PC-Rare 用户手册

（Rémy Griffais MD PhD）

“八聚体频率出现的不一致性的方法寻找引物的最佳选择”

目录

简介
- 硬件配置.
- 安装 **PC-Rare**

使用 **PC-Rare**

菜单：
- Sequence 菜单包括：
- Works and Results 菜单：
- DNA context 菜单
- Windows 菜单 列出了工作屏幕的显示选项，包括 **PC-Rare** 的不同窗口：
- Sequence 窗口 （Sequence 菜单, sequence 窗口子菜单）.
- DNA Context 窗口,用来选择您想要操作的数据库
- User Oligonucleotide 窗口（Sequence 菜单 - User Oligonucleotide 子菜单）.
- printer setup 窗口 （sequence 菜单）
- Frequency 窗口 （sequence 菜单, Frequency 窗口子菜单）
- The parameters 窗口 （Works and results 菜单, Research 子菜单）
- The Results 窗口 （Work and Results 菜单, Results 子菜单）

在 **PC-Rare** 的打印结果中你可以获得什么？
- 在 sequence 窗口中
- 在 User Oligonucleotide 窗口中
- 在 frequency 窗口中
- Results printed from the results 窗口

感谢

参考书目

简介

PC-Rare 是为选择 PCR 的引物而设计的（Mullis *et al.*,1985; Saiki *et al*,1985; Saiki *et al.*,1988）。使用 **PC-Rare** 选择引物的方法在“K-tuple frequency in the Human genome and the Polymerase chain reaction （1991）”中有描述。请阅读这篇文章，它用简单的术语描述了这种新的非常有效的选择引物的

方法。最近的一篇文章（Chenal *et al.*, 1996）也描述了这种方法。**PC-Rare** 使用的这种方法被称为“Octamer Frequency Disparity （O.F.D，八聚体频率出现的不一致性的方法）”方法。The method used by **PC-Rare** is called the "Octamer Frequency Disparity" （O.F.D） Method. Please when using primers with rare octamer at their 3' end, use as citation （Griffais *et al.*, 1991）.

硬件配置

PC-Rare 同时提供 MS Windows 和 Macintosh 两种版本。每个版本分别适用于 Windows 31 和 Windows 95 或 Mac 和 Power-Mac。**PC-Rare** 的 Windows 版本在有 WABI 时可以用于 Sun 的 Solaris 的 Windows 仿真器。

在基于 Power-Mac 的计算机上，请将软件使用的内存调至 2M（默认为 512Kb）以获得最佳效果。 **PC-Rare** 可以操作很大的序列，所以在运行程序之前请保留足够的内存（参考您的 Macintosh 用户手册）。象其它一些科学程序一样，一个协处理器可以显著的提高某些操作的次数。

安装 **PC-Rare**

象从网上获得的其它软件一样，**PC-Rare** 和特定的数据库在使用之前必须解开。相关操作的信息可以在下载本软件的 FTP 上找到。

使用 PC-Rare

要运行 PC-Rare，打开 **PC-Rare** 文件夹，双击 **PC-Rare** 图标。**PC-Rare** 打开的窗口被载入到你的计算机屏幕上。要继续载入程序，请单击 "First Try..."按钮。

在 Macintosh 版本上，需要感谢我的朋友 Dr. Pierre Soubiran，他立即答应了本软件 Mac 版本的开发。Pierre 借了一台 Macintosh LC 计算机供我使用了几个月。

要运行本软件，也是通过双击程序的图标。这个程序被直接载入而不显示开始屏幕。下面几分钟我们请你查看…

菜单行:

Sequence，Works and results，DNA context Windows

Sequence 菜单包括:

- **Input** – 用来载入一个序列文件。对话框是标准的。**PC-Rare** 使用 TEXT（ASCII）格式的序列文件来工作。这种文件可以用大部分程序来创建。
- **Sequence** 窗口 –未被激活。当载入一个序列文件时被激活，可以用来显示 Sequence 窗口。
- **Frequency** 窗口 – 未被激活。当选择一个序列文件或一个数据库时被激活。用来显示 examining frequencies 窗口。
- **User oligonucleotide** –未被激活。当选择一个数据库时被激活。用来通过键盘编辑研究一个短的序列（约 30 个核苷酸）
- **About** – 关于软件的各种信息，联系作者的 E-mail 地址，对于 EUROGENTEC 以前的帮助的感谢。
- **Printer setup** – 一个标准的打印机调整对话窗口。

重要提示：
每次使用 **PC-Rare** 时，记住调整您的打印机。**PC-Rare** 以图象格式打印 A4 格式（210*297mm）的文件。

Works and Results 菜单 包括：

- **Research**–未激活。当载入一个序列文件或一个数据库（DNA Context）时才被激活。用来寻找 PCR 目标。

- **Results** 窗口–未被激活。在寻找 PCR 目标或是从文件载入一个结果时被激活。（参见下面的 Load results）.

