
- 记录浏览器The Record Viewer
 - 目标序列Target Sequence
 - 显示格式Display Format
 - 摘要Summary
 - 图形Graphic
 - 队列Alignment
 - 序列Sequence
 - GenBank
 - EMBL
 - FASTA
 - ASN.1
 - 桌面Desktop
 - 完成Done
 - 屏幕下端的控制按钮Controls on the bottom of the screen
- 描述符Descriptors
 - 概述Overview
 - 注释菜单Annotate menu
 - 更新数据Update date
 - 生成数据Create date
 - GenBank block
 - 领域Region
 - 评述Comment
 - 标题Title
 - 分子描述Molecule description
 - 分子Molecule
 - 完整性Completedness
 - 技术方法Technique
 - 分类Class:
 - 分子构象Topology:
 - 链Strand:
 - 生物学来源Biological Source
- 特征 Feature s
 - 概述Overview
 - 注释菜单Annotate menu
 - 特性页Properties Page
 - 常规子页General Subpage
 - 评述子页Comment Subpage
 - 引用子页Citations Subpage

- 参照子页Cross-refs Subpage
 - 定位页Location page
- 衰减子attenuator
- C_区C_region
- CAAT_信号CAAT_signal
- 编码区CDS
- 编码区页Coding Region Page
 - 常规子页General Subpage
 - 其它选项子页Exceptions Subpage
 - 蛋白质子页Protein Subpage
 - 产物子页Product Subpage
- 冲突 conflict
- D-环D-loop
- D_片段D_segment
- 增强子enhancer
- 外显子exon
- GC_信号GC_signal
- 基因gene
- 干涉DNAiDNA
- 内含子intron
- J_片段J_segment
- 长终止重复LTR
- 成熟肽mat_peptide
- misc_binding
- misc_difference
- misc_feature
- misc_recomb
- misc_RNA
- misc_signal
- misc_structure
- 修饰的碱基modified_base
- mRNA
- N_区N_region
- 原序列old_sequence
- 加尾信号polyA_signal
- polyA_位点polyA_site
- 前体RNA precursor_RNA
- 初级转录本 prim_transcript
- 引物结合位点primer_bind
- 启动子promoter
- 蛋白质结合位点protein_bind
- 核糖体结合位点RBS
- 重复区repeat_region
- 重复单元repeat_unit
- 复制原点rep_origin
- rRNA
- 转换区S_region
- 卫星序列satellite
- 细胞质小分子RNA scRNA
- 信号肽sig_peptide
- 核内小分子RNA snRNA
- 核仁小分子RNA snoRNA

- 来源source
- 茎环结构stem_loop
- 序列标签位点STS
- TATA_信号TATA_signal
- 终止子terminator
- 转运肽transit_peptide
- tRNA
- 不确定的unsure
- 可变区V_region
- V片段V_segment
- 变异variation
- 3' 发夹区 3'clip
- 3' 非翻译区 3'UTR
- 5' 发夹区 5' clip
- 5' 非翻译区 5'UTR
- -10_信号-10 信号_signal
- -35_信号-35 信号_signal
- 生物学来源Biological Source
 - 概述：描述符或特征Overview: Descriptor or Feature?
 - 物种页Organism Page.
 - 名称子页Names Subpage
 - 定位子页Location Subpage
 - 序列位置Location of Sequence
 - 序列来源Origin of Sequence
 - 遗传密码子页Genetic Codes Subpage
 - 细胞核的Nuclear
 - 线粒体的Mitochondrial
 - 谱系子页Lineage Subpage
 - 限定词子页Modifiers Page
 - 来源子页Source Subpage
 - 物种子页Organism Subpage
 - 综合页Miscellaneous Page
 - 同义词子页Synonyms Subpage
 - 对照子页Cross-refs Subpage
- 出版物Publications
 - 概述：描述符或特征Overview: Descriptor or Feature?
 - 注册表中引用项Citation on Entry Form
 - 状况Status
 - 类别Class
 - 范围Scope
 - 引用信息表格（常规）Citation Information Form (General)
 - 作者页Authors Page
 - 名称子页Names Subpage
 - 联系子页Affiliation Subpage
 - 引用信息表格（选择期刊作为分类）Citation Information Form (If selected Class was Journal)
 - 标题页Title Page
 - 期刊页Journal Page
 - 备注页Remark Page
- 文件菜单File Menu
 - 关于SequinAbout Sequin

- 帮助Help
- 打开Open
- 将FASTA序列直接导入序列编辑器FASTA Nucleotide Direct to Sequence Editor
- 关闭Close
- 输出到GenBank Export GenBank
- 复制浏览Duplicate View
- 存储Save
- 另存为Save As
- 恢复Restore
- 准备提交Prepare Submission
- 打印Print
- 退出Quit
- 编辑菜单Edit Menu
 - 复制Copy
 - 清除Clear
 - 复制副本Duplicate
 - 编辑序列Edit sequence
 - 编辑队列Edit alignment
 - 编辑提交者信息Edit Submitter Info
 - 提交子页Submission page
 - 联系子页Contact page
 - 引用子页Citation page
 - 更新序列Update Sequence
 - 特征传递Feature Propagate
 - 添加序列Add Sequence
- 搜索菜单Search menu
 - 查找ASN.1 Find ASN.1
 - 查找原始文件Find FlatFile
 - 按基因查找Find by Gene
 - 按蛋白质查找Find by Protein
 - 按位置查找Find by Position
 - 确认Validate
 - 拼写检查Spell Check
 - CDD BLAST
 - ORF探测器ORF Finder
 - 选择对象Select Target
- 综合菜单Misc Menu
 - 样式管理Style Manager
 - 网络配置Net Configure
 - GenBank记录更新Updating existing GenBank records:
 - 在Medling中执行查询Performing a Medline lookup:
 - 执行分类学查找Performing a Taxonomy lookup:
 - 访问NCBI首页Accessing the NCBI DeskTop:
 - Entrez查询Entrez Query
 - NCBI桌面NCBI DeskTop
- 注释菜单Annotate Menu
- 选择菜单Options Menu
 - 字体Font
 - 图例Legend
- 序列编辑器Sequence Editor
 - 打开序列编辑器Starting the Sequence Editor

- 序列编辑器内的指针移动Moving around the Sequence Editor
- 编辑序列Editing sequence
- 改变显示方式Changing the display
- 编辑CDS（编码区）Editing a CDS (coding sequence)
- 处理多个队列序列Working with sets of aligned sequences
- 产生新队列Creating new alignments
- 序列编辑器文件菜单Sequence Editor File menu
 - 读取序列文件Read Sequence File
 - 从Entrez下载Download from Entrez
 - 输入队列Import Alignment
 - 输出Export
 - 接受更改Accept Changes
 - 取消Cancel
- 序列编辑器编辑菜单Sequence Editor Edit menu
 - 取消Undo
 - 剪切Cut
 - 粘贴Paste
 - 复制Copy
 - 更新Refresh
 - 删除序列Delete Sequence
- 序列编辑器浏览菜单Sequence Editor View menu
 - 浏览模式View mode
 - 编辑模式Edit mode
 - 标签Label
 - 字体Font
 - 选项Preferences
 - 互补链Complement
 - 读码框Reading frames
 - 翻译方式Translation Style
 - 查找Find
 - 查找前一个Find previous
 - 查找下一个Find next
- 序列编辑器特征菜单Sequence Editor Features menu
- 序列编辑器队列菜单Sequence Editor Alignment menu
 - 选择参考Select Reference
 - 选择全部Select all
 - 显示替换/显示全部Show substitutions/Show all
 - 选择变异区/保守区Select Variations/Conservation
 - 点矩阵Dot Matrix
- 序列编辑器窗口按钮Sequence Editor window buttons
 - 转到Go to
 - 查找Look at
 - 合并/分裂特征模式Merge feature mode/Split feature mode
 - 隐藏/显示特征Hide feat./Show feat.
 - 更新Refresh
 - 接受Accept
 - 取消Cancel

1. 介绍 (Introduction)

Sequin 程序是为协助向 GenBank, EMBL 和 DDBJ 序列数据库提交序列而设计的。它是在 NCBI 编写的。NCBI 是 NIH 中 NLM 的一部分。帮助文档的这一

部分提供了关于如何使用 Sequin 表格进行数据提交的基本概括，接下来的几部分将就如何向各个表格中输入信息进行详细的解说。

1.1 帮助文档 (The help documentation)

sequin 帮助文档有两种格式：在线和 WWW (<http://www.ncbi.nlm.nih.gov/sequin/sequin.hlp.html>) 格式。在线版的文本伴随你在 Sequin 表格里的进程而滚动。在窗口上部的“查找”命令下可输入待查询的词或词组。在线文档也可以以文本格式保存下来，或直接由打印机输出。欲保存文档为文本格式，点击含有帮助文档的窗口，在 Sequin 文件菜单下，选择“输出帮助”(ExportHelp)。若不保存而直接打印，点击帮助窗口，由 Sequin 文件菜单中选择“打印”(Print)。

1.2 表格的组织 (Organization of Forms)

信息经由数个表格输入进 Sequin。每个表格由数页组成，这些页在表格上部以文件夹标签的形式表示。你可以通过点击文件夹标签进入你的目的页，也可以通过点击屏幕下方的“后一页”(Next page)或“前一页”(Prevpage)按钮而进行页间的转换。通过点击表格第一页上的“前一表格”(Prev form)或表格最后一页上的“后一表格”(Next form)按钮，可以完成在表格之间的转换。

文件夹标签分为两个级别。大号粗体字的标签表示每次进入都需要填写的页，小号细体字表示含有不经常用到的参数的页。

有很多方法可以向表格的某页输入信息。许多常用方法，例如文本框，单选按钮，选择框和滚动框等等，在其它程序中经常用到，这里不再详细描述。Sequin 还使用两种不太常用的数据输入法：弹出式菜单和电子数据表

弹出式菜单：当鼠标点击这种菜单时，可选择项将被列出，通过移动鼠标至目的选项并放开而完成选择。只可以选一项。

扩展页：是一个可以根据输入信息的量而改变大小的区域。当你完成向一个区域输入数据后，会出现另一个区域，你可以根据自己意愿向该区域输入附加信息。

1.3 Sequin 总览 (Overview of Sequin)

如果是第一次使用 Sequin，你将需要填写四个表格：“欢迎进入 Sequin”表格，“提交者”表格，“序列格式”表格以及“种属和序列”表格。

当你填完这些表格后，会出现一个含有 Sequin 记录浏览器的窗口。

由这一浏览器你可以调用许多其它表格，编辑你在三个初始表格中填写的内容，或者添加你认为有必要提交的附加信息。

以下将给出如何填写表格和使用记录浏览器的详细指南。

2. Sequin 欢迎界面 (Welcome to Sequin Form)

首先，有三个按钮指示你想要提交序列的数据库：GenBank，EMBL 或 DDBJ。如果你是第一次进行序列提交，点击“开始新的提交”(Start New Submission) 如果你想修改已存在的提交记录，点击“读取现有记录”。

如果你想退出 Sequin 程序，点击 J “退出程序”。

你也可以用“读取现有记录”命令读入 FASTA 格式的序列文件进行分析。序列可以用 Sequin 显示并可以用象 CCD BLAST 一类的软件进行分析，但由于该序列没有适当的注解因此不应被提交。

如果你在联网状态下运行 Sequin，你将看到另一个标明“从 Entrez 下载”的按钮。这一选项允许你通过 Sequin 更新现有的数据库记录。

这些记录将通过 NCBI 的 Entrez retrieval system 由 GenBank 下载到 Sequin 中。记录的内容将出现在 Sequin 中，

你可以根据需要通过更新序列或注解对记录进行编辑。

如果在“欢迎进入 Sequin”界面上你看不到“从 Entrez 下载”的按钮，则你没有在联网状态下运行 Sequin。若欲使 Sequin 处于联网状态，见帮助文档后部的解说。

你只可以更新你自己提交过的序列，不能更新他人提交的序列。要更新现有记录，首先选择你将对其进行更新的数据库。

这一数据库应该是你提交原始序列的数据库。如果你不知道应该使用哪一个数据库，请将记录送交 GenBank 那么 NCBI 工作人员将把它转发到适当的数据库中。接下来，点击“从 Entrez 下载”按钮，在第一个表格中，输入序列的收录号(Accession number)或 GI。

如果你想以更新的形式向某一数据库提交该记录，那么输入“yes”。

填写“提交者”表格。这一表格的指南可以在 Sequin 帮助文档中，Sequin 文件菜单(Sequin File)下的“编辑提交者信息”处找到。这样 Sequin 将打开这一记录。关于如何添加注解或更新序列可以在本文档的“编辑记录”(Editing the record)和“序列编辑器”(Sequence editor)找到。你将不会看到“提交者”表格，“序列格式”表格和“种属和序列”表格。

请注意，无论是新记录还是更新，都应通过电子邮件提交给适当的数据库。Sequin 不支持通过因特网的直接提交。

在“其它”(Misc)菜单下还有一些附加的配置。首先，您可以在单机操作和联网状态之间切换。Sequin 的默认设置是单机操作，这可以满足大多数序列提交的要求。当 Sequin 处于联网状态时，可进行与 NCBI 的数据交换，例如：从 Entrez 检索序列或进行 BLAST 搜索。在以后的“网络设置”(Net Configure)部分将详细描述 Sequin 的联网状态。其次，当您在联网状态下运行 Sequin 时，可以查询 NCBI 的 Entrez 数据库。

关于如何查询 Entrez 请点击此处。第三，您可以启动 NCBI 桌面。该桌面是为高级 Sequin 使用者准备的，详细介绍请见。

3. 提交者表格 (Submission Authors Page)

这一表格的内容将作为序列的引文。它可以包含与正式发表的序列之引文相同的信息。每一表格的底部都有两个按钮，点击“上一表格”(Prev Form)(在表格之首页)或“上一页”(Prev Page)(在表格的后继页)可以切换到前一表格或上一页。点击“下一表格”(Next Form)(在表格之末页)或“下一页”(Next Page)(在表格的前数页)可以切换到下一表格或下一页。

表格中的页可以通过“文件”(File)菜单下的“输出”(Export)功能分别进行保存。如果您在运行多序列提交，可以使用“文件”(File)菜单下的“导入”(Import)功能直接向页中贴入以前曾输入过的信息。

联络页，作者页和机构页可以作为一个模块被保存起来以便下次提交时使用。第一次使用 Sequin 进行序列提交时，在“提交者”表格中填入所需的信息，然后进行已准备好的序列的提交。在记录浏览器中，当提交基本完成时，点击“编辑”(Edit)菜单下的“编辑提交者信息”(Edit Submitter Info)，在产生的“提交指南”表格的文件菜单下，点击“输出提交者信息”(Export Submitter Info)来保存这一信息。

在以后使用 Sequin 提交序列时，如果您已经储存了提交者信息，只需点击“提交者”表格中提交页(Submission page)之“文件”(File)菜单下的“导入提交者信息”(Import Submitter Info)即可。但您仍需要在提交页(Submission page)输入稿件题目。

3.1 提交页 (Submission Page)

3.1.1 我们是否可以在发表前公布此记录? (May we release this record before publication?)

请在两个选项中选一个。如果您选择了 yes，那么当数据库工作人员将这一数据条目加入到数据库中后，

它将被立即公布。如果您选择了 no，会出现一个区域请您填入序列可被公布的日期，这样，被提交的序列将被保管直至序列或 GenBank 标识号被正式发表，或直至您所要求的发布日期，以先出现事件为准。

3.1.2 稿件的临时标题 (Tentative title for manuscript)

请填入一个可以适当描述该序列条目的题目。这一题目将被用于序列提交。您可以选择不使用其作为序列正式发表时的题目。

当提交程序进行到后半段，您将会有机会修改这一信息并添加描述这一序列的、已发表或正被发表的文章的参考文献。

3.2 联系页 (Contact Page)

请填入提交序列者的姓名、电话、传真号码和电子邮件地址。该人将作为关于这一序列提交的联络者。他(她)不必是测序作者之一。电话、传真及电子邮件地址将不会显示在数据库记录中。

3.3 作者页 (Authors Page)

请填入为该序列的测序做出学术贡献的人们的姓名。联络页中的联络人将被自动列为第一作者，您可以根据需要对这一信息进行修改。作者姓名应以“名，中名，姓”的顺序填入。您可以根据自己意愿填入任意数目的作者。

当您输入第三位作者的姓名后，文本框将变成一个扩展页，您可以使用空格键换行。

3.4 机构页 (Affiliation Page)

请填入关于进行测序操作的机构的信息。该机构可以不是联络页中联络者的工作单位。这一信息将在记录的参考信息部分出现，并以“直接提交”(Direct Submission)为抬头。

4. “序列格式”表格 (Sequence Format Form)

使用这一表格描述您所提交的序列的类型和格式。

Sequin 可以处理单个核酸序列或一批相关的序列。若一批序列在同一篇文章中发表或来源于同一生物，那么您可以进行“批量提交”(Batch Submission)。生物学意义上相关的一批序列可以被归类为“环境样本”(Environmental samples)，种群(Population)，系统发育(Phylogenetic)，突变(Mutation)研究或序列片段集合(Segment Sets)。序列片段集合包含覆盖特定基因区域的不互相重叠的序列集。在所有情况下，虽然一批序列被视为一次提交，但每一序列将会得到它自己的数据库编号(Accession Number)并可以单独被注解。Sequin 可以将作为 Phylogenetic, Population, 突变集或环境样本而提交的一批序列以序列比对的形式展示出来。这些序列可以以 FASTA, Contiguous (FASTA+GAP, NEXUS, MACAW)或 Intercleaved (PHYLIP, NEXUS)格式被提交。

如果序列是 FASTA 格式的, Sequin 将会生成一个序列比对(Alignment), 如果序列已经是 FASTA+GAP, PHYLIP,

MACAW 或 NEXUS 格式的比对, 则 Sequin 不会改变原有的比对。如果您的序列比对中的某一序列已经存在於 GenBank/EMBL/DDBJ 数据库中, 您必须对该序列进行标记以避免其被给予新的国际序列号(Accession Number)。如果该序列已有的 AC Number 是 U12345, 那么请将其标识为 accU12345, 而不要给它一个新的序列标识(Sequence Identifier)。

单个序列、序列片段集以及批量提交都必须以 FASTA 格式进行。

4.1 提交类型(Submission type)