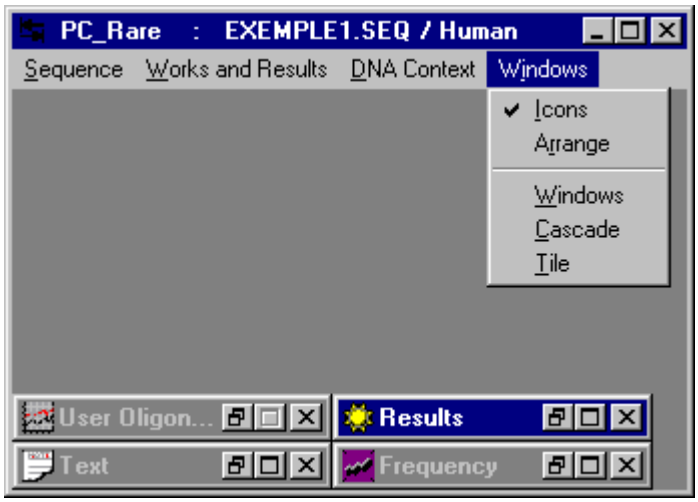
- **Save results**–未被激活。在找到 PCR 目标后被激活。显示一个标准的对话框来保存一个结果文件。保存的文件具有.RES 扩展名，可以通过 Load Results 在以后重新载入。

- **Load results** –用来载入通过 Save Results 命令保存在磁盘上的结果。

The DNA context 菜单

- **Choice** – 可以选择你想要操作的数据库
- **Statistics** - （仅限于 Macintosh 版本）:未被激活。当载入一个序列文件，用户进行相关的研究可以使这个相关研究的结果显示出来后，方被激活。

Windows 菜单 列出了工作屏幕的显示选项，包括 **PC-Rare** 的不同窗口：



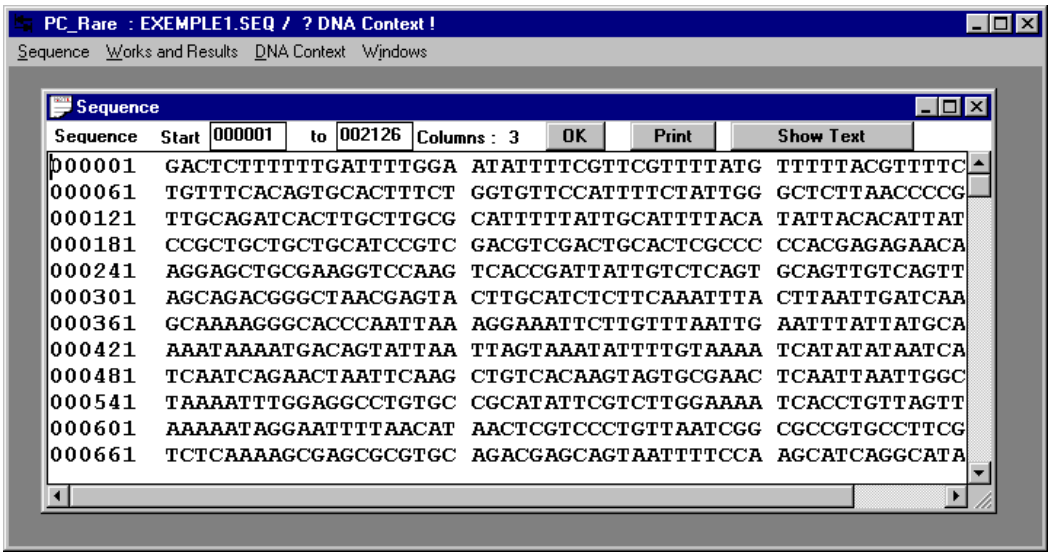
- **Show Window** –当程序运行时，该选项已经被预先选择了。如果这个选项被选中……
- **Open All**–打开所有有用的和可以访问的窗口。
- **Close All**–关闭所有窗口。
- **Windows**-如果该选项被选中，每个新开的窗口都会占有程序工作空间的一部分。
- **Cascade**–重新组织已经打开的窗口。
- **Arrange**–在工作过程中重新组织窗口。

PC Rare 的不同窗口

下面我们来接触一下本软件各个窗口的细节。

Sequence 窗口（Sequence 菜单, sequence 窗口子菜单）.