选择您正在生成的提交的类型。

- 单个序列: 一条 mRNA 序列或基因组 DNA 序列。如果您在提交在同一文章中被发表的多个序列, 请考虑使用“批量提交”。如果您决定提交多个分别含有一个或多个序列的 Sequin 文件, 请将每一文件用单独的电子邮件提交。
- 片段序列: 一系列覆盖特定基因区域的、不重叠、不连续的序列。典型的例子是一组编码某基因外显子的基因组 DNA 序列以及它们的内含子片段。但是, 如果片段序列是一个比对的一部分, 请适当的选择种群, 系统发育或突变研究组。
- 种群研究: 通过测序同种生物不同个体的同一基因而得出的一组序列。
- 系统发育研究: 通过测序不同生物的同一基因而得出的一组序列。
- 突变研究: 通过测序单一基因的多个突变而得出的一组序列。
- 环境样本: 通过测序一群未分类的或未知生物的同基因而得出的一组序列。
- 批量提交: 不属于种群, 突变或系统发育研究的一组相关序列。这些序列应该在某种意义上相关联, 例如来源于同一篇文章或同种生物。您应该有计划使这些序列在同一天被发布。

4.2 序列数据格式 (Sequence data format)

使用“单选按钮”选择一种数据格式。如果您在提交单个序列、片段序列或在进行批量提交, 您的序列必须是FASTA格式。如果您所提交的一组序列是作为种群, 系统发育或突变研究的一部分, 则您可以选择序列格式。您可以将这一组序列用FASTA格式作为独立序列提交, 但是如果您的序列已经比对过, 那么您可以将这些序列作为一个比对的一部分进行提交。目前Sequin接受FASTA+GAP, PHYLIP, MACAW, MEXUS Interleaved, 和NEXUS Contiguous等格式。所有格式将在以后的“核酸页”(Nucleotide Page)中给予描述。

4.3 提交类别 (Submission Category)

利用单选按钮指出你的序列是否符合“原始提交”(original submission)或“第三方注释提交”(third-party annotation submission)。如果是在你的实验室直接测得的核苷酸序列, 你的提交就被作为“原始提交”(original submission)。

如果已经从GenBank中下载了序列, 而且添加到注释中, 你的条目就可以提交到“第三方注释数据库”(Third-Party Annotation Database, TPA)。

为了将序列释放到 TPA 数据库, 序列必须出现在生物学期刊的同等出版物中。在提交过程中需要提供原始序列的“GenBank 收录号”(GenBank Accession number), 以便完成 TPA 提交。

5. 物种和序列表格 (Organism and Sequences Form)

这个表格由三页组成。

在物种页中，您要说明序列是从哪种生物中取得。但是，正如下面要解释的，当一组序列要做为一个系统发育研究的一部分来提交时，物种名应该由序列文件自己表明或在接下来的来源修改表格中标明。第二也是核苷酸表格，要求您从一个单独的文件中把核苷酸序列输入 Sequin 中。第三页的属性随前一个序列格式页的提交种类而改变。如果您输入的是一个单独的或分段的序列，那么将出现一个蛋白质页面，该页面会要求您输入相关的氨基酸序列。如果你输入的是一个种群，系统发育或突变研究，这一页将变成注释(annotation)页面，你将可以加如核苷酸序列的注释。

5.1 物种页 (Organism Page)

有关序列来源的物种资料应该在此页输入。或者在任何种类的提交中，物种名和任何附加资料都可以在核苷酸序列的文件中标明。事实上，如果你提交的是一系列来做系统发育研究的序列，您将不能在物种名页面输入任何信息。相反，您必需在序列文件中或接下来的来源修改表格中标明。标明物种名如何按格式修改该物种资料的详细介绍存于下面的核苷酸页的文件夹中。

5.1.1 物种 (Organism)

滚动式名单包括多种生物的系统命名。想要找名单中的任何系统命名，只要在适当位置输入该系统命名的开头数个字母，名单会自动移到正确位置供您选择物种。当你选好了物种的名称，系统命名栏和用来翻译的遗传编码(geneticcode)将被自动填写。如果该物种有常用名，常用名栏也将会自动填写。你也可以用空格键将页面移动到名单的正确位置。如果你对于系统命名或常用名有任何问题，请参阅 NCBI 的分类学(taxonomy)浏览器(<http://www.ncbi.nlm.nih.gov/Taxonomy/tax.html/>)。

如果你要找的物种名没有列出，就直接输入。如果您不知道系统命名，请尽可能精确进行描述，并同时加上能够特异性识别的标志，比如象某个克隆，分离株(isolate)，菌株或销售商编号(Voucher number)或栽培种名称(cultivar name)，比如：Nostoc ATCC29106, uncultured spirochete Im403, Lauraceae sp. Vasquez 25230 (MO), Rosa hybrid cultivar'Kazanlik'。同时如果可行，如果名字是未发表的，可标明提交时间。更多的象亚种，菌株，分离株或血清型(serotype)的资料可在提交的后来阶段输入。

5.1.2 序列位置 (Location of Sequence)

在供选择的名单中，请输入含有你的序列的基因组位置。绝大多数序列是位于“基因组”中。下面是对这个弹出式菜单所含选择的简单介绍。

- 基因组：染色体。这一分类还包括线粒体和叶绿体中由核基因组编码的蛋白。
- 叶绿体：一种含叶绿素的质体。
- 有色体(chromoplast)：一种不含叶绿素的有色质体，存在于水果和花中。
- 动基体(Kinetoplast)：只在 Kinetoplastida(例如利氏曼原虫属, Leishmania)中存在的一种特殊的线粒体。注意：“knetoplast”应该只用在 kinetoplastida 成员中(trypansomes and bodonids)。
- 线粒体：一种存在于多数真核细胞的细胞质中的半自主，可自身复制的细胞器。
- 质体(plastid)：任何一类具有双层膜结构，集光的细胞器(或其衍生细胞

器)。注意：只有在无法给予更确切的名称例如“叶绿体”时，才使用“质体”。

- 巨大细胞核(macronuclear)：一种仅在有纤毛的原生生物(例如：四膜虫属)中被发现的特殊的细胞核。注意：巨大细胞核只能用于纤毛属的生物。
- 染色体外成分(Extrachromosomal)：其它没有在这里列出的染色体外核酸成分，例如 B 染色体或 F 因子。
- 质粒(Plasmid)：在细菌中被发现的染色体外基因组成分。注意，这不包括为复制目的基因而使用的克隆载体。
- 蓝色小体(Cyanelle)：只存在于 glaucocystophytes(例如 Cyanophora)中的一种特殊质体。注意：“蓝色小体”只能用于 Glaucocystophyceae 属的生物。
- 原病毒：整合到宿主细胞染色体上的病毒。
- 病毒体：一个完整的病毒颗粒。
- 多态核(Nucleomorph)：在 Chlorarachniophyceae 属(例如 Chlorarachnion)和 Cryptophyta 属(例如隐滴虫，Cryptomonas)中发现的退化的核的残迹。注意：nucleomorph 只能用于 Chlorarachniophyceae 属或隐芽属(Cryptophyta)的物种。
- Apicoplast：存在于 apicomplexans 属生物(如疟原虫，Plasmodium)中的退化的质体。注意：“apicoplast”只能用于 apicomplexans 属的生物。
- 白色体(Leucoplast)：一种没有任何色素的质体。
- 前质体(Proplastid)：不成熟的质体。
- 内源病毒(Endogenous_virus)：一个永久性整合到宿主基因组中并随宿主种系而传代的病毒。

5.1.3 翻译用的遗传密码(Genetic Code for Translation)

如果提交类型是“系统发育研究”，那么这一选项将显示“默认的遗传密码”。请使用这一区域选择应被用来翻译该核酸序列的遗传密码。真核生物的遗传密码是“标准”密码。如果您在以上体到的滚动式物种名单中选择一个系统命名，那么这一区域将被自动选择。如果您在序列文件的首行标明了物种名称，Sequin 将在您导入序列文件后自动填写这一区域。但是，如果改物种不是 GenBank 给出的前 500 种物种之一，那么这一区域将不会被自动填写，则您必须选择一种遗传密码。在 NCBI 的分类学页(taxonomy page) (<http://www.ncbi.nlm.nih.gov/htbin-post/Taxonomy/wprintgc?mode=c>)中，可以找到更多有关蛋白质翻译表格的信息。

5.2 核酸页(Nucleotide Page)

在这一页中您需要给出核酸序列以及与之相关的描述性信息。Sequin 也可以依据在核酸序列文件首行的信息对物种名称，株系，染色体号码以及许多其它的修饰项。

如果您想将序列作为系统发育，突变，种群研究或环境样本集的一部分进行提交，您可以选择在文件中或核酸页中标明物种名称，株系，染色体号码以及其它修饰项。

另外，您可以在物种和序列表格后的来源限定词表格(Source Modifiers form)中输入这些信息。

如果您要提交一系列比对的序列并且这些序列之一已经存在於 GenBank/EMBL/DDBJ 数据库中，您必须对该序列进行标记以避免其被给予新的国际序列号(AccessionNumber)。如果该序列已有的序列号是 U12345，那么请将其

标识为 accU12345，而不要给它一个新的序列标识(Sequence Identifier)。

5.2.1 分子(Molecule)

数据库中的序列可以是几种分子类型之一，请在弹出式菜单中选择被测序的分子之类型。

- 基因组 DNA(Genomic DNA): 直接来源于生物的 DNA 序列。注: rRNA 基因的 DNA 序列属于这种分子类型。
- 基因组 RNA(Genomic RNA): 直接来源于生物的 RNA，如病毒。
- 前体 RNA(Precursor RNA): 剪接处理为 mRNA、rRNA、tRNA，或其他胞内 RNA 前的初始转录物。
- mRNA[cDNA]: 一条从 mRNA 衍生的 cDNA 序列。
- 核糖体 RNA(Ribosomal RNA): 来源于核糖体的 RNA(rRNA)的序列，如来源于 rRNA 的 cDNA 序列。
- 转运 RNA(Transfer RNA): 来源于转运 RNA(tRNA)的序列，如来源于 tRNA 的 cDNA 序列。
- 核小 RNA(Small nuclear RNA): 来源于核小 RNA(snRNA)的序列，如来源于 snRNA 的 cDNA 序列。
- 小胞质 RNA(Small cytoplasmic RNA): 来源于小胞质 RNA(scRNA)的序列，如来源于小胞质 RNA 的 cDNA 序列。
- 其它遗传物质[质粒](Other-Genetic[plasmid]): 既不是遗传物质也不是转录产物的序列。如质粒，B 染色体，F 因子。
- cRNA: 由 DNA 转录产生的互补 RNA 序列，主要用于病毒序列的提交。
- 核仁小 RNA(Small nucleolar RNA): 来源于核仁小 RNA(snoRNA)的序列，如来源于 snoRNA 的 cDNA 序列。

5.2.2 拓扑结构(Topology)

请在弹出式菜单中选择分子的拓扑结构，线性(Linear)或环状(Circular)。多数的序列有线性的拓扑结构。如果序列是完全的并且是环状的，例如是一个质粒或一个完整的线粒体基因组，请选择 Circular。

5.2.3 5' 端未完成或 3' 端未完成(Incomplete at 5' end/Incomplete at 3' end)

如果序列在 5' 或 3' 端是未完成的，请选择适当的选项框。如果提交的序列是完整的，例如一个基因的完整编码序列，请不要选择任何一个选项框。

5.2.4 FASTA 定义行以序列号起始(FASTA def line starts with sequence ID)

如果你在序列格式表单中选择交叉存取或连续的格式，那么你将不会看到这个窗口。

当标准的简单提交时，建议您按照下述的要求按 FASTA 格式建立核酸序列。核酸的 FASTA 格式为：第一行为由 ">" 符号开头的对该序列的简单描述，第二行开始为核酸序列。你应按照这个规则，并且将该序列的简单描述按下述填写为：

>SeqID [org=organism scientific name] title

选定这个按钮。序列的唯一标志就是在 ">" 之后的单词。

如果你没有包括序列号，请不要选择这个按钮，那么 Sequin 会自动为其分配一个序列号。然而，请确定在这第一行的描述信息之前确实有 ">"。

5.2.5 核酸序列的 FASTA 格式(FASTA format for nucleotide sequences)

你所提交的序列必须在另一文件中，不能直接在本页中敲入。序列必须以 FASTA 格式提交。每一行序列均不能超出 80 个字符。

注意：如果提交的是作为“环境样本”(Environmental samples)，种群(Population)，系统发育(Phylogenetic)，突变(Mutation)研究或序列片段集合(Segment Sets)中一部分的多序列，那么每个序列你必须均是FASTA格式。然而，它们在同一文件或多个文件中都可以。你可以按下述的说明将有关序列的生物体，染色体或细胞株等的相关信息编码。或者，也可以在“物种和序列表单之后的通过填写“来源修饰语表格”提交。

The line directly above the sequence (the first line in the file, for a single sequence) should read

序列之前的第一行（在单序列文件中就是文件的第一行）的格式为：

>SeqID [org=organism scientific name] [modifier=modifier name] title

例如，

>DNA.new [org=Homo sapiens] [chromosome=17] [map=17q21] Human breast and ovarian cancer susceptibility (BRCA1) mRNA, complete cds.

>: ">"符号必须在描述信息之前。不以">"开始的行均会被程序认为是核酸序列。

SeqID: 你对所提交序列的唯一标识。只有你在核酸页上选择了"FASTA def line starts with sequence ID."选择钮之后才需要你填写。这个名字不能与其它核酸或蛋白序列的名字重复。如果你不为它命名，Sequin 将自动为它产生唯一标识。在提交过程中，数据库的工作人员将会将此标识发为接收号(accession number)。若提交的是一系列序列，而其中的一个已被 GenBank/EMBL/DDBJ 数据库接收，那么你必须将其标出，以避免重新为它编号。只需以其接收号作为标识即可，如 accU12345。

- [org=organism scientific name]: 此域使你能直接在序列文件中将其物种标出。如果序列是作为系统发育学(phylogenetic study)内容提交，则必须填写此项。若是其它类，则既可在这一指明也可以通过前面的物种页说明。必须填写完整的科学名(不能缩写)。本域应如下所示以完整的括号完成。在"="两端无空格，例如一般来说应包括[org=Homo sapiens]，[org=Mus musculus]，[org=Saccharomyces cerevisiae]，and [org=Drosophila melanogaster]。NCBI 在 taxonomy database 中有物种科学和普通名称的相关信息。
- [modifier=modifier name]: 在序列的第一行也可以填写额外的信息。这些额外的信息包括 moltype, 分子(molecule), 染色体chromosome, map, 克隆clone, 组织类型tissue-type, 菌株strain, 以及样品证明specimen-voucher. 有关信息可在生物资源修饰信息页的 Source and Organism 页中找到。只要你将这些信息分别用括号分开，可以按自己的意愿添加。格式必须严格按上面的例子。这是一个多修饰信息的范例：
[strain=BALB/c] [chromosome=5] [sex=male]
[tissue-type=testis][moltype=mRNA]. 它们也可能在提交过程中作为 as Biological Source 的特征在随后提交。
- [lineage=lineage]: 若在 NCBI 分类学数据库中没有你提交的物种的种系相关数据，你可在此处提交完整的种系。
- title: 序列的定义或描述。Sequin 可按照注解(Annotate)菜单的普通定义功能对序列进行自动命名。命名也可在后面的提交过程中通过选择描述菜单 -->Title而进行，菜单 -->Title在记录观察器中的注释菜单中(Annotate menu)，描述如下(below)。请认真选择名称，因为它将作为提

交序列的定义行。如果你所提交的是 一系列“环境样本”(Environmental samples), 种群(Population), 系统发育(Phylogenetic), 突变(Mutation)研究或序列片段集合(Segment Sets), 那么你可以将此域空出。而在 Annotation 页添加序列的名称。

5.2.6 核酸的定义行(标题) (Nucleotide Definition Line (Title))

核酸的定义行, 或标题, 均遵照结构化的格式。对于一个包含编码区的序列, 其定义格式如下:

物种蛋白名(基因代号) mRNA/基因, 全部 / 部分编码区序列

然而, 一般的格式并不能满足所有的要求, 正如下列例子所述:

- 人类乳腺和卵巢癌相关蛋白(BRCA1) mRNA, 完整 cds 序列。
- 人类乳腺和卵巢癌相关蛋白(BRCA1)基因, 第 4 外显子。
- Gallus gallus red-sensitive pigment mRNA, complete cds.
- Bos taurus retinal pigment (RPE1) mRNA, 3' UTR.
- Saccharomyces cerevisiae cystathionine gamma-lyase (CYS3) gene complete cds.
- Arabidopsis thaliana pyruvate dehydrogenase E1 alpha subunit mRNA, nuclear gene for mitochondrial product, complete cds.
- Homo sapiens Down syndrome region, chromosome 13, genomic sequence.
- Ophraella conferta isolate 62 16S ribosomal RNA gene, mitochondrial gene for mitochondrial product.

5.2.7 核酸 FASTA 格式的第一行的例子 (Example of the first line of a nucleotide FASTA sequence)

```
>DNA.new [org=Homo sapiens] [chromosome=17] [map=17q21]
Homo sapiens breast and ovarian cancer susceptibility protein
(BRCA1) mRNA, complete cds.
```

```
>DNA.新 [物种=人类] [染色体=17] [染色体图谱=17q21] 人类乳腺和
卵巢癌相关蛋白(BRCA1) mRNA, 完整 cds 序列.
```

5.2.8 核酸序列对比的 FASTA+GAP 格式 (FASTA+GAP format for aligned nucleotide sequences)

许多程序可以 FASTA 格式输出序列对比的结果。为了比较这些序列, 需要经常插入一些空白区。在 FASTA+GAP 格式中, 这些空白区以“-”表示。禁止使用“?”来表示在序列对比中不确定的碱基, 因为 Sequin 将会在其输入序列时自动去除非 IUPAC 字符。所有序列, 包括其空白区, 均应长度相同。这些空白区只会出现在序列对比中, 而不会在单个的序列。

出现在类似FASTA序列中的FASTA+GAP格式。在上面的FASTA format for nucleotide sequences 中已以对格式化FASTA序列有了详细描述。所有的以FASTA+GAP格式的序列应保存于同一文件中。

下面是一个 FASTA+GAP 格式的例子:

```
>A-0V-1-A [org=Gallus gallus] [strain=C]
TCACTCTTTGGCAACGACCCGTCGTCATAATAAAGATAGAGGGGCAACTAAAGGAAGCTCTA
TTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT
>A-0V-2-A [org=Drosophila melanogaster]
[strain=D]TCACTCTTTGGCAAC---GCGTCGTCACAATAAAGATAGAGGGGCAACTAAAGGA
AGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT
>A-0V-3-A [org=Caenorhabditis elegans]
[strain=E]TCACTCTTTGGCAAC---GCGTCGTCACAATAAAGATAGAGGGGCAACTAAAGGA
AGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT
```

```

>A-0V-4-A [org=Rattus norvegicus]
[strain=F]TCACTCTTTGGCAACGACCCGTCGTCACAATAAAGATAGAGGGGCAACTAAAGGA
AGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT
>A-0V-7-A [org=Aspergillus nidulans]
[strain=G]TCACTCTTTGGCAACGACCCGTCGTCACAATAAAGATAGAGGGGCAACTAAAGGA
AGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT

```

5.2.9 以 PHYLIP 格式输出的核酸序列对比 (PHYLIP format for aligned nucleotide sequences)

许多程序以 PHYLIP 格式输出的核酸序列对比:

下面是一个 PHYLIP 格式的例子:

```

      5      100
A-0V-1-A   TCACTCTTTG GCAACGACCC GTCGTCATAA TAAAGATAGA GGGCAACTA
A-0V-2-A   TCACTCTTTG GCAAC---GC GTCGTCACAA TAAAGATAGA GGGGCAACTA

A-0V-3-A   TCACTCTTTG GCAAC---GC GTCGTCACAA TAAAGATAGA GGGGCAACTA
A-0V-4-A   TCACTCTTTG GCAACGACCC GTCGTCACAA TAAAGATAGA GGGGCAACTA
A-0V-7-A   TCACTCTTTG GCAACGACCA GTCGTCACAA TAAAGATAGA GGGGCAACTA
            AAGGAAGCTC TATTAGATAC AGGAGCAGAT GATACAGTAT TAGAAGAAAT
            AAGGAAGCTC TATTAGATAC AGGAGCAGAT GATACAGTAT TAGAAGAAAT

```

在这个例子中, 第一行说明这里有 5 个序列, 每个均为 100 个核苷酸。接下来是 5 个含有序列号的 5 行, 然后接核酸序列。比如第一个序列的序列号是 A-0V-1-A。注意, 接下来的序列区中均不再含有序列号。

禁止使用”?”来表示在序列对比中不确定的碱基, 因为 Sequin 将会在其输入序列时自动去除非 IUPAC 字符。这些不确定的碱基应以 IUPAC 字符如 N 来表示。PHYLIP 文件应以”-“, 而不是”?”来表示 5'和 3'端缺失的碱基。

你可以修饰这种格式, 使 Sequin 可以得到序列正确的物种以及其它信息。在下面的 Source Modifiers for PHYLIP and NEXUS 中将会有这种范例。

当然, 你也可以只以标准的 PHYLIP 格式提交, 而在 Source Modifiers form 中交待物种, 种系, 染色体等信息。

5.2.10 以 NEXUS 格式输出的核酸序列对比 (NEXUS format for aligned nucleotide sequences)

许多程序以下面的两种 PHYLIP 格式之一输出的核酸序列对比, 交插 NEXUS 和 连续 NEXUS。

禁止使用”?”来表示在序列对比中不确定的碱基, 因为 Sequin 将会在其输入序列时自动去除非 IUPAC 字符。这些不确定的碱基应以 IUPAC 字符如 N 来表示。NEXUS 文件可以”?”来表示 5'和 3'端缺失的碱基。

下面是一个交插 NEXUS 格式的例子:

```

#NEXUS
[This data assembled using Sequencher*, from Gene Codes
Corporation.]
begin data;
  dimensions ntax=5 nchar=100;
  format datatype=dna gap=: interleave;
  matrix

A-0V-1-A   TCACTCTTTG GCAACGACCC GTCGTCATAA TAAAGATAGA GGGCAACTA
A-0V-2-A   TCACTCTTTG GCAAC---GC GTCGTCACAA TAAAGATAGA GGGCAACTA
A-0V-3-A   TCACTCTTTG GCAAC---GC GTCGTCACAA TAAAGATAGA GGGCAACTA

```

```
A-0V-4-A   TCACTCTTTG GCAACGACCC GTCGTCACAA TAAAGATAGA GGGCAACTA
A-0V-7-A   TCACTCTTTG GCAACGACCA GTCGTCACAA TAAAGATAGA GGGCAACTA
```

```
A-0V-1-A   AAGGAAGCTC TATTAGATAC AGGAGCAGAT GATACAGTAT AGAAGAAAT
A-0V-2-A   AAGGAAGCTC TATTAGATAC AGGAGCAGAT GATACAGTAT AGAAGAAAT
A-0V-3-A   AAGGAAGCTC TATTAGATAC AGGAGCAGAT GATACAGTAT AGAAGAAAT
A-0V-4-A   AAGGAAGCTC TATTAGATAC AGGAGCAGAT GATACAGTAT AGAAGAAAT
A-0V-7-A   AAGGAAGCTC TATTAGATAC AGGAGCAGAT GATACAGTAT AGAAGAAAT
```

在这个例子中，最初的几行提供了对这些对比序列的一些说明。接下来的5行含有序列号，然后接核酸序列。比如第一个序列的序列号是 A-0V-1-A。注意，接下来的序列区中也含有序列号。

下面是一个联续 NEXUS 格式的例子：

```
#NEXUS
BEGIN DATA;
DIMENSIONS NTAX=5 NCHAR=100;
FORMAT MISSING=? GAP=- DATATYPE=DNA ;
MATRIX

A-0V-1-A
TCACTCTTTGGCAACGACCCGTCGTCATAATAAAGATAGAGGGGCAACTAAAGGAAGCTCTA
TTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT

A-0V-2-A
TCACTCTTTGGCAAC---GCGTCGTCACAATAAAGATAGAGGGGCAACTAAAGGAAGCTCTA
TTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT

A-0V-3-A
TCACTCTTTGGCAAC---GCGTCGTCACAATAAAGATAGAGGGGCAACTAAAGGAAGCTCTA
TTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT

A-0V-4-A
TCACTCTTTGGCAACGACCCGTCGTCACAATAAAGATAGAGGGGCAACTAAAGGAAGCTCTA
TTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT

A-0V-7-A
TCACTCTTTGGCAACGACCAGTCGTCACAATAAAGATAGAGGGGCAACTAAAGGAAGCTCTA
TTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT
```

在这个例子中，最初的几行提供了对这些对比序列的一些说明。接下来的5行含有序列号，然后接核酸序列。比如第一个序列的序列号是 A-0V-1-A。注意，接下来的序列区中也含有序列号。

你可以修饰这种NEXUS格式，使Sequin可以得到序列正确的物种以及其它信息。在下面的[Source Modifiers for PHYLIP and NEXUS](#)中将会有这种范例。

当然，你也可以只以标准的NEXUS格式提交，而在[Source Modifiers form](#)中交待物种，种系，染色体等信息。

5.2.11 对 PHYLIP 和 NEXUS 的一些补充信息 (Source Modifiers for PHYLIP and NEXUS)

你可以修饰 PHYLIP 或 NEXUS 格式，使 Sequin 可以通过文件末尾的说明得到有关序列正确的物种及其它信息。第一行针对第一个序列，第二行针对第二个序列，以此类推。每一个序列必须有单一的一行来说明。这些插入行的作用与 FASTA 文件中在序列前的均相同，但不需要以序列号开头。相反序列号只应象上面所述出现在序列的开头。

每一个初始行均应以">"打头。以括号将物种的科学名括起来。其它说明也应括起来。可以为每一行均增加个别序列标题，也可以在注视页中对所有序列增加相同的标题。详细信息参照[FASTA format for nucleotide sequences](#)。对于标题的格式，参照[Nucleotide Definition line \(title\)](#)。

以下的有关物种和种系的相关信息应紧接上面的 PHYLIP 和 NEXUS 的例子之后。

```
>[org=Gallus gallus] [strain=C]
>[org=Drosophila melanogaster] [strain=D]
>[org=Caenorhabditis elegans] [strain=E]
>[org=Rattus norvegicus] [strain=F]
>[org=Aspergillus nidulans] [strain=G]
```

当然，你也可以只以标准的 PHYLIP 或 NEXUS 格式提交，而在 [Source Modifiers form](#) 中交待物种，种系，染色体等信息。

5.2.12 导入系列对比序列片段 (Importing aligned sets of segmented sequences)

Sequin 也可以读入以 FASTA 或 FASTA+GAP 格式进行对比的序列片段。每一片段均应有自己的序列号，但对物种和来源的说明只能在每一序列的第一个片段中说明。通过方括号来限定每一片段。例如，

```
[
>A-0V-1-Apart1 [org=Gallus gallus] [strain=C]
TCACTCTTTGGCAAC
>A-0V-1-Apart2
GACCCGTCGTCATAATAAAGATAGAGGGGCAACTAAAGGAAGCTCTA
TTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT
]
[
>A-0V-2-Apart1 [org=Drosophila melanogaster] [strain=D]
TCACTCTTTGGCAAC
>A-0V-2-Apart2
GAAGCGTCGTCACAATAAAGATAGAGGGGCAACTAAAGGAAGCTCTA
TTAGATACAGGAGCAGATGATACAGTATTAGAAGAAAT
]
```

现在尽管不连续序列可以产生提交文件，但 Sequin 不能正确显示它们。

5.2.13 向 Sequin 导入核酸序列 (To import a nucleotide sequence into Sequin)

当你的序列以合适的 FASTA 或 PHYLIP 格式显示，请击 "Import Nucleotide FASTA" 或 "Import Nucleotide PHYLIP"。这是会出现一个新窗口，显示可用的目录和文件。选择正确的序列文件，然后击 OK。如果导入的序列错误，则选择 EDIT 菜单下的 Clear 命令清除错误的序列。

当导入之后，会出现一个消息框。第一行是导入核酸序列的数量，以及序列的长度。每一个片段均被编号，会例出其长度，唯一的标识(SeqID)，和标题（定义行）。如果有缺失任何信息，请检查序列文件，然后重新导入。

5.2.14 向 Sequin 导入一系列核酸序列 (To import a set of nucleotide sequences into Sequin)

你有两种选择来导入一系列 FASTA 格式的核酸序列。如果所有的序列均在同一文件中，请击 "Import Nucleotide FASTA"。如果它们在不同的文件中，请顺序击 "Import Nucleotide FASTA" 导入它们。无论哪种形式，请确认每个序列都为正确的 FASTA 格式

5.3 蛋白质页 (Protein Page)

注意：本页用于向单一或片段序列增加蛋白序列。如果你提交的是来自种群 (Population)，系统发育(Phylogenetic)，突变(Mutation)研究或序列片段集合 (Segment Sets)中一部分序列，请用 [Annotation Page](#) 页提交，详细说明见下文。

本页允许你提供由你输入的核酸序列翻译而成的蛋白序列。若该核酸序列可被不同剪切或它含有不同的阅读框，可在本页输入所有的蛋白质序列。每个蛋白序列均会在数据库以编码序列 (CDS) 出现。Sequin将会自动决定哪些核酸序列为蛋白的编码区，并且在数据库中标出核酸序列中的非编码区。Sequin也会提供工具，使你能以图像的形式在你的核酸序列中标出所有的开放阅读框，并把它们转变成CDS属性。这些工具将在下面的 [ORF Finder](#) 中有详述。

5.3.1 由蛋白测序确认的概念性翻译 (Conceptual translation confirmed by peptide sequencing)

大多数蛋白质均是计算机产生的对核酸序列的概念性翻译。若你是通过直接测序得到的整个蛋白或由蛋白衍生的多肽的序列数据，请选择这个选择钮。

5.3.2 N-端不完整/C-端不完整 (Incomplete at NH3 end/Incomplete at CO2 end)

若序列缺少蛋白的 N-端或 C-端，请选择这个选择钮。

5.3.3 FASTA 定义行以序列号起始 (FASTA def line starts with sequence ID)

当标准的简单提交时，建议您按照下述的要求按 FASTA 格式建立蛋白序列。FASTA 格式为：第一行为由 ">" 符号开头的对该序列的简单描述，第二行开始为蛋白序列。你应按照这个规则，并且将对序列的简单描述按下述填写为：

>SeqID [gene=locus;description] [prot=name;description] title

选定这个按钮。序列的唯一标志就是在 ">" 之后的单词。

如果你没有包括序列号，请不要选择这个按钮，那么 Sequin 会自动为其分配一个序列号。然而，请确定在这第一行的描述信息之前确实有 ">"。

5.3.4 建立有 CDS 间隔的原始 mRNA (Create initial mRNA with CDS intervals)

若选择这个按钮，Sequin 将会以 CDS 属性的相同间隔作为 mRNA 属性。当记录被装载完成之后，你可以编辑 mRNA 属性，增加 5' UTR 和 3' UTR 区。这些工作可用 mRNA 编辑器或序列编辑器完成。

5.3.5 蛋白序列的 FASTA 格式 (FASTA format for protein sequences)

你所提交的序列必须在一文件中，而不能直接在本页中敲入。序列必须以 FASTA 格式提交。如果提交的是多序列，那么每个序列必须均是 FASTA 格式。每一行序列均不能超出 80 个字符。在向 Sequin 导入序列前，去除所有含有诸如 "Z" 或 "*" 之类的终止子。

序列之前的第一行（在单序列文件中就是文件的第一行）的格式为：

>SeqID [gene=locus;description] [prot=name;description] [comment=text] title

例如，

```
>Prot.new [gene=BRCA1] [prot=Breast and ovarian cancer  
susceptibility protein]
```

```
Homo sapiens breast and ovarian cancer susceptibility protein  
(BRCA1) protein, complete sequence.
```

- >: ">" 符号必须在描述信息之前。不以 ">" 开始的行均会被程序认为是蛋白序列。

SeqID: 你对所提交序列的唯一标识。只有你在核酸页上选择了 "FASTA def line

starts with sequence ID."选择钮之后才需要你填写。这个名字不能与其它核酸或蛋白序列的名字重复。如果你不为它命名, Sequin 将自动为它产生唯一标识。在提交过程中, 数据库的工作人员将会将此标识发为收录号 (accession number)。

- [gene=gene locus]: 输入[gene=gene locus]。如下所示本域应以完整的括号完成。在 "="两端无空格, 例如[gene=eIF4E].
- [prot=protein name]: 输入[prot=protein name]。如下所示本域应以完整的括号完成。在 "="两端无空格, 例如[prot=eukaryotic initiation factor 4E-I].
- title: 序列的定义或描述。Sequin 可按照你所提供的信息对序列进行自动命名。数据库的工作人员会在必要的时候对它进行修饰。
- 必须在登录时包括蛋白的名字, 其它均是可选的。

5.3.6 蛋白质的定义行 (标题) (Protein Definition line (title))

核酸的定义行, 或标题, 均遵照结构化的格式。对于一个包含编码区的序列, 其定义格式如下:

物种蛋白名 (基因名) 蛋白, 全部 / 部分编码区序列 (Genus species Protein name (gene name) protein, complete/partial sequence)

使用以物种种系格式存在的名字。根据基因的性质选择是全部还是部分编码区序列。

然而, 一般的格式并不能满足所有的要求, 正如下列例子所述:

- Homo sapiens breast and ovarian cancer susceptibility protein (BRCA1) protein, complete equence.
- Gallus gallus red-sensitive pigment protein, complete sequence.
- Bos taurus retinal pigment (RPE1) protein, partial sequence .
- Saccharomyces cerevisiae cystathionine gamma-lyase (CYS3) protein, complete sequence.
- Arabidopsis thaliana pyruvate dehydrogenase E1 alpha subunit protein, complete sequence, mitochondrial gene for mitochondrial protein.

5.3.7 蛋白质 FASTA 格式的第一行的例子 (Example of the first line of a protein FASTA sequence) :

```
>Prot.new [gene=BRCA1] [prot=Breast and ovarian cancer  
susceptibility protein]  
Human breast and ovarian cancer susceptibility protein (BRCA1)  
protein, complete sequence.
```

5.3.8 向 Sequin 导入蛋白序列 (To import a protein sequence into Sequin)

当击"Import Protein FASTA."时会出现一个新窗口, 显示可用的目录和文件。选择正确的序列文件, 然后击 OK。如果导入的序列错误, 则选择 EDIT 菜单下的 Clear 命令清除错误的序列。

当导入之后, 会出现一个消息框。第一行是导入蛋白序列的数量, 以及序列的长度。每一个片段均被编号, 会例出其长度, 唯一的标识(SeqID), 和标题 (定义行)。如果有缺失任何信息, 请检查序列文件, 然后重新导入。

5.3.9 向 Sequin 导入一系列蛋白序列 (To import a set of protein sequences into Sequin)

当你想向 Sequin 输入一系列非连续的蛋白质序列时, 例如你要导入一段核酸序列的多个开放阅读框。你有两种选择来导入一系列 FASTA 格式的核酸序列。如果所有的序列均在同一文件中, 请击"Import Protein FASTA."。如果它们在不同的文件中, 请顺序击"Import Nucleotide FASTA."导入它们。无论哪种形式, 请确认每个序列都为正确的 FASTA 格式。

5.4 序列导入确认窗口 (Sequence Import Check Window)

当你导入蛋白序列之后, 出现一个新窗口, 你可以在这里编辑有关蛋白序列的信息。若你没有为序列输入唯一的标识 (SeqID), Sequin 会自动产生。必须填 SeqID 和蛋白姓名, 但蛋白名称也可在随后的提交过程中修改。其它信息是可选的。

5.5 注解页 (Annotation Page)

注意: 本页用于添加对种群(Population), 系统发育(Phylogenetic), 突变 (Mutation), 环境样本等多序列的注释。若你提交的是单一或片段序列, 本页就是指上述的 [Protein page](#)

5.5.1 对一系列序列增加基因, rRNA 和 CDS 属性 (Adding Gene, rRNA and CDS features to a set of sequences)

使用本页顶部的单选按钮选择向你的序列增加的属性。在本页上增加的任何注释针对所有的序列, 并会适用于每个序列的全长。若你只想针对于其中的某一些序列, 你必须在随后的提交过程中手动说明。你可以只选择其中的一些按钮。

- **Gene:** 选择此按钮向所有的序列添加一个基因属性。一旦你选定这个钮, 在表单上将会出现一些输入行。若基因比提交的序列长, 请选择 5'端不完整或 3'端不完整或均不完整。输入基因标志, 若需要可添加注释。
- **rRNA:** 若想向所有的序列添加 rRNA 属性, 请选择 rRNA。这个属性应包括序列的全长, 从第一个核酸到最后一个核酸。一旦你选定这个钮, 在表单上将会出现一些输入行。若序列不编码完整的 rRNA 分子, 请选择 5'端不完整或 3'端不完整或均不完整。输入 rRNA 的名称, 如 16s 核糖体 RNA。若你添加了基因属性, 程序会自动添加基因标志选钮。注释框使你可向 rRNA 添加额外的注释。
- **CDS:** 若想向所有的序列添加开放阅读框属性, 请选择 CDS。Sequin 可自动选择序列上最长的开放阅读框, 并标为 CDS 属性。若你序列含有多个开放阅读框, 或该阅读框并不是全长的, 你可在随后的提交过程中使用 [Coding Region](#) 属性窗口编辑。一旦你选定这个钮, 在表单上将会出现一些输入行。我们鼓励你填写蛋白和基因域。在注释框中输入信息是可选的。
- **None:** 若你不想向序列添加 rRNA 或 CDS 属性, 请选择 none。

5.5.2 对一组序列添加标题 (Adding a title to a set of sequences)

通过指令组 (box), 给所提交的每一条序列添加标题。注意相同标题将赋予所有的序列。在 FASTA 格式中已具有标题的, 这里不必再添加。有关数据库首选标题格式的具体介绍请参考在上面的核酸页 (Nucleotide page)。

标题通常应该由生物的名称开始。如果所有序列均来源于同一生物, 直接将生物体名称写入标题。但是, 如果所提交序列得自不同生物, 比如在系统发生研究, 标题不要写有生物名称, 这时要选中 "Prefix title with organism name (在标题加入生物名称前缀)", Sequin 将在每一个序列名称前添加相应的生物名称。

利用 "生成定义行" (Generate Definition Line) 的 "注释" ([Annotate](#)) 菜单, Sequin 还可以自动生成标题。另外, 在提交序列的后期, 选择记录浏览器里 "注释" (Annotate) 菜单下的 "描述符" (Descriptor) --> "标题" (Title) 也能够编辑标题, 详见下文。

6. 来源限定词表格 (Source Modifiers Form)

当以系统发生(phylogenetic)、突变(mutation)、群体研究(population study)、成套环境样本(environmental samples set)或者批处理提交序列(as a batch submission)时,你只能见到本表。

本表是你添加或者修改物种(organism)等必需信息,或者有关提交序列的株系、染色体等可选择的信息。从弹出菜单(pop-up menu)的顶部,选择你注释需要的限定词。左侧柱状表(The left column)列出的是序列的SeqID或者所提交序列的唯一标志符。在标有“评价”(Value)的相应空格内敲入序列限定词。例如,如果选择了物种限定词“Organism modifier”,你可以在第一个“评价”(Value)空格内键入Mus musculus,在第二个键入Homo sapiens等等。

如果你还没有提供生物体的学名,在这个表格中填写。不要用缩写。

在生物来源限定词页(Biological Source Modifiers page)的来源(Source)子页and物种(Organism)子页中,可以查找限定词的完整描述。可以添加你希望的所有限定词。

7. 组配跟踪(Assembly Tracking)

如果提前指定条目是“第三方注释提交”(Third-Party Annotation submission),你只能看到这张表格。你必须提供原始序列的“GenBank 收录号”(GenBank Accession number)来组配TPA序列。我们不能接受与参考序列相同或来自私人数据库的原始序列。关于这方面的更多信息请参考TPA主页。

如果在组配跟踪表格的第一行收入了“GenBank 收录号”(GenBank Accession number),GenBank的职员会为你量身定作。你不必填写“from”和“to”的相应栏。注意为将提供组配序列全部覆盖到,要输入多个收录号。

你也可以通过“注释”(Annotate)菜单下的记录浏览器来产生组配跟踪表格,从下拉菜单中选择描述符和TPA组配(Descriptors and TPA Assembly)。

8. 编辑记录(Editing the Record)

8.1 概述(Overview)

填写完“物种与序列表格”(Organism and Sequences Form)后,Sequin将根据输入信息对你的条目进行处理。此时所看到的窗口叫“记录浏览器”(record viewer)。当你对已有的记录进行更新时,也是使用同样的窗口。因此,无论你是提交新记录还是对已有记录进行更新,直到是相同的。

在记录浏览器的默认窗口中,你所看到的条目与其在数据库中出现情况近似。提交过程早期输入的大部分信息都在记录浏览器中,像联系等其他信息也能够看到,但在数据库条目中不会出现。如果你提供了核酸序列的概念性翻译,该序列将会以CDS特征列出。尽管核酸序列包括内含子和外显子,Sequin自动决定哪些核酸编码蛋白并列出来。

在“文件”(File)“菜单下,选择“保存”(Save)或“另存为”(Save As)将条目保存为文件。这不同于提交到数据库的条目的存储。无论你在编辑过程中做了任何改变,此次保存能够使你重新调出原始拷贝。如果你要以后再编辑该条目,你可以在“Sequin 欢迎界面”(Welcome to Sequin)中点击“读取已存在的提交条目”(Read existing Submission),并选择该文件。

提交到数据库的条目现在可以进行处理了。也可能你还要通过“描述符”(Descriptors)或者“特征”(Features)表格,再添加些信息。一般来说,Descriptors用于对一条完整序列或者一组序列的注释,而Features用于对某一段

序列的注释。例如，可以通过改变参考描述符（Reference Descriptor）增加一条已发表文稿，或者通过增加特征来注释序列的一段信号肽或者 poly A 信号。

记录浏览器中的信息能够以不同方式编辑。其一是通过双击你所要编辑信息的区域来增加或者修改。有很多区域，如“定义”（Definition）、“来源”（Source）、“参考”（Reference）、或“者特征（Features）等都可以编辑。例如，如果你想为序列增加一个参考资料，只需双击 Reference，进入相应的表格。

另一个添加或修改信息的途径是，在“综合”（Misc）或者“特征”（Features）菜单中，选择相应的表格产生新描述符。这些选项将在本帮助文件的后面介绍。

最后，你该编辑序列了。有关序列编辑的说明（[Instructions](#)）在序列编辑器（Sequence Editor）中介绍。

8.2 向数据库提交最终记录(Submitting the finished record to the database)

当添加了所有的相关信息后，就要将你的条目提交到数据库。选择“查找”（Search）菜单下的“确认”（Validate）指令，该功能是检测你所提交的格式与数据库所要求的格式有否差异。

如果 Sequin 发现了你的记录格式有问题，在屏幕上你会看到错误信息和修改建议。单击错误信息，记录浏览器就会滚动到导致错误的特征处。双击错误信息，则启动新的表格，你可以输入信息、修改错误。如果你对一组序列注释，用“shift+单击”寻找目标序列和特征。你也可以不理睬建议，自己处理。当你认为修改了所有的错误，点击“再次确认”（Revalidate）。

信息：选择“详细”（Verbose）、“常规”（Normal）、和“简洁”（Terse）。Verbose 提供更加详细的错误信息。

筛选：选择你想看的错误信息。你可选择全部（ALL）、SEQ_INST（序列自身的错误，如序列类型和长度）、SEQ_FEAT（描述符错误）、或你的记录所特有的错误。

程度：选择你想看到的错误信息的类型。你可以查看所选的信息类型，也可以是任何更加严重的问题信息。

错误信息有 4 类：“消息”（Info）、“警告”（Warning）、“错误”（Error）、和“拒绝”（Reject）。“消息”（Info）是最轻的，“拒绝”（Reject）是最严重的。你可以提交有错误的记录。但我们建议你尽可能地修改错误。注意有些信息只是建议，不是完全否定。某些“警告”（Warning）信息可能是某一连接位点与多数序列不同。这可以是合理结果，但是还是希望你再次检查一下序列。某些“错误”（Error）信息可能是所提供序列的概念性翻译不编码开放读码框。这时你应该用正确的开放读码框来翻译序列。某些“拒绝”（Reject）信息可能是你缺失了序列来源的物种名称。数据库接受的条目必需有生物体名称。

如果 Sequin 没有检查出你的记录有任何格式错误，你将看到“确认检测成功”（Validation test succeeded）。点击提交界面的“完成”（Done）按钮，或者选择“文件”（File）菜单的“准备提交”（Prepare Submission）。你将被提示保存文件，将该文件 E-mail 到数据库。你必须电邮该文件，因为 Sequin 不在网络上自动提交文件。数据库的电子邮件地址见下：

- GenBank: gb-sub@ncbi.nlm.nih.gov
- -EMBL: datasubs@ebi.ac.uk
- -DDBJ: ddbjsub@ddbj.nig.ac.jp

完成你的条目，关闭记录浏览器，你将返回“Sequin 欢迎界面”（Welcome to Sequin），可以开始输入新的条目。

9. 记录浏览器 (The Record Viewer)

9.1 目标序列 (Target Sequence)

弹出菜单显示Sequin条目相关的所有核酸和蛋白质序列的SeqIDs列表。借助该菜单选择记录浏览器中所显示的序列，以及你需要的目标序列 (target)，即需要添加描述符的序列 (参见Sequin帮助文件的“描述符” (Descriptors))。你可以根据名称选择某一个序列，或一组序列 (如所有序列)，或选择SEG_dna (针对一组核苷酸片段)。随时你可以改变选择。

9.2 显示格式 (Display Format)

记录浏览器的显示格式可以被更改为下面任何一种。在一种显示格式下编辑的区域会更改所有格式下的相应区域。

9.2.1 摘要 (Summary)

该格式以图表摘要形式显示条目。除了各行没有标注外，显示与图形浏览器浏览相同。顶部的条代表核酸序列。行代表序列的不同特征，如编码序列 (CDS) 或一组序列的附加序列。双击箭头或条打开编辑窗口。

9.2.2 图形 (Graphic)

该格式以图的形式显示条目。顶部的条代表核酸序列。底部箭头或者条代表序列的不同特征。双击箭头或者条打开相应的编辑窗口。“序列编辑器” (Sequence Editor) 中任何突出显示的序列都由方框 (box) 标记 (参考前文的“提交类型” (Submission type))，目标序列要设定为“SEG_dna”。

“样式弹出菜单” (Style pop-up menu) 可以用多种样式和颜色显示结果。

“刻度弹出菜单” (Scale pop-up menu) 能够以不同尺寸显示结果。数值越小，显示比例越大。

9.2.3 队列 (Alignment)

该格式显示一组队列序列。例如当“目标序列” (Target Sequence) 弹出菜单中设置为“所有序列” (All Sequences) 时，输入的那些作为“群体” (population)、 “系统发生的” (phylogenetic)、 “变异” (mutation)、 “环境样本组” (environmental samples set) 等一部分的序列。每一个序列以一条表示。序列间的差异以红色垂直线表示。双击代表某一序列的条，则打开该单条序列浏览窗口。如果要打开队列编辑器，查看所有序列的排列配比，设置“所有序列” (All Sequences) 的“目标序列” (Target Sequence)，设置队列显示格式 (Display Format to Alignment)，单击序列周围的框使其突出显示，再选择 Sequin “编辑” (Edit) 菜单中的“编辑队列” (Edit Alignment)。

9.2.4 序列 (Sequence)

该格式依照所注解的特征 (如 CDS 或者 mRNA) 显示记录的核苷酸序列。选择不同的选项显示有所变动。使用“序列” (Sequences) 的弹出菜单选择你要编辑的序列。如果记录中有多条序列，选择“队列” (Aligned) 查看所有序列。主序列的全部序列都会显示出来。对于其它序列，与主序列相同的区域以点表示，不同的区域用字母表示。如果多个序列是作为“系统发生的”

(phylogenetic)、 “群体” (population)、 “变异研究” (mutation study) 部分序列，主序列可以通过在“目标序列” (Target Sequence) 的弹出菜单中选择不同序列。你可以使用“特征” (Features) 弹出菜单改变所显示的特征。在“目标序列” (Target Sequence) 的弹出菜单选择序列或队列所要显示的特征，或不显示特征。利用“编号” (numbering) 弹出菜单可以选择序列编号的位置，在窗口的侧面还是每条序列行的顶部，或者不编号。

9.2.5 GenBank

该格式显示提交到 GenBank 或者 DDBJ 的格式。

9.2.6 EMBL

该格式显示提交到 EMBL 的格式。

9.2.7 FASTA

该格式仅显示序列和定义行，而没有任何注释，称之 FASTA 格式。很多分子生物学分析程序使用这个格式。在这种显示方式下不能编辑。

9.2.8 ASN.1

该格式以“抽象语法符号 1”（Abstract Syntax Notation 1）显示条目，是 NCBI 使用的数据库描述语言。在这种显示方式下不能编辑。

9.2.9 桌面（Desktop）

NCBI 桌面显示用 Sequin 浏览记录时的内部结构。“桌面”（DeskTop）将在“综合”（Misc）菜单下解释。

9.3 完成（Done）

该按钮使你在完成提交时确认条目。参见“提交记录到数据库”（Submitting the finished record to the database）

9.4 屏幕下端的控制按钮（Controls on the bottom of the screen）

如果在 Entrez 下载序列，你会在屏幕底部看到附加的控制按钮。从 Entrez 下载的序列具有 Entrez 链接按钮。例如，选择“核苷酸”（Nucleotide）弹出菜单，可以看到与“目标序列”（Target Sequence）弹出菜单中类似的核苷酸序列。同样，选择 Medline 可以查看任何文献链接。

10 描述符（Descriptors）

10.1 概述（Overview）

“描述符”（Descriptors）用于注释给定条目的一条完整序列，或者一组序列。因为其应用于整个序列，它们没有序列上的特定位置。与“特征”（Features）相比较，后者应用于序列内部的特定范围。

你可以用下面两种方法之一编辑描述符。

（1）在记录界面，双击描述符正文启动一个表格，用来添加信息。

（2）在注释（Annotate）菜单中选择 Descriptors。

10.2 注释菜单（Annotate menu）

该菜单使你进入生成新的、或者修改已有的描述符的功能。选择你所想修改的描述符。

你选了一个描述符时，可以看到“Descriptor Target Control”窗口。使用该目标控制弹出菜单，选择你该描述符适用的序列。在核苷酸或者氨基酸序列输入 Sequin 时，其 SeqID(s)与所列出的名称相对应。该菜单的缺省选项是设置“Target Sequence”弹出菜单为记录界面。可以选择该描述符使用于单条还是一组序列。如果生成一条新的描述符，选择 Create New，如果修改一条以前的描述符，选择“Edit Old”

下面列出可以添加的描述符。有关“出版物”（Publications）和生物学（Biological Source）来源的两项描述符在其他章节中介绍。

10.2.1 更新数据（Update date）

此仅限数据库职员使用。请不要修改其数据。

10.2.2 生成数据（Create date）

此仅限数据库职员使用。请不要修改其数据。

10.2.3 GenBank block

此仅限数据库职员使用。

10.2.4 区域 (Region)

该描述符提供了有关该序列的遗传背景的一般信息。例如，如果序列是从 Huntington's Disease 基因附近克隆的，你可以在这里输入相关信息。为描述符提供信息是可选的。

10.2.5 评述 (Comment)

该描述符用于列出你想提供的有关序列的任何其他信息。你可以不用此项功能。

10.2.6 标题 (Title)

该描述符包含的信息将出现在数据库条目的定义行。你所向Sequin输入的核苷酸序列的名称就在这个位置。如果你要修改定义行，或者提交序列时没有给标题，编辑该描述符。有关生成正确定义行的更多信息，请参见[Organism and Sequences Form](#)。

10.2.7 分子描述 (Molecule description)

该描述符简要介绍获得序列分子的特征。在这里可以编辑已经输入的信息。

10.2.7.1 分子 (Molecule)

GenBank 中的序列包括代表多种类型的不同分子。在 Molecule 弹出菜单中输入分子类型。该菜单选项的简单说明前面已经列出。

10.2.7.2 完整性 *Completedness*

在菜单中选择恰当的选项。

- **Complete:** 在提交一个完整单元时，如一个基因的全部编码序列，使用这个指令。
- **Partial:** 在提交一个不完整的单元时，该序列的哪一端不完整尚不清楚，如一个基因的部分编码序列，使用此指令。
- **No left:** 在提交一个不完整的单元时，该序列的左端的 5' 末端或者氨基末端不完整，如一个基因的部分编码序列或者部分蛋白质序列，使用此指令。
- **No right:** 在提交一个不完整的单元时，该序列的右端的 3' 末端或者羧基末端不完整，如一个基因的部分编码序列或者部分蛋白质序列，使用此指令。
- **No ends:** 在提交一个不完整的单元时，该序列的 5'、3' 末端或者氨基、羧基末端均不完整，如一个基因的部分编码序列或者部分蛋白质序列，使用此指令。
- **Other:** 在上述任何情况都不附时，使用此指令。

10.2.7.3 技术方法 (*Technique*)

在弹出式的菜单中选择用于生成序列的技术方法。

- **Standard:** 标准测序技术
- **EST:** 表达序列标签：从 cDNA 一次测序获得的低质量的 mRNA 序列。这些序列在 EST 部分。single-pass, low quality mRNA sequences derived from cDNAs. These sequences will appear in the EST division.
- **STS:** 序列标签位点：已经绘制到基因组上的 EST。这些序列在 STS 部分。Sequence Tagged Site: an EST sequences which has been mapped onto the genome. These sequences will appear in the STS division.
- **Survey:** 概览，一次测通的基因组序列。这些序列在基因组概览 (GSS)

部分。single pass genomic sequence. These sequences will appear in the Genome Survey Sequence (GSS) division.

- **Genetic Map:** 遗传图谱, 例如遗传图谱信息, 在基因组 (genome) 部分。genetic map information, for example, in the Genomes division.
- **Physical Map:** 物理图谱, 例如物理图谱信息, 在基因组 (genome) 部分。physical map information, for example in the Genomes division.
- **Derived:** 从短序列组装成一个重叠克隆群而获得的序列。A sequence assembled into a contig from shorter sequences.
- **Concept-trans:** 蛋白质序列, 通过适当的遗传密码翻译来的。a protein translation generated with the appropriate genetic code.
- **Seq-pept:** 蛋白质序列, 通过肽链直接测序得到的。protein sequence was generated by direct sequencing of a peptide.
- **Both:** 蛋白质序列, 通过概念性翻译产生, 并经肽链测序验证。protein sequence was generated by conceptual translation and confirmed by peptide sequencing.
- **Seq-pept-Overlap:** 蛋白质序列, 对多个肽链测序, 并由这些肽链序列的重叠排序得到的。protein sequence was generated by sequencing multiple peptides, and the order of peptides was determined by overlap in their sequences.
- **Seq-pept-Homol:** 蛋白质序列, 对多个肽链测序, 并由这些肽链序列的相同序列得到的。protein sequence was generated by sequencing multiple peptides, and the order of peptides was determined by homology with another protein.
- **Concept-Trans-A:** 由登录作者提供的核苷酸序列通过概念性翻译得到的。conceptual translation of the nucleotide sequence provided by the author of the entry.
- **HTGS 0:** 高通量基因组序列, 0 阶段。这些序列是由高通量测序计划得到的, 在 HTG 部分。High Throughput Genome Sequence, Phase 0. These sequences are produced by high-throughput sequencing projects, and will be in the HTG division.
- **HTGS 1:** 高通量基因组序列, 1 阶段。这些序列是由高通量测序计划得到的, 在 HTG 部分。High Throughput Genome Sequence, Phase 1. These sequences are produced by high-throughput sequencing projects, and will be in the HTG division.
- **HTGS 2:** 高通量基因组序列, 2 阶段。这些序列是由高通量测序计划得到的, 在 HTG 部分。High Throughput Genome Sequence, Phase 2. These sequences are produced by high-throughput sequencing projects, and will be in the HTG division.
- **HTGS 3:** 高通量基因组序列, 3 阶段。这些序列是由高通量测序计划得到的, 在 HTG 部分。High Throughput Genome Sequence, Phase 3. These sequences are produced by high-throughput sequencing projects, and will be in the HTG division.
- **FLI_cDNA:** 全长插入 cDNA。序列是全长的 cDNA 序列, 但不是必要的完整转录本。这些序列是大规模测序计划获得的。Full Length Insert cDNA. Sequence corresponds to entire cDNA but not necessarily entire transcript. These sequences are produced by large sequencing projects.

- **HTC:** 高通量 cDNA。这些序列是大规模测序计划获得的。High Throughput cDNA. These sequences are produced by large sequencing projects.
- **Other:** 用于指定那些以上描述以外的序列。use this designation when none of the above descriptions apply.

10.2.7.4 分类 **Class:**

从弹出式菜单中，选择测序分子的类型。From the pop-up menu, select the type of molecule which was sequenced.

- **DNA:** DNA
- **RNA:** RNA
- **Protein:** 蛋白质
- **Nucleotide:** 不要选择这项
- **Other:** 不要选择这项

10.2.7.5 分子构象 (**Topology**):

从弹出式菜单中，选择测序分子的分子构象。

- **Linear:** 线性分子 (大多数序列)
- **Circular:** 环状分子 (如完整的质粒或者线粒体)
- **Tandem:** 不要选择此项。
- **Other:** 不要选择此项。

10.2.7.6 链 (**Strand**):

从弹出式菜单中，选择序列来源物种的基因组是单链的或是双链的。该项主要用于病毒提交。

- **Single:** 物种只含有单链基因组，如 ssRNA 病毒。
- **Double:** 物种只有双链基因组，如 dsDNA 病毒。
- **Mixed:** 不要选择此项。
- **Mixed Rev:** 不要选择此项。
- **Other:** 不要选择此项。

10.2.8 生物学来源 (**Biological Source**)

以下 ([below](#)) 将详细介绍生物学来源描述符。

11. 特征 (**Features**)

11.1 概述 (**Overview**)

特征 (Feature) 用于对一个存在一个或多个区间的序列进行注解，这是和“描述符” ([Descriptors](#)) 相对应的，描述符 (Descriptors) 用于对一个完整序列或一整套完整序列进行注解。特征 (Feature) 会被加入到在记录浏览器的弹出式菜单中选定的目标序列中。即使要参考氨基酸序列特点，大多数特征 (Feature) 所指代的还是核苷酸序列水平的。

可以通过以下三种方法中的一种来添加或修改特征 (Feature)。

(1) 在记录浏览器中，双击已存在特征 (Feature) 的文本，弹出一个表格，在表格中可以添加信息。

(2) 从注释菜单中选择特征 (Feature) 选项。

(3) 从序列编辑器特征菜单上选择特征 (Feature) 选项。

列于注释菜单中的特征选项和列于序列编辑器中的特征选项是相同的。如果通过注释菜单来添加特征，必须提供特征的核苷酸序列区域。而通过序列编辑器

来添加特征，就不必知道对应的核苷酸序列。除此之外，两种方法是相同的。通过简单地突出显示特征关联的序列，序列的位置就会自动地存入特征的定位框。

11.2 注释菜单 Annotate menu

这个菜单可以添加或修改记录浏览器目标序列弹出式菜单中选定序列的特征。特征被分为六类。选择喜欢的特征标记你的序列，一个新的表格就会出现。

特征表格具有共同的设计。第一页是特殊特征特有的，例如基因的编码区。第二页列有特征的特性。第三页描述特征的定位。在第一页上填上合适的信息。

11.2.1 特性页 Properties Page

这页列出引用描述特征的特性。

11.2.1.1 常规子页 General Subpage

在这里输入特征的一般性评述。

选择必要的任何旗帜。如果这个序列只是包含有特性的部分描述，选中“Partial”框。如果特性注释的是核苷酸序列的转录后修饰，例如核糖体滑动或RNA编辑，选中“Exception”框。利用弹出式菜单选择支持特性存在的某种证据。如果是实验证明的，选中“Experimental”；如果没有支持特性的实验证据，选中“Non-Experimental”；如果不想做出以上选择，则可以选中空白行。

大多数特征与一个特殊基因相关联，一般来说基因是通过核苷酸序列获得的。选择弹出式菜单中基因的名字。如果要添加一个新基因的名字，选择“new”，并输入基因名称和可选择的描述。在默认情况下，特征和基因之间的图谱绘制是基于它们之间的重叠性的。也就是说，和一个特征相关联的基因就应该在位置上和特征的位置相交迭。在某些情况下，举例来说，如果是两个基因交迭得到的序列，你可能希望将特征应用到一个不同的基因。在这时，选择对照子页

“cross-reference”，在弹出式菜单输入新基因的名称。如果不想将特征与任何基因相关联，选择禁止“suppress”。你也可以通过点击“Edit Gene Feature”，弹出基因特征表格来编辑信息。

11.2.1.2 注释子页 Comment Subpage

添加特征的注释，特别是选中常规子页中的“Exception”框。

11.2.1.3 引用子页 Citations Subpage

这页是用来列出特别用于注释特征的注释。引用必须已经输入到Sequin帮助文件的记录中（参考Publications）。点击编辑应用“Edit Citations”，选中你要引用出版物旁边的复选标记。为了降低数据库的容量，我们不鼓励轻易使用特征的引用项。

11.2.1.4 参照子页 (Cross-refs Subpage)

这页是用于对照本地条目和外部数据库（GenBank、EMBL/EBI和DDBJ以外的数据库）的条目，例如dbEST或者是FLYBASE。大多数用户不用填写这页。要得到这个主题的更多信息，请参考国际核苷酸序列数据库协作页 [International Nucleotide Sequence Database Collaboration page](#)。

11.2.2 定位页 (Location page)

这页用于选择引用特征对位置。每一个特征必须有一个关联的区间序列。大多数情况下，Sequin 需要知道特征是用于核苷酸序列还是蛋白质序列，并相应减少你需要选择的选项。

Sequin 是核苷酸数据库的提交工具。因此，大多数特征的位置只是指向核苷酸序列。举例来说，虽然成熟肽、信号肽、转运肽描述的是蛋白质序列，但它们的位置却指向 DNA。

如果你的核苷酸序列在 5' 末端或 3'末端是缺失的,那就要分别选中 5' Partial 或 3' Partial 框。如果你的氨基酸序列在氨基端或羧基端是缺失的,那就要选中 NH2 Partial 或 COO Partial 框。

输入特征的序列范围。如果 SeqID 代表核苷酸序列,数值要和核苷酸序列区间相对应;如果 SeqID 代表蛋白质序列,数值要和氨基酸序列区间相对应。如果特征包括序列的多个不连续区间,指出每一个区间的开始点和结束点。如果每个区间是分开的,描述特征时就不要彼此结合,选中“存在间隙的散布区间”

“Intersperse intervals with gaps”(例如在注释多引物结合位点时)。如果特征是由几个彼此相连的区间组成,不要选中此项(例如注释 mRNA 序列或基因组 DNA 序列)。

只针对核酸特征:从弹出式菜单中,选择发现特征的链。

- Plus: 正链或编码链。
- Minus: 负链或非编码链。
- Both: 双链。
- Reverse: 不要选择此项。

利用弹出式菜单选择通过位置描述的序列的 SeqID。

以下是有用特征的简单描述。如何使用编码区特征的详细介绍也包含在内。DDBJ/EMBL/GenBank 特征表定义页 (definition page) 提供了其它特征等详细信息。

11.3 衰减子 (Attenuator)

1) 使转录终止的 DNA 区段,它控制一些细菌操纵子的表达;2) 位于启动子和第一个结构基因之间的序列片段,可引起转录的部分终止。

11.4 恒定区 (C_region)

免疫球蛋白轻链、重链以及 T 细胞受体 α 、 β 、 γ 链的恒定区。依据特殊的链有一个或多个外显子。

11.5 CAAT 信号 (CAAT_signal)

CAAT 框;真核转录单元起始位点上游约 75bp 的一段保守序列,可能是 RNA 聚合酶的结合位点。保守区序列为 GG(C or T)CAATCT。

11.6 编码区 (CDS)

编码区;和一种蛋白质的氨基酸序列对应的核苷酸序列(包括终止密码子)。特征包括氨基酸的概念性翻译。

11.6.1 编码区页 (Coding Region Page)

在填写物种和序列表格 (Organism and Sequences form) 时,大多数用户会加上序列的编码区,但需要编辑编码区或填写附加项。选择在特征菜单 (Features menu) 下的编码区和转录子菜单 (Coding Regions and Transcripts submenu of the Features menu) 中的 CdRgn 或者双击记录浏览器编辑一个存在的 CDS。如果要在物种和序列表格 (Organism and Sequences form) 中添加一个编码区的部分序列,你应该编辑编码区特征,以避免产生错误的编码区位置信息。

11.6.1.1 常规子页 (General Subpage)

选择用于翻译核苷酸序列的遗传密码。要得到更多信息或翻译表,参考NCBI的“分类页”(taxonomy page)。

选择用于翻译的可读框。

点击“运行产物浏览器”(Launch Product Viewer)得到编码区的记录浏览器。

点击“编辑蛋白质信息”（Edit Protein Information）弹出蛋白质特征表格，添加蛋白质的其它信息。但是蛋白质的名称必须已经在蛋白质页上填过。

如果“接受”（Accept）的“重译”（retranslate）项已被选中，当你点击“接受”（Accept）时，Sequin 就会根据“定位页”（Locations page）指定的区域信息翻译核苷酸序列。这个新的翻译结果会替代以前提供的所有翻译结果。如果区域指定合适的话，这不是什么问题。但如果你不需要 Sequin 重新翻译序列时，不要选中这项。

如果提供的编码区序列是部分序列，就要选中“定位子页”（Location subpage）的“Partial”框。最好选择“Synchronize Partials”框，这时 Sequin 会使其它相关特征（例如蛋白质）也标记为“partial”。

11.6.1.2 其它选项子页（Exceptions Subpage）

其它选项描述没有转录后修饰的区域。输入发生转录后修饰的氨基酸位置，在弹出式列表中选择蛋白质中实际出现的氨基酸。Sequin 会将氨基酸数目转变为核苷酸距离。

11.6.1.3 蛋白质子页（Protein Subpage）

此页用于输入或编辑蛋白质产物的名称或说明。对于一个新序列，直接将信息输入框中。如果要编辑已经存在序列的说明，点击“编辑蛋白质信息”（Edit Protein Information），弹出“蛋白质特征表格”（Protein feature form）即可修改。

11.6.1.4 产物子页（Product Subpage）

如果要浏览序列，点击“产物”（Product）弹出式菜单中该序列的名称。你也可以通过选择文件（file）菜单下的“Import Protein FASTA”输入一个新的蛋白质序列，但序列必须符合上面“物种和序列表格”（Organism and Sequences form）所要求的格式。

输入一个蛋白质序列后，点击“Predict Interval”。这项功能会预测编码区指定核苷酸序列的区域。如果不选择这项功能，区域有可能是错的，当你确认记录时就会得到错误信息。如果序列的 5' 或者 3' 是不完整的，你必须首先在“定位页”（Locations Page）中手动指出。

点击“Translate Product”，Sequin 可以将核苷酸序列翻译为蛋白质序列。但 Sequin 会将核苷酸全长序列进行翻译，包括潜在的 5' 和 3' 非翻译区，这在生成记录时，有可能导致错误信息，因此要在“定位页”（Locations Page）中正确指定编码区，在“常规子页”（General subpage）中选择正确的可读框。

11.7 冲突（conflict）

相同序列的独立判断在这个位点或区域表现不同。

11.8 D-环（D-loop）

置换环；线粒体 DNA 内部的一个区域，在该区域一段短的 RNA 与 DNA 的一条链互补配对，将原来配对的 DNA 链置换出来。也用于描述由 RecA 蛋白催化的，其它单链 DNA 在双链 DNA 的某个区域侵入产生的置换反应。

11.9 D-片段（D_segment）

Diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain

免疫球蛋白重链和 T 细胞受体 β 链上的多样性片段。

11.10 增强子（enhancer）

一个顺式作用序列，可增强（一些）真核生物启动子的利用效率，增强子功能与其取向（可以正向或反向）和位置（可以在启动子的上游或下游）无关。

11.11 外显子 (exon)

编码剪接后 mRNA 部分的基因组区域。可能包括 5'非翻译区 (UTR)，全长编码区，和 3' 非翻译区 (UTR)。

11.12 GC_信号 (GC_signal)

GC 框；是一段保守的富含 GC 的区域，位于真核转录单元起始位点上游，可能是多拷贝的，与取向无关；保守区域为：GGGCGG。

11.13 基因 (gene)

生物学上关注的部分被确定为一个基因，名称已经被指定。

11.14 干涉 DNA (iDNA)

干涉 DNA；通过几种重组方式中的一种来消除 DNA。

11.15 内含子 (intron)

一段 DNA 片段；能被转录，但在翻译时被出去，而将其两边的序列（外显子）剪接到一起。

11.16 J_片段 (J_segment)

免疫球蛋白的重链和轻链之间，T 细胞受体的 α 、 β 、 γ 链之间的连接片段。

11.17 长终止重复 (LTR)

长终止重复；一段在特定序列两端直接重复的序列，比较有代表性的是逆转录酶病毒中发现的。

11.18 成熟肽 (mat_peptide)

成熟肽或蛋白质编码序列；经翻译后修饰的成熟的或最终的肽链或蛋白质的编码序列。

11.19 misc_binding

核酸与另外一半共价结合或非共价结合的位点，但不能用其它结合关键词（初级键或蛋白质键）来描述。

11.20 misc_difference

特征序列不同于条目中所出现的，也不能用其它任何差异关键词（冲突、不确定、原序列、突变、变异、等位基因或碱基修饰）来描述。

11.21 misc_feature

生物学上关注的部分，但不能用其它特征关键词来描述。

11.22 misc_recomb

发生普遍的、具有位点特异性的重复重组事件的位点，在该位点发生双链 DNA 的断裂和重接，但不能用其它重组关键词（干涉 DNA iDNA 和病毒体 virion）或适当来源关键词（插入序列 insertion_seq，转座子 transposon 或前病毒 proviral）来描述的。

11.23 misc_RNA

任一转录本或 RNA 产物，但不能用其它 RNA 关键词来描述 (prim_transcript, precursor_RNA, mRNA, 5'clip, 3'clip, 5'UTR, 3'UTR, exon, intron, polyA_site, rRNA, tRNA, scRNA, snoRNA 和 snRNA) 的。

11.24 misc_signal

包含控制或改变基因功能或基因表达信号的区域，但不能用其它信号关键词来描述 (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator 和 rep_origin) 的。

11.25 misc_structure

任何二级或三级结构或构型，但不能用其它结构关键词 (stem_loop 和 D-loop) 来描述的。

11.26 修饰的碱基 (modified_base)

特指修饰过的核苷酸，可以被特定分子替换（给定修饰碱基的合适值）。

11.27 mRNA

信使 RNA；包括 5'非翻译区 (5'UTR)、编码区 (CDS, 外显子) 和 3'非翻译区 (3'UTR)。

11.28 N_区 (N_region)

在重排的免疫球蛋白片段间插入的额外的核苷酸。

11.29 原序列 (old_sequence)

当前序列修正了这个位点以前序列的版本。

11.20 加尾信号 (polyA_signal)

为加上 polyA 尾，对 RNA 转录本进行剪切的核酸内切酶的必要识别区域。保守序列为 AATAAA。

11.31 polyA_位点 (polyA_site)

RNA 转录本上的位点，在转录后添加腺苷酸形成多聚腺苷酸。

11.32 前体 RNA (precursor_RNA)

没有成熟 RNA 产物中的任一 RNA 种类；可能包括 5' 发夹区 (5'clip)、5' 非翻译区 (5' UTR)、编码区 (CDS, 外显子)、插入序列 (内含子)、3' 非翻译区 (3' UTR) 和 3' 发夹区 (3' clip)。

11.33 初级转录本 (prim_transcript)

初级的 (起始的、未加工的) 转录本；包括 5' 发夹区 (5'clip)、5' 非翻译区 (5' UTR)、编码区 (CDS, 外显子)、插入序列 (内含子)、3' 非翻译区 (3' UTR) 和 3' 发夹区 (3' clip)。

11.34 引物结合位点 (primer_bind)

复制、转录、反转录的非共价结合位点。包括合成位点，例如 PCR 引物单元。

11.35 启动子 (promoter)

DNA 分子上 RNA 聚合酶结合起始转录的区域。

11.36 蛋白质结合位点 (protein_bind)

核酸上蛋白质非共价结合位点

11.37 核糖体结合位点 (RBS)

核糖体结合位点。

11.38 重复区 (repeat_region)

基因组中包含重复单元的区域。

11.39 重复单元 (repeat_unit)

单一重复成分

11.40 复制原点 (rep_origin)

复制原点；核酸复制 (得到相同的两份拷贝) 的起始位点。

11.41 rRNA

成熟的核糖体 RNA；核蛋白分子 (核糖体) 的 RNA 组分，可以将氨基酸组装成蛋白质。

11.42 转换区 (S_region)

免疫球蛋白重链的转换区。与重链 DNA 重排有关，可以导致同一 B 细胞的不同免疫球蛋白表达。

11.43 卫星序列 (satellite)

一个短的基本重复单元的多次串联重复（等同的或相近的）。多数含有一个基本成分或其它特性而不同于一般的基因组，因此可以使卫星序列与大多数基因组 DNA（主带）分开。

11.44 scRNA

细胞质小分子 RNA；真核细胞质或（有时）细胞核中出现的几种小的细胞质 RNA。

11.45 信号肽 (sig_peptide)

信号肽编码序列；分泌的蛋白质 N 末端结构域的编码序列，这个结构域可以使出生肽附着在膜上。前导序列。

11.46 核内小分子 RNA (snRNA)

核内小分子 RNA，与前体 RNA^d 剪接与加工有关。

11.47 核仁小分子 RNA (snoRNA)

核仁小分子 RNA，一般与核糖体 RNA (rRNA) 的修饰与加工有关。

11.48 来源 (source)

确定序列指定区域的生物学来源。这个关键字是必须有的，每一个条目都要有。最小的条目可以只有一个来源关键词而生成一条序列。每条序列可以有一个以上的来源关键词。

11.49 茎环结构 (stem_loop)

发夹结构；在单链 RNA 或 DNA 中，邻近的（反向）互补序列通过碱基配对形成的双链螺旋区域。

11.50 序列标签位点 (STS)

序列标签位点。短的、单拷贝的 DNA 序列，在基因组中作为图谱坐标，可以用 PCR 进行检测。基因组中的一段区域可以通过确定一系列 STS 的顺序而定位到图谱中。

11.51 TATA_信号 (TATA_signal)

TATA 框；Goldberg-Hogness 框；真核 RNA 聚合酶 II 转录单元起始位点前约 25bp 处发现的一个保守的富含 AT 的七碱基多聚体，可能与酶正确起始的结合位置有关。保守序列为 TATA(A or T)A(A or T)。

11.52 终止子 (terminator)

一段 DNA 序列，位于转录本的末端或与一个启动区相邻，可使 RNA 聚合酶终止转录，可能是抑制蛋白质的结合位点。

11.53 转运肽 (transit_peptide)

转运肽编码序列；核编码的细胞器蛋白质的 N 末端结构域的编码序列，这个结构域与翻译后蛋白质进入细胞器有关。

11.54 tRNA

成熟的转运 RNA，一种小的 RNA 分子（75-85 个碱基长），介导核苷酸序列翻译成氨基酸序列的过程。

11.55 不确定的 (unsure)

作者不确定这个区域的准确序列。

11.56 可变区 (V_region)

免疫球蛋白的轻链、重链，T 细胞受体的 α 、 β 、 γ 链上的可变区。编码部分可变氨基末端。由 V_segments、D_segments、N_regions 和 J_segments 组成。

11.57 V 片段 (V_segment)

免疫球蛋白的轻链、重链，T 细胞受体的 α 、 β 、 γ 链上的可变区。编码可变区 (V_region) 的大部分和前导肽的最后几个氨基酸。

11.58 变异 (variation)

属于同一基因但含有稳定突变的链 (例如 RFLPs, 多态性等等), 和这个位点 (也可能是其它位点) 的现有序列不同。

11.59 3' 发夹区 (3' clip)

前体转录本 3' 端的大部分区域, 在加工过程中被去掉。

11.60 3' 非翻译区 (3' UTR)

在成熟转录本的 3' 端或接近 3' 端, 通常在终止密码子之后, 不被翻译成蛋白质。尾部。

11.61 5' 发夹区 (5' clip)

前体转录本 5' 端的大部分区域, 在加工过程中被去掉。

11.62 5' 非翻译区 (5' UTR)

在成熟转录本的 5' 端或接近 5' 端, 通常在起始密码子之前, 不被翻译成蛋白质。头部。

11.63 -10_信号 (-10_signal)

Pribnow 框; 细菌转录单元起始位点上游 10bp 左右的一个保守区, 可能与 RNA 聚合酶结合有关。保守序列为 TAtAaT。

11.64 -35_信号 (-35_signal)

细菌转录单元起始位点上游 35bp 左右的一个六碱基多聚体保守区, 保守序列为 TTGACa or TGTTGACA。

12. 生物学来源 (Biological Source)

这个注释是非常重要的。一个条目只有含有序列来源物种的一些基本信息才能被数据库接受。这些基本信息在提交前物种和序列表格 (Organism and Sequences Form) 中输入。更详细的物种信息表格 (Organism Information form) 可以帮你改变或添加以前输入的数据。

12.1 概述: 描述符或特征 (Overview: Descriptor or Feature?)

Sequin 接受两类生物学来源信息: “生物学来源描述符” (Biological Source Descriptors) 和 “生物学来源特征” (Biological Source Features)。“生物学来源描述符” (Biological Source Descriptors) 像其它描述符一样, 提供一个条目中的一条完整序列或一套完整序列的物种信息。“生物学来源特征” (Biological Source Features) 像其它特征一样, 提供一个给定序列特殊区域的物种信息。

大多数情况下使用 “生物学来源描述符” (Biological Source Descriptors), 因为条目中所有序列来源相同。然而, 如果对一个妄想的分子测序, 例如部分是酵母的, 部分是小鼠的, 就应该使用 “生物学来源特征” (Biological Source Features) 来注释哪些序列来自酵母, 哪些序列来自小鼠。

要添加一个 “生物学来源描述符” (Biological Source Descriptors), 选择 “注释菜单” (Annotate menu) “描述符” (Descriptor) 部分上的 “生物学来源” (Biological Source)。要添加一个 “生物学来源特征” (Biological Source Features), 选择 “注释菜单” (Annotate menu) “书目和注解” (Bibliographic and Comments) 部分上的 “生物学来源” (Biological Source)。

注释一个 “生物学来源描述符” (Biological Source Descriptors) 或一个 “生物学来源特征” (Biological Source Features) 是与注释任何 “描述符” (Descriptors) 和 “特征” (Feature) 相同的。在生成 “描述符” (Descriptors) 和 “特征” (Feature) 时, 可以在帮助文档中找到相应的帮助。以下用于指导填写生物学来源专用表格。

12.2 物种页 (Organism Page)

12.2.1 名称子页 (Names Subpage)

这个滚动列表包含有许多物种的科学名称。为找到列表中的一个名称，可以键入科学名称的前几个字母，或使用拇指条查找。点击列表中的一个名称填满科学名称栏。如果物种有一个普通名称，该栏会被自动填充。你也可以直接键入科学名称。如果对物种的科学名称或普通名称有疑问，可以参考NCBI网站上的“分类浏览器” ([taxonomy browser](#))

12.2.2 定位子页 (Location Subpage)

12.2.2.1 序列位置 (Location of Sequence)

通过选择列表，请输入含有序列的基因组位置。大多数条目会选择基因组“Genomic”。弹出式菜单中选项的简短描述在前面列出。

12.2.2.2 序列来源 (Origin of Sequence)

此菜单是数据库专职人员使用的，不用填写。

12.2.3 遗传密码子页 (Genetic Codes Subpage)

12.2.3.1 细胞核的 (Nuclear)

利用此栏选择用于核酸序列翻译的遗传密码。真核生物的遗传密码是标准的“Standard”。如果你已经在上面的滚动列表中选择了一个物种的名称，此栏会自动填充。

要得到关于翻译表的更多信息，请参考NCBI的分类页 ([taxonomy page](#))。

12.2.3.2 线粒体的 (Mitochondrial)

不用选择此项。

12.2.4 谱系子页 (Lineage Subpage)

此页信息通常由数据库管理人员填写，他们利用NCBI/GenBank维持的“分类数据库” ([taxonomy database](#)) 来完成。如果你要输入一个与NCBI数据库不同的分类系统，请在这里填写。

如果在“登陆网络” ([network-aware](#)) 模式下运行Sequin，会有一个标有查找分类“Lookup Taxonomy”的按钮。点击此按钮，可执行物种分类系统的自动查询。Sequin会进入NCBI的“分类数据库” (Taxonomy database) 执行查询，并填写“分类系统与分类栏” (Taxonomic Lineage and Division)。

如果对Sequin的分类系统有任何评述，请在你登陆时一同提交。在Sequin的“文件菜单” (File menu) 中，选择“编辑提交者信息” (Edit Submitter Info)。在“提交页” (Submission page) 中题头标有“对数据库管理人员的特殊说明” (Special Instructions to Database Staff) 的对话框输入你的评述。当收到你的提交后，NCBI会有专人和你联系。

12.3 限定词页 (Modifiers Page)

此页是用于输入有关来源和/或物种更多信息的。输入信息是选择完成的。

12.3.1 来源子页 (Source Subpage)

从该页左侧下拉菜单中选择一个限定词，在页的右侧输入适当的名称。如果在滚动列表中没有发现合适的限定词，可以在该页底部栏中用文本格式输入其它来源信息。限定词可以选择多项。

以下是通用修饰词的描述：

- 细胞系 (Cell-line)：序列来源的细胞系。
- 细胞类型 (Cell-type)：序列来源的细胞类型。
- 染色体 (Chromosome)：基因定位的染色体。

- 克隆 (Clone): 获得序列的克隆名称。
- 克隆-库 (Clone-lib): 获得序列的文库名称。
- 国家 (Country): 用于流行病或种群研究的 DNA 样品来源的国家。
- 发育阶段 (Dev-stage): 物种发育阶段。
- 内病毒名称 (Endogenous-virus-name): 非活性病毒名称, 病毒已经整合到寄主染色体上, 因此将表现出垂直传播。
- 环境样品 (Environmental-sample): 序列是从不确认物种中直接通过分子手段分离。
- 频率 (Frequency): 特征发生频率。
- 基因型 (Genotype): 物种的基因型。
- 种系 (Germline): 序列表明是 DNA, 且是免疫球蛋白家族的一个成员。这个限定词用来表示序列来自非重组 DNA。
- 单倍体 (Haplotype): 物种的单倍体形式。
- 隔离来源 (Isolation-source): 描述序列来源物种的当地地理来源。
- 实验室宿主 (Lab-host): 序列来源物种的实验室繁殖宿主。
- 图谱 (Map): 基因在图谱上的位置。
- 质粒名称 (Plasmid-name): 获得序列的质粒名称。
- 质体名称 (Plastid-name): 获得序列的质体名称。
- 群体变种 (Pop-variant): 获得序列的群体变种名称。
- 重组的 (Rearranged): 序列表明是 DNA, 且是免疫球蛋白家族的一个成员。这个限定词用来表示序列来自重组 DNA。
- 片段 (Segment): 断裂成两个或更多个核酸分子的病毒基因组名称。
- 性别 (Sex): 序列来源的物种性别。
- 亚克隆 (Subclone): 获得序列的亚克隆名称。
- 组织文库 (Tissue-lib): 获得序列的组织文库。
- 组织类型 (Tissue-type): 序列来源的组织类型。
- 可转基因的 (Transgenic): 可以作为转基因受体的物种。

12.3.2 物种子页 (Organism Subpage)

从该页左侧下拉菜单中选择限定词, 并在右侧键入适当的名称。如果在滚动列表中没有发现合适的限定词, 可以在该页底部栏中用文本格式输入其它物种信息。限定词可以选择多项。

以下是通用修饰词的描述:

- 首字母缩写 (Acronym): 基于正式名称词首的标准首字母缩写 (通常用于病毒)
- 无性型 (Anamorph): 用于真菌无性阶段的科学名称。
- 权威 (Authority): 获得序列来自物种名称的创造者。
- 生物型 (Biotype): 参考 biovar。
- 生物类型 (Biovar): 物种 (通常是真菌、细菌或者病毒) 的种类。通过一些特别的生物学特性 (常常是地理上的、生态上的或者是生理上的) 来划分的。与 biotype 相同。
- 品种 (Breed): 获得序列的特定种类 (通常是指驯养的哺乳动物)。
- 化学类型 (Chemovar): 物种 (通常是真菌、细菌或者病毒) 的种类。通过其自身的化学特性来划分的。
- 通用名称 (Common): 获得序列物种的通用名称。

- 栽培种 (Cultivar): 获得序列植物的栽培品种。
- 剂量 (Dosage): 不用选择此项。
- 生态型 (Ecotype): 获得序列的特定生态型 (适应当地环境的群体), 通常指拟南芥的群体。
- 型 (Forma): 获得序列物种的型 (命名代码中的最小的分类单位)。这个术语通常用于植物和真菌。
- 专型 (Forma-specialis): 获得序列的生理上的特殊类型。通常限指某几个寄生真菌。
- 群 (Group): 不用选择此项。
- 隔离 (Isolate): 获得序列特殊个体的描述或确认。举例来说, X14 号病人。
- 自然宿主 (Natural Host): 提交序列来自一个物种, 这个物种又与有些第二物种存在共生、寄生或其它特殊关系。这个自然宿主 “natural host” 限定词可以被用于确定宿主种类的名称。
- 原名 (Old name): 不用选择此项。
- 病原体类型 (Pathovar): 物种 (通常是真菌、细菌或者病毒) 的种类。通过病原菌的生物学靶向来划分的。例如: 番茄病原体类型的假单胞菌 (*Pseudomonas syringae pathovar tomato*) 和烟草病原体类型的假单胞菌 (*Pseudomonas syringae pathovar tabaci*)。
- 血清组 (Serogroup): 参考 serotype。
- 血清类型 (Serotype): 物种 (通常是真菌、细菌或者病毒) 的种类。通过其抗原特性来划分的。与 serogroup 和 serovar 相同。
- 血清类型 (Serovar): 参考 serotype。
- 样品证明人 (Specimen-voucher): 原始物种的个体或群体以及当前保存地的证明人, 通常是一个公共机构。
- 株系 (Strain): 获得序列物种的株系。
- 亚组 (Subgroup): 不用选择此项。
- 亚种 (Sub-species): 获得序列物种的亚种。
- 亚株系 (Substrain): 获得序列物种的亚株系。
- 亚类型 (Subtype): 获得序列物种的亚类型。
- 同义词 (Synonym): 获得序列物种名称的同义词 (可替换的科学名称)。
- 有性型 (Teleomorph): 真菌有性阶段的科学名称。
- 类型 (Type): 获得序列物种的类型。
- 品种 (Variety): 获得序列物种的品种。

12.4 综合页 (Miscellaneous Page)

12.4.1 同义词子页 (Synonyms Subpage)

如果来源物种有可替换的名称, 在这里输入。请注意: 这只是适合填写物种的可替换名称的栏, 而不适用基因或蛋白质可替换的名称。

12.4.2 对照子页 (Cross-refs Subpage)

这页是用于对照本地条目和外部数据库 (GenBank, 、EMBL/EBI和DDBJ以外的数据库) 的条目, 例如dbEST或者是FLYBASE。大多数用户不用填写这页。要得到这个主题的更多信息, 请参考“国际核苷酸序列数据库协作页”

([International Nucleotide Sequence Database Collaboration page](#))。

13. 出版物 (Publications)

13.1 概述: 描述符或特征 Overview: Descriptor or Feature?

Sequin 接受两类出版物信息: “出版物描述符”(Publication Descriptors)和“出版物特征”(Publication Features)。“出版物描述符”(Publication Descriptors)是参考文献,像其它描述符一样,包括一个条目中的一条完整序列或一套完整序列的参考文献。“出版物特征”(Publication Features)也是参考信息,像其它特征一样,包括一个给定序列特殊区域的参考信息。

在数据库条目中的“参考”(Reference)栏输入出版物。参考是引用的和提交序列有关的未发表的、印刷中的或已出版的论著。出版物应该提供序列记录中关于克隆原理和序列判定的信息。

一般来说,当用一个出版物描述序列时,应使用“出版物描述符”(Publication Descriptors)。为输入一个“出版物描述符”(Publication Descriptors),选择“注释”(Annotate)菜单中的“出版物”(Publication),单击“出版物描述符”(Publication Descriptors)。

然而,当一个出版物描述基因的5'端克隆,另一个出版物描述基因的3'端克隆时,应使用“出版物特征”(Publication Features)。为输入“出版物特征”(Publication Features),首先选择“注释”(Annotate)菜单中的“出版物”(Publication)部分的“出版物特征”(Publication Features),将出版物输入记录。接着输入出版物的信息,最后指定出版物涉及的序列区域(在定位页)。

为尽量降低数据库的容量,我们不鼓励轻易在条目中使用“出版物特征”(Publication Features)。

如果你计划将参考文献添加为发表刊物的文章,要在网络状态运行Sequin。在这种模式下,如果已经提供了一些基本信息,程序会通过查询Medline数据库自动填写文章“名称”(Title)、“作者”(Author)和“刊物页码”(Journal page)。Sequin上网的“指导”(Instructions)在“综合菜单”(Misc menu)的文档中给出。下面是在Medline数据库中执行查询的“指导”(Instructions)。

“出版物描述符”(Publication Descriptors)或“出版物特征”(Publication Features)的注释和“描述符”(Descriptors)或“特征”(Features)的注释很相似。生成“描述符”(Descriptors)或“特征”(Features)的帮助,请参考帮助文档的适当章节。以下是填写出版物表格的细节指导。

13.2 注册表中引用项 (Citation on Entry Form)

13.2.1 状况 (Status)

利用单选按钮,从下面三项中选择一项:

- 未发表: 选择选项(1)没有计划发表稿件来描述提交序列;(2)稿件已经写完,但未提交;(3)稿件已经提交,但未接受。
- 印刷中: 文章已接受,但出版物还未印出。
- 已出版: 文章已经出版。

13.2.2 类别 (Class)

利用单选按钮,选择序列出现的出版类型:

- 定期刊物
- 书籍章节
- 书籍
- 论文/专论
- 会议记录章节: 会议摘要

- 会议记录：会议
- 专利
- 投稿

13.2.3 范围 (Scope)

请利用单选按钮选择一个或多个选项。

- 涉及完整序列：引用中列出的出版物描述了这个条目的完整序列。最常见的是在出版物中序列被首次提到。
- 涉及部分序列：引用中列出的出版物只描述了序列的一部分。当序列在一段时间内得到的只是片段时，选择此项。
- 引用序列特征：引用中列出的出版物描述或指代序列中发现的一个特征（例如一类启动子或一个基序）。

当你填写完注册表中引用项后，点击下一步“Proceed”进入下一个表格。

13.3 引用信息格式 (概要) (Citation Information Form (General))

在每页的底部有两个选项：接受和取消。点击"Accept"，表格内输入的信息将替换所有存在的引用。点击"Cancel"将取消改变引用的操作。

13.3.1 作者页 (Authors Page)

13.3.1.1 名字子页 (Names Subpage)

请输入作者的名字。注意第一作者列第一位。你可以加入多个作者到该页。当你输入第三个作者的名字后，指令组成为数据表格，你可以用翻阅条滚动到下一行。

13.3.1.2 联系子页 (Affiliation Subpage)

请输入主要作者所属机构的信息。

依据“输入条目引用” (Citation on Entry Form) 选择的出版物“分类” (Class) 不同，“引用信息表” (Citation Information Form) 的其它页是不同的。这里介绍填写期刊的“引用信息表” (Citation Information Form) 的指导。

13.4 引用信息格式 (如果选定类型为杂志) (Citation Information Form (If selected Class was Journal))

13.4.1 标题页 (Title Page)

在指令栏中给原稿输入标题。

13.4.2 期刊页 (Journal Page)

把正确的期刊名、卷数、出版社、页数、日期和年份信息输入到指令栏内。在弹出菜单中选择月份。如需要，从“拼写检查” (Erratum) 弹出式菜单中选择并解释错误项。如果知道“Medline 唯一标志符检索号” (Medline Unique Identifier, MUID)，请输入。

如果在“登陆网络” (network-aware) 模式下运行Sequin，提供一些必要的信息，程序就可从Medline数据库查找题目、作者和期刊名。例如，在相应的栏中输入已知的出版MUID，选择“Lookup By MUID”，Sequin会自动查找并填写其余信息。查找出版物MUID的一种方法是查找NCBI中Entrez的参考文献。如果不知道MUID，输入期刊名、卷数、页数和年份，并选择“Lookup Article”，Sequin会找到题目和作者的信息。

13.4.3 备注页 (Remark Page)

该页仅限于数据库成员使用

14. 文件菜单 (File Menu)

14.1 关于 Sequin (About Sequin)

介绍当前版本 Sequin 的详细资料。

14.2 帮助 (Help)

打开帮助文档。

14.3 打开 (Open)

打开一个已有的条目。该命令可打开在 Sequin 中已经保存的一个记录。若要进行分析，还可打开一个 FASTA 格式的序列文件。这个序列将在 Sequin 内打开，并可以使用 CDD BLAST 工具分析，但不能被提交，因为没有正确的注释符。

14.4 FASTA 格式的序列直接打开到序列编辑器 (FASTA Nucleotide Direct to Sequence Editor)

Opens a FASTA formatted sequence directly into the Sequence Editor. The sequence can then be analyzed with tools such as the "Find" command.

打开一个FASTA格式的核酸直接进入序列编辑器。这个序列可用"Find"指令等工具加以分析。

14.5 关闭 (Close)

关闭该条目

14.6 Export GenBank (从 GenBank 中输出)

将 GenBank 原始文件格式输出到一个文件

14.7 复制浏览 (Duplicate View)

复制条目进行浏览。你可以同时以不同的显示格式查看这个项目。

14.8 保存 (Save)

保存条目。注意：保存项目仅仅是为了以后编辑它。不会提交到数据库，也就是该条目不能生效。

14.9 另存为 (Save As)

参考 Save 项。

14.10 恢复 (Restore)

用以前保存的文本替换当前的记录。记录上次保存后，如果你不想保存所做改变时可以使用该项。

14.11 准备提交 (Prepare Submission)

准备把该条目提交给数据库。参见Sequin帮助文档的“提交最终记录到数据库” (Submitting the finished record to the database) 部分

14.12 打印 (Print)

打印当前选定的窗口。选定的窗口可以是 Sequin 的表格、页或帮助文档的一部分。

14.13 退出 (Quit)

退出 Sequin。

15. 编辑菜单 (Edit Menu)

15.1 复制 (Copy)

复制选定项目

15.2 清除 (Clear)

清除选定项目

15.3 复制副本 Duplicate

对选定部分复制副本。

15.4 编辑序列 (Edit sequence)

要编辑单个序列，在目标序列弹出菜单中选择序列标识符并点击编辑序列，就可以启动序列编辑器。关于 [sequence editor](#) 的详细说明见下面。

15.5 编辑队列 (Edit alignment)

如编辑一批队列序列，在目标序列弹出菜单中选择“所有序列” (All Sequences) 项，并在“显示格式” (Display Format) 中选择Alignment项。通过点击序列条周围的视窗来突出显示“队列” (alignment)，然后点击Edit alignment项启动队列编辑器。关于 [alignment](#) 编辑器的更多详细说明见下面。

15.6 编辑提交者信息 (Edit Submitter Info)

打开“提交指示表格” (Submission Instructions form)，该表格可以输入提交记录人的附加信息。多数信息是在 Sequin 的第一表格，“提交作者表格” (Submitting Authors form) 内输入。

也可以通过这里的提交作者表格 (Submitting Authors form) 保存相关信息，以便能在随后 Sequin 提交中用到。点击文件菜单下的“编辑提交者信息” (Edit Submitter Info)，打开“提交指示表格” (Submission Instructions form)，点击“输出提交者信息” (Export Submitter Info) 将信息存到一个文件。如果已经保存了提交者信息，那么在随后的 Sequin 提交时，通过点击提交作者表格

(Submitting Authors form) “提交页” (Submission page) 文件 (File) 菜单下的“输入提交者信息” (Import Submitter Info)。

15.6.1 提交页 (Submission page)

说明提交类型。如果是新的提交，选择“新建” (New)。如果你想更新原有记录而重新提交到数据库，选择“更新” (Update)。通过选择“**Yes**”或“**No**”的单选按钮来说明在发表之前是否公开。如果选择“**Yes**”，数据库成员把它加入到数据库后就被公开。如果选择“**No**”，你需要在相应栏中填写序列将被公开的日期。提交将会被数据库成员秘密保存，当序列正式发表、在 GenBank 收录号正式发表或到达公开日期时 (无论哪一个先出现) 再公开。如果你有其他特殊要求，在页底部栏中输入。

15.6.2 联系页 (Contact page)

更新记录提交者的名字、单位或联系方式

15.6.3 引用页 (Citation page)

为条目中发表这些序列而获得科学荣誉的工作者更新名字和单位。地址需要列出执行测序或序列分析的主要机构。如果多个实验室参与了这个项目，该页将记录主要作者的工作单位。如果你要提交记录作为数据库的更新，在“描述子页” (Description subpage) 中解释更新原因。

15.7 更新序列 (Update Sequence)

该项允许你用另外一个序列替换现有序列，利用末端重叠合并两个序列，修补添加一段正确的片段到当前序列或将一个序列的特征复制到另外一个序列。

用“读取FASTA文件” (Read FASTA file) 输入一个FASTA格式的序列。用“读取序列记录” (Read Sequence Record) 输入一个ASN.1 格式文件 (例如 Sequin已经保存的序列记录)。如果你在“登陆网络模式” ([Network Aware mode](#)) 下运行Sequin格式下，你可以从Entrez中利用“下载收录号” (Download Accession) 输入序列。如果你已经完成QBLAST搜索，你可以从在“图像或队列” (Graphic或Alignment) 视窗下已经选定的队列中通过“选择队列” (Selected Alignment) 输入序列。

无论怎样，输入序列与原始序列要有足够的相似区才能把这两个序列组成一个队列。如果相似性很低而不能产生队列，Sequin 将通知你相似性太低而不能完成更新。

当输入更新序列后，将打开一个新的窗口，包括两个图表视图和新老序列队列文本。第一个图表显示两个序列的相对长度和两个序列重叠区的长度。第二个图表显示在新老序列重叠区内的任何插入、缺失或点突变。点击图表内区域将在文本下面滚动出对应的队列核酸序列。

队列关系栏表示更新序列的方式。例如替换、扩展 5' 端、扩展 3' 端或修补作用。修补功能可以替换序列的内部片段而对两翼区域没有影响。

当利用“下载收录号”（Download Accession）或“读取序列记录”（Read Sequence Record）方式更新时，更新操作框需要你明确是序列还是特征被更新。因为新老序列含有相同特征，你可以选择保留新特征或者老特征或合并新老特征。

15.8 特征传递（Feature Propagate）

这个选项可以把一个序列注释的“特征”传递给同队列组的其它序列。例如，如果队列组内一个核酸序列有 CDS 的“特征”，对于同组内间隔相同区段的其它核酸序列，你可以注释一个相似的 CDS。

默认的将被传递的“特征”的来源是一组中的第一个成员。如果你希望利用一个不同的条目作为“特征”来源，在选择“编辑菜单”（Edit menu）“特征传递”（Feature Propagate）之前在“目标序列菜单”（Target Sequence menu）内定位该序列。

“特征传递”（Feature Propagate）窗口允许你选择哪些“特征”要被传递，这些特征在队列的缺口处是否将被扩展或分裂。在缺口处选择分裂将产生两个特征，队列内缺口两侧各一个。如果要传递一个 CDS 特征，你需要定义翻译末端或扩展通过内部终止子。也可以在“特征”来源条目中，通过选择通过不完整 3' 端翻译 CDS 而延伸通过终止密码子进行翻译。

15.9 添加序列（Add Sequence）

这个选项可以在已有的群体、突变、系统发育或环境样本集合中加入一条新序列。你可以 FASTA 或 ASN.1 格式（如在 Sequin 中保存的序列）输入一个新条目。

16. 搜索菜单（Search menu）

16.1 查找 ASN.1（Find ASN.1）

利用这条指令，你可以查找或替换在一些区域提交的成串字符，包括人工输入数据。可以更改的字符包括座位、定义、收录、关键词、来源、参考或特征。利用该项，要选择“查找”（Find）项并用正确的文本填写查找替换行。注意这种方式不能用于编辑序列。

16.2 查找原始文件（Find FlatFile）

利用这条指令，你可以查找在所有领域提交的成串字符。

16.3 按基因查找（Find by Gene）

这个选项可以在记录浏览器中快速移动到含有特殊基因符号的基因特征。

16.4 按蛋白质查找（Find by Protein）

这个选项可以在记录浏览器中快速移动到含有特殊产物名称的 CDS 特征。

16.5 按位置查找（Find by Position）

这个选项可以在记录浏览器中快速移动到注释为指定核苷酸位置的任意特征。

16.6 验证 (Validate)

该项用于检查你提交的格式与数据库要求的输入格式之间的差异。如果有差异，它将提示给你改正的方式。参见Sequin帮助文档的“提交最终记录到数据库” ([Submitting the finished record to the database](#))。

16.7 拼写检查 (Spell Check)

根据 PubMed 数据库对记录进行拼写检查。

16.8 保守域比对 (CDD BLAST)

依据NCBI的保守区数据库 ([Conserved Domain Database](#)) 对选择序列进行CDD比对。要进行CDD BLAST, Sequin必须在“登陆网络” (network aware) 模式下。

CDD 目前包含的结构域主要来自两个收集站: Smart 和 Pfam, 也有 NCBI 研究人员的贡献。来源数据库也提供描述和引用链接。因为保守域对应的是简洁结构单元, CDs 随时可能通过 Cn3D 包含与 3D 结构的链接。

CDD 的搜索结果将在记录浏览器中显示。这些结果仅供使用, 在提交前将被删除。

16.9 ORF 探测器 (ORF Finder)

ORF探测器以图表的形式显示核酸内所有的开放读码框。通过这个工具可以选择 ORFs, 而且 ORFs 的序列记录特征表现为编码序列。

ORFs, 以带颜色的框标示, 是指从起始码子开始到终止子结束的最长核酸序列。如果一整条核酸序列是一个开放读码框, 但不包括前端起始密码子或末端终止种植子, 它也将被作为 ORF。所有六个读码框被显示, 上面三个代表正链, 下面三个代表负链。ORFs 的核酸序列间隔在窗口右侧以递减长度顺序显示。互补 (负) 链的间隔以 a c 表示。选择 ORFs 可以直接点击它们或序列间隔。ORF 长度按钮可以选择显示的 ORF 的长度。例如, 默认值为 10 时, 显示所有不超过 10 核苷酸长度的 ORFs。点击标有 ORF 的框可以改变显示方式。潜在的起始密码子标为白色, 终止密码子标为红色。也可以在这种显示选择 ORFs。起始密码子和终止密码子的界定依据其选择的遗传密码。在启动 ORF Finder 之前要确保为翻译的序列选择了合适的遗传密码。

ORF探测器与序列编辑器联合工作。一旦选定了一个ORF, 这个序列就在编辑器内突出显示。使用序列编辑器的工具, 你可以把突出显示序列转换为CDS, 翻译并在记录里以CDS特征保存它。参见“序列编辑器” ([Sequence editor](#)) 内的[Editing a CDS](#)文件。

16.10 选择对象 (Select Target)

这项用于更改目标序列弹出窗口选定的序列。在指定栏中键入序列的 SeqID, 记录浏览器将更新以显示这条序列。

17. 综合菜单 (Misc Menu)

17.1 样式管理器 (Style Manager)

样式管理器可以使你选择不同形式的图表显示格式 (Graphical Display Format)。通过“记录浏览器” (record viewer) 上的图表显示格式 (Graphical Display Format) 选择图表显示。利用样式管理器, 你可以复制或修改形式以达到你的要求。

17.2 网络配置 (Net Configure)

在默认情况下, Sequin 可在单机工作 (stand-alone) 模式运行。但 Sequin 也可通过因特网与 NCBI (GenBank) 交流信息。Sequin 的“登陆网络”(network-aware) 模式与单机工作 (stand-alone) 模式基本一致, 但包括更多有用的选项。

只有在计算机与网络直接连接时, Sequin 的“登陆网络”(network-aware) 模式才能使用。以电子邮件登陆因特网是无效的。一般来说, 如果你能安装和使用“网页浏览器”(WWW-browser), 就能安装和使用“登陆网络”(network-aware) 模式的 Sequin。如果不确定是否与网络直接相连, 资源系统管理员或网络服务商。

有两种方法可以把 Sequin 转换到“登陆网络”(network-aware) 模式。如果你在最初的“Sequin 欢迎界面”(Welcome to Sequin form) 模式下, 选择“综合”(Misc) 菜单下的“网络设置”(Net Configure)。如果你已经在“Sequin 提交”(Sequin submission) 模式下工作, 并在利用记录浏览器查看记录, 选择“综合”(Misc) 菜单下的“网络设置”(Net Configure)。

大多数用户可以直接使用“网络设置页”(Network Configuration page) 的默认(常规)设置, 选择 Accept 完成配置过程。

如果默认连接不起作用, 可以使用防火墙连接(Firewall Connection)。联系你的系统管理员输入相应“代理服务器”(Proxy)和“端口”(Port)。如果你不能登陆“域名服务器”(DNS), 不要选择此框。

“间隔时间”(Timeout)弹出菜单用于设定 Sequin 从 NCBI 服务器获得答复等待的时间长度。如果你在美国之外或网络连接不好, 需要设定更长一些时间(例如 60 秒或 5 分钟)。

如果建立网络配置有问题, 请联系 [mailto: info@ncbi.nlm.nih.gov](mailto:info@ncbi.nlm.nih.gov)

如果想让 Sequin 改回单机工作 (stand-alone) 模式, 从“综合菜单”(Misc menu) 中选择“网络设置”(Net Configure), 点击: None.

Sequin 在“登陆网络”(network-aware) 模式下附加有一些重要功能。这些功能将作为附加菜单条目出现。这些功能的简短描述见下面。在帮助文档有更多功能描述。

17.2.1 更新 GenBank 已有记录 (Updating existing GenBank records)

Sequin 在“登陆网络”(network-aware) 模式下, 可以利用 GenBank 收录号或 GI 身份证号 (NID) 从 Entrez 中下载 GenBank 已有记录。那样就可以用 Sequin 对记录做一些必要的修改并作为更新序列重新提交到 GenBank。关于提交更新序列的指导 ([Instructions](#)) 在“Sequin 欢迎界面”(Welcome to Sequin Form) 下。你可以从 Entrez 下载任何记录并在 Sequin 下查看, 但只能对自己以前提交的记录做有效的数据库更新。

17.2.2 在 Medline 中检索 (Performing a Medline lookup)

在“登陆网络”(network-aware) 模式下, Sequin 可以把一条 Medline 记录的相关部分直接输入到序列提交记录内。只需输入一些必需信息, 例如: “Medline 唯一标志符”(Medline Unique Identifier, MUID) 或者是期刊名、卷数、年份或页数, 而不用输入完整引用。关于“Medline 检索”(Medline lookup) 可参见标有出版物的部分文档。

17.2.3 执行分类学查找 (Performing a Taxonomy lookup)

在“登陆网络”(network-aware) 模式下, Sequin 可以从 NCBI 的分类学数据库中查找物种的分类学谱系。一般情况下, 当记录已提交到 GenBank 后, 这种

查找是由NCBI数据库管理人员执行的。如果在提交前进行分类学查找，你可以在记录内不一致意见的部分作些标注。关于 [taxonomy lookup](#) 的介绍参见“生物学来源” (Biological Source)：物种页 (Organism pag)：谱系子页 (Lineage subpage) 的部分文档。

17.2.4 登陆 NCBI 桌面 (Accessing the NCBI DeskTop)

NCBI桌面显示Sequin中浏览记录的内部结构，关于桌面 ([DeskTop](#)) 的解释参见“综合” (Misc) 菜单。

17.3 Entrez 查询 (Entrez Query)

这项用于在NCBI的“Entrez 数据库” ([Entrez database](#)) 中执行查询。

如果从 Entrez 中下载了一个序列，在窗口的底部将出现 Entrez 链接按钮。如要查看相似的核酸序列，可以选择“目标弹出菜单” (Target pop-up menu) 下的“核苷酸” (Nucleotide)，然后点击左侧链接按钮获得 Entrez 相关序列列表。如要查看记录中文献或 CDSs 的 Entrez 相关记录，选择“Medline 或蛋白质数据库” (Medline or Protein database)，点击“查找” (Lookup) 按钮。有关基因组或结构的记录也可以浏览。

17.4 NCBI 桌面 (NCBI DeskTop)

只有在“登陆网络” ([network-aware](#)) 模式下，Sequin才可以使用该项。

NCBI 桌面提供 Sequin 记录的内部结构浏览——ASN.1，ASN.1 数据模型的所有结构以类似“维恩图”的形式表示出来。

此外，可以通过桌面获得NCBI的一些非正式软件工具。这些工具是NCBI软件包的组成成分。你可以用自己的软件工具来自定义软件包这些功能。可以通过匿名 [ftp: //ftp.ncbi.nih.gov/toolbox/README](ftp://ftp.ncbi.nih.gov/toolbox/README) 得到NCBI的软件包和它的说明。

桌面只是为经验十分丰富的使用者提供的。这里我们不提供这些特殊功能的文件说明。

18. 注释菜单 (Annotate Menu)

该菜单用于输入序列“特征” (features) 和“描述符” (descriptors)。

“特征” (features) 的前六个选项包括“基因定义区” (Genes and Named Regions)、 “编码转录区” (Coding Regions and Transcripts)、 “结构RNAs” (Structural RNAs)、 “文献和注释” (Bibliographic and Comments)、 “位置链接” (Sites and Bonds) 和“保留特征” (Remaining Features)。这六个选项的特征都能被添加到序列上。关于特征的更多描述见上面的“特征” ([Features](#)) 部分。

第七个选项为出版物，“出版物特征” (Publication Feature) 或“出版物描述符” (Publication Descriptor)。关于出版物的更多描述见上面的“出版物” ([Publications](#)) 部分。

第八选项描述符，用于添加记录的“出版物描述符” (Publication Descriptor)。关于描述符的更多描述见上面的“描述符” ([Descriptors](#)) 部分。

第九选项是“生成定义行” (Generate Definition Line)，根据记录提供信息为序列生成一个标题。这个选项对单条序列或序列组都起作用，并能对多个特征作出综合注释。这个标题遵循GenBank惯例，但如果不是很恰当，数据库管理人员可对其修改。在这里输入的标题将替代提交中他处的标题，例如和核苷酸序列相关的任一标题。关于“生成定义行” (Generate Definition Line) 的描述参见上面的“核苷酸定义行” ([Nucleotide Definition line \(title\)](#))。

19. 选项菜单 (Options Menu)

19.1 字体 (Font)

使用此项可改变显示字体。从弹出菜单中, 选择字体类型和大小。对于其他修饰项, 例如加黑、斜体或下划线指令组, 选中相应框。默认字体为 10 点

19.2 图例 (Legend)

只有在图表浏览时才能使用, 该项可以显示出图表中出现的各种特征, 也就是图表使用的颜色、形式和字体

20. 序列编辑器 (Sequence Editor)

序列编辑器对你条目中的核酸或氨基酸序列进行修改。例如, 可以添加或删除核苷酸或添加或删除条目的 CDS (编码序列) 特征。

虽然序列编辑器也可以取消你对序列所做的变动, 但我们强烈建议你在向序列编辑器提交序列前要保存一份拷贝, 以在需要时对其恢复。

20.1 运行序列编辑器 (Starting the Sequence Editor)

在序列编辑器中出现的序列取决于“目标序列弹出菜单”(Target Sequence pop-up menu) 中选择的序列。对于核酸序列可以采用两种方法运行序列编辑器。第一, 你可以双击“记录浏览器”(record viewer) 里任意格式的序列, 含有 DNA 序列的窗口将会出现。第二, 在“记录浏览器”(record viewer) 的“目标序列弹出菜单”(Target Sequence pop-up menu) 中选择你所需要的序列, 然后点击“编辑”(Edit) 菜单下的“编辑序列”(Edit Sequence)。对于蛋白质序列运行序列编辑器也同样可以采用两种方式。如果在“目标序列弹出菜单”(Target Sequence pop-up menu) 选择了蛋白质序列, 双击该蛋白质序列, 含有蛋白质序列的窗口将会出现。如果在“目标序列弹出菜单”(Target Sequence pop-up menu) 中选择了核苷酸序列, 在 CDS (编码区) 特征中双击打开“编码区特征表格”(Coding Region feature form) [参见 Sequin 帮助文档的“特征”(Features)]。点击“运行产物浏览器”(Launch Product Viewer) 打开序列编辑器。进入蛋白质序列编辑器后, 两种方法将显示相同的窗口。

20.2 序列编辑器的指针移动 (Moving around the Sequence Editor)

通过鼠标或键盘可以移动指针。显示窗口将调整指示指针位置。序列每行的第一个残基都将在窗口的左侧显示。指针的位置, 或鼠标选定序列范围, 将在窗口的左上角显示。如果你想将指针移动到一个特定位置, 将号码输入到序列编辑器窗口左上角的框内, 然后点击“转到”(Go to) 按钮。如果你想查找特定的序列, 且不需要移动指针时, 将号码输入到窗口右上方的框内, 点击“查找”(Look at) 按钮。

20.3 编辑序列 (Editing sequence)

选择一段序列并用鼠标突出显示。要选择完整序列, 点击窗口左侧的序列定位号码。在序列编辑器中突出显示的任何序列, 在以“图表显示格式”(Graphic Display Format) 查看时都将作为一栏显示。

一种插入或删除残基的方式是利用鼠标。移动指针到合适的位置并输入, 写入的文本将插入到指针左侧。删除序列可用退格键 (backspace) 或删除键 (delete key)。指针左侧文本将被删除。删除一段序列, 可以用鼠标选定后再用删除键或退格键删除。

另一种插入或删除残基的方式是利用序列编辑器内编辑菜单下的选项。选定残基后用剪切项 (Cut) 移出或用复制项 (Copy) 复制, 再用粘贴项 (Paste) 粘贴到任何位置。用清除项 (Clear) 可永久地清除选定序列。

为了保存对序列所做的更改, 点击序列编辑器显示窗口底部的接受 (Accept) 按钮。如果不想保存这些修改, 点击序列编辑器显示窗口底部的取消 (Cancel) 按钮。无论你选择接受按钮或取消按钮后都将退出序列编辑器, 返回到记录浏览器。请注意: 如果不在记录浏览器中保存, 任何修改都将不能成为 Sequin 记录的永久成分。序列编辑器显示窗口底部的 Save 按钮仅用来保存 CDS 特征。

20.4 更改显示 (Changing the display)

核酸序列的默认显示方式为编码链。如果要查看这个序列的互补链, 也就是你想查看这个序列的双链模式, 选择序列编辑器“视图”(View) 菜单下的“互补”(Complement) 项。你也可以选择查看此链的翻译结果。选择序列编辑器“视图”(View) 菜单下的 Reading Frames +) 项查看上面的(编码)链的三项翻译情况。选择序列编辑器视图菜单下的“阅读框-”(Reading Frames -) 项查看下面的(非编码)链的三项翻译情况。蛋氨酸残基用彩色标出。

Complement, Reading Frames +, and Reading Frames - 三项可同时或单独选用。

仅第一条编码链可以被编辑。任何更改都将在“互补链”(Complement) 或“读码框”(Reading Frames) 项中反映出。

20.5 编辑 CDS (编码序列) [Editing a CDS (coding sequence)]

序列编辑器的一个突出特点是可以在核酸序列上设计一个新的 CDS (编码区) 特征。如设计新的 CDS 特征, 选定这些残基, 再选择 Features-->Coding Regions 和 Transcripts-->CDS。这将打开 CDS 编辑器。突出显示序列的位置将自动添加到“定位页”(Location page)。访问“定位页”(Location page), 进入 Coding Region-->Protein in subpage 下的蛋白质名称, 并点击接受按钮。选择序列编辑器底部“显示特征”(Show Features) 选项查看彩色条状的 CDS。点击 CDS 区或 CDS 在页左侧空白区的名字都可以选中 CDS。要清除一个 CDS, 可以先选定再点击 Sequin 编辑菜单下的清除键。当 Sequin 问你是否想要删除相关蛋白产物时, 点击 yes 键。

你也可以用记录浏览器查找菜单下的“ORF探测器”(ORF Finder) 来查找编码序列。点击 ORF Finder 项查找序列的 ORFs (开放读码框)。关于 ORF Finder 的更多描述已在上面 (above) 提到。在 ORF Finder 中, 点击要添加到序列上的 ORF, 在序列编辑器中查看序列时, 其 ORF 将被突出显示。你可以通过上面所做的介绍[编码区 (Coding Region) (CDS) 特征表格]把序列转变成 CDS。最简单大方法是, 在“编码区页”(Coding Region page) 的“蛋白质子页”(Protein subpage) 中输入蛋白质的名称。对于更多指导, 参见上面的 CDS 特征。当你点击编码区表格的接受按钮后, Sequin 将接受 CDS 为一个新的特征。记录浏览器和其他窗口也将更新。CDS 的颜色将改为粉色。选择记录浏览器的“图表显示格式”(Graphic Display Format) 可以看到特征的图示。

保存的 CDS 特征也可以被编辑。按上面的描述可以改变已保存 CDS 的长度或位置。然而, 一个保存后的 CDS 不能在序列编辑器中去除。要去除一个已保存的特征, 进入记录浏览器的“图表显示格式”(Graphic Display Format), 选择 CDS 并选择编辑窗口下的清除键。

当在 CDS 的一个区域内添加或清除核酸序列时，你要选择 CDS 是否被这些更改所打断。在 Sequin 主窗口下，选择割裂（Split）特征模式是指插入或删除序列可以打断 CDS。选择合并（Merge）特征模式是使 CDS 与改变的序列合并。

20.6 处理多个队列序列（Working with sets of aligned sequences）

Sequin 可以处理密切相关的队列核苷酸序列，这些序列是群体、系统发育的或突变研究的一部分。如果序列以“前队列”（pre-aligned）格式输入，如 PHYLIP 格式，Sequin 就采用原有队列格式。如果序列以 FASTA 格式单个的输入，Sequin 将产生自己的队列。

你可以在“序列队列编辑器”（Sequence Alignment Editor）内浏览队列序列。在记录浏览器中，选择目标序列菜单下的“所有序列”（All Sequences），并选择“队列显示格式”（Alignment Display Format）。点击序列条周围的框突出显示队列。然后从编辑菜单中选择“编辑队列”（Edit alignment）。这个队列可以应用多种格式浏览。可以参见下面（[below](#)）序列编辑器队列菜单的介绍。

输入一系列核酸序列后，如果你要将 CDS（编码区）特征添加到队列中的一个或多个序列，必须先将 CDS 特征添加到一个序列上（参见上面（[above](#)）的 Editing a CDS）。如果要在序列编辑器中选取单个序列，双击“队列浏览”（Alignment view）内该序列的名称。接着可以将特征传递给其他序列（参见编辑菜单下的特征传递（Feature Propagate））。

20.7 产生新队列（Creating new alignments）

Sequin 可以输入一些序列，并将它们与原有的队列或单条序列合并成新队列。特征可以在队列序列中相互传递。序列可以来自一个文件也可以从 Entrez 上下载。（参见下面的序列编辑器的文件菜单）

20.8 序列编辑器文件菜单（Sequence Editor File menu）

20.8.1 读取序列文件（Read Sequence File）

从文件中导入 FASTA 格式序列或 ASN1 记录，有三种形成队列的方法。BLAST 提供最好的局部队列形成方法；扩展的 BLAST 提供一个基于 BLAST 局部队列形成的全局队列形成方法；全局队列形成方法通过动态编程算法来提供不同的全局队列形成方式。

20.8.2 从 Entrez 中下载序列（Download from Entrez）

利用收录号（accession number）或 gi 从 Entrez 中导入一个序列。

20.8.3 输入队列（Import Alignment）

以 Phylip、NEXUS 和 FASTA+Gap 格式输入序列队列。文件中的第一个序列与原记录序列具有相同的序列和序列标识符。

20.8.4 输出（Export）

以 Text 或 FASTA 格式输出单条序列，以 FASTA+Gap、Phylip 或 ASN1 格式输出队列序列。这项功能允许选择输出序列的范围。

20.8.5 接受更改（Accept Changes）

关闭序列编辑器，提交在序列编辑器中所做的所有更改。保存所作更改可选择 Sequin 主菜单中文件下的 Save 选项（File-->Save）。

20.8.6 取消（Cancel）

关闭序列编辑器，取消所作更改

20.9 序列编辑器编辑菜单（Sequence Editor Edit menu）

20.9.1 撤消（Undo）

撤消序列编辑器上次保存后的所有操作。

20.9.2 剪切 (Cut)

移动突出显示的序列，可以任意粘贴到其它位置。

20.9.3 粘贴 (Paste)

将剪切或拷贝的序列粘贴到鼠标指针右侧。

20.9.4 复制 (Copy)

复制突出显示的序列，可以粘贴其它任何位置。

20.9.5 更新 (Refresh)

通过重新载入数据更新窗口。注意这项不会撤消任何编辑操作。

20.9.6 删除序列 (Delete Sequence)

删除队列中选择的序列

20.10 序列编辑器浏览菜单 (Sequence Editor View menu)

20.10.1 浏览模式 (View mode)

该模式只能对序列进行浏览。

20.10.2 编辑模式 (Edit mode)

该模式可以对序列进行编辑。

20.10.3 标签 (Label)

用于对序列标识符进行更改。一般情况下，序列标识符仅以队列序列显示。如果从 Entrez 上下载序列的定义行中有不同的名字，可以更改可见的名字。你可以查看标有以下名字每个序列：FASTA short, FASTA long, Locus, Accession 或 Report。

20.10.4 字体 (Font)

选择字体。

20.10.5 选项 (Preferences)

Allows you to change the style of the display, including the colors, font type, and font size.

选择显示风格，包括颜色、字体和字号

20.10.6 互补链 (Complement)

在原始链下面显示提交链（原始链）的互补链

20.10.7 读码框 (Reading frames)

显示选定编码序列的同步翻译。你可以选择 6 个读码框的任意一个或全部。

20.10.8 翻译方式 (Translation Style)

用于选择浏览编码序列翻译结果的三种方式。默认方式是显示所有氨基酸。如果你选择***，Sequin 将把所有蛋氨酸残基显示为 M，所有终止密码子显示为*。如果你选择 orf 选项，Sequin 将显示所有的开放读码框，单个可读框的 M 和* 用一个“~”来连接。

20.10.9 查找 (Find)

查找命令可以在输入序列中查找 DNA 序列或氨基酸序列样本。这种查找是不区分大小写的。如要准确匹配查找 DNA 序列样本，将该样本输入栏中。你也可以采用不准确匹配查找的模式。如果要查找样本的反向互补序列，选中相应框。例如：

TCAGGGC 用于查找序列 TCAGGGC;

[TCA]CAGGGC 用于查找 T 或 C 或 A，且后面跟有 CAGGGC;

NCAGGGC 用于查找 T 或 C 或 G 或 A，且后面跟有 CAGGGC;

TCA(3)GC 用于查找 TCAGGGC;

TCA(1: 3)GC 用于查找序列 TCAGC、TCAGGC、TCAGGGC;

TCA(1: 3)NC 用于查找 TCA, 后面跟有 1-3 个任意碱基 (G、A、T 或 C), 且后面又跟有 C 的序列。例如: TCATC 或 TCATTC 或 TCAATGC。

为了查找与氨基酸样本准确匹配的序列, 在栏中输入该序列, 并点击“翻译序列”(translate sequence)项。Sequin 会查找三个正链开放读码框中所有可能的情况。开放读码框将被显示, 编码蛋白质的 DNA 序列将被突出显示。你也可以采用不准确匹配查找的方式。例如:

CDLPEYC 用于查找序列 CDLPEYC;

[CRQ]DLPEYC 用于查找 C 或 R 或 Q, 且后面跟有 DLPEYC;

XDLPEYC 用于查找任何氨基酸, 且后面跟有 DLPEYC;

CDL(3)EYC 用于查找 CDLEEEYC;

CDL(1: 3)PE 用于查找 CDLPE、CDLPPE 和 CDLPPPE;

CDL(1: 3)X 用于查找 CDL, 其后面跟有 1-3 个任意氨基酸, 且又跟有 C 的序列。例如: CDLAAE、CDLRSE 或 CDLAPQE。

20.10.9.1 查找前一个 (Find previous)

查看前面出现的样本。

20.10.9.2 查找下一个 (Find next)

查看后面出现的样本。

20.11 序列编辑特征菜单 (Sequence Editor Features menu)

浏览单个序列和浏览队列序列的特征菜单是不同的。

如果是浏览单个序列, 菜单包括可用于注释序列的所有“特征”列表。这些“特征”与“Sequin 注释主菜单”(main Sequin Annotate menu)中的是一样的。

你可以在“注释”(Annotate)菜单或“序列编辑器”(Sequence Editor)中注释序列“特征”。如果使用“注释”(Annotate)菜单来注释序列“特征”, 你必须提供“特征”相关的核苷酸序列的位置。如果你使用“序列编辑器”

(Sequence Editor)来添加“特征”就不需要知道相关的核苷酸序列。简单地突出显示一下特征适用的序列, 序列的位置将自动进入特征定位栏中。关于注释特征的更多解释可参见“特征”(Features)部分。

20.12 序列编辑器队列菜单 (Sequence Editor Alignment menu)

这些菜单选项只有在编辑队列时才能用到。当来自系统发生的、群体的或突变研究的一组序列提交时, 就会产生队列。也可以选择“序列编辑器编辑菜单”(Sequence Editor Edit menu)下的“Align with”选项, 将另外的参考序列输入到 Sequin 中而形成队列。

20.12.1 选择参考 (Select Reference)

选择队列序列中的一条作为主序列。在序列编辑器中点击所需序列, 然后选择“Select master”改变显示形式。默认情况下, 主序列将是列出的第一个序列。主序列的标识符将用彩色显示。

20.12.2 选择全部 (Select all)

选择所有核苷酸序列

20.12.3 显示替换/显示全部 (Show substitutions/Show all)

更改序列显示方式。当“Show all”被选择时(那时“Show substitutions”是可见的), 显示的是每个条目的完整序列。当选择“Show substitutions”时, 显示主序列的完整序列, 队列中的其它条目序列则是与主序列一致的序列以点的形式显示, 不一致的序列以字母显示出来。

20.12.4 选择变异区/保守区 (Select Variations/Conservation)

“Select Variations”用于突出显示和参考序列不同的区域，“Conservation”用于突出显示和参考序列一致的区域。

20.12.5 点矩阵 (Dot Matrix)

这个选项用于显示队列中选定两个序列的点阵列图示结果。该方法正在完善中。

20.13 序列编辑器窗口按钮 (Sequence Editor window buttons)

20.13.1 转到 (Go to)

移动指针到指定位置

20.13.2 查找 (Look at)

不用移动指针而将窗口移动到指定位置。

20.13.3 合并/分裂特征模式 (Merge feature mode/Split feature mode)

在合并模式中，输入到一个已有“特征”区的任何新的序列都将成为那个“特征”的一部分。例如，你在一个 CDS 中间输入新序列，这段序列将成为 CDS 的一部分。在分裂模式中，新序列将打断原有“特征”。例如，你在 CDS 中间输入新序列，该 CDS 将被这段新序列打断（可参见记录浏览器的 CDS 定位）。

20.13.4 隐藏/显示特征 (Hide feat./Show feat.)

隐藏/显示序列特征的触发开关。要隐藏特征，点击显示为“Hide feat”时的开关；要显示特征，点击显示为“Show feat”时的特征。

20.13.5 更新 (Refresh)

通过重新载入数据更新窗口。注意这项不会撤消任何编辑操作。

20.13.6 接受 (Accept)

关闭序列编辑器，保存对序列或特征的所有更改。

20.13.7 取消 (Cancel)

关闭序列编辑器，但对序列或特征不作任何更改。

问题和评述？

写信给 [NCBI Service Desk](#)

2003 年 10 月 20 日修订。