在载入一个序列后它才是可用的，它是用来显示序列和它的头部的。这个窗口实际由两个不同的部分构成。

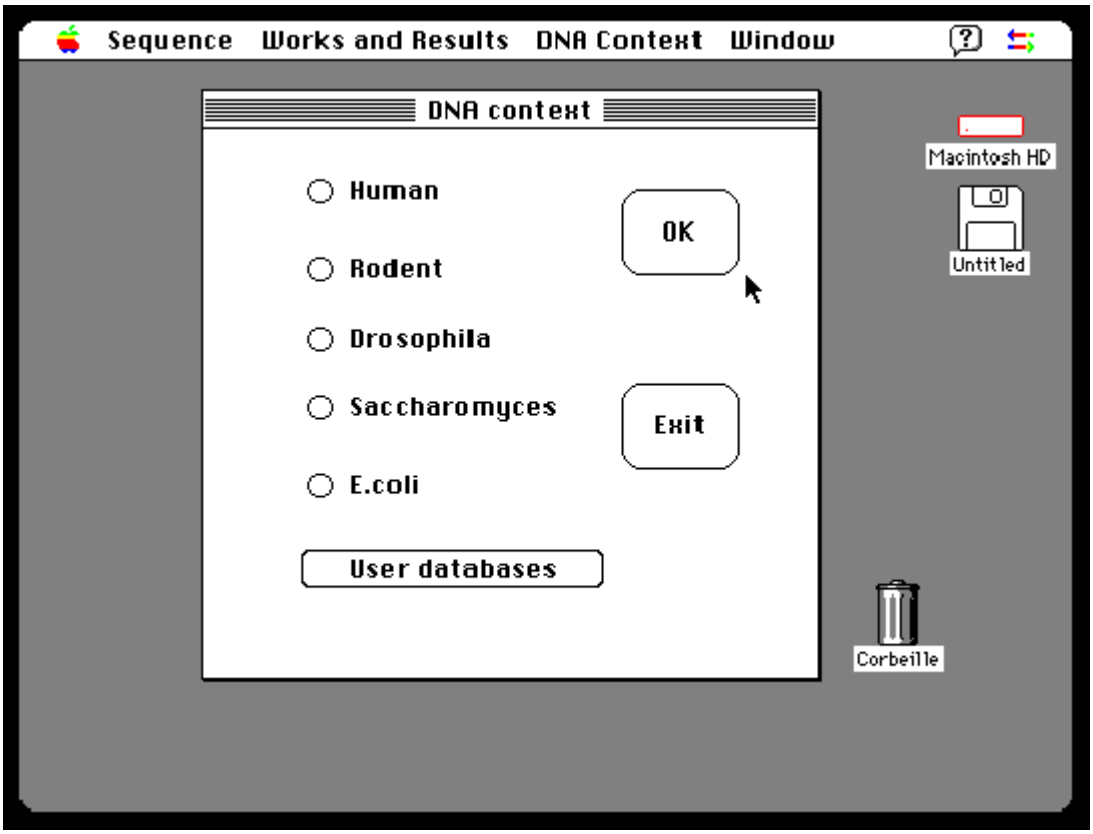


（Macintosh 版本）：一个当 sequence 窗口被选中时才可见的浮动窗口。

- **Show Text**：单击 Show text 在查看 edit heading 模式下查看你的序列的头部。
- **Columns**：每行显示多少个 **20** 个碱基块（单击 **coloumns** 改变此值）。
- **Start**：序列第一个显示的碱基（用户可调节）。
- **To**：该序列最后显示的碱基（用户可调节）。
- **OK**：就用上述各值。
- **Exit**：退出 sequence edition 窗口。
- **Print**：打印窗口内容。

此窗口显示的是一个可操作的标准窗口。在 Macintosh 中单击空格会使浮动窗口显示到最前。

DNA Context 窗口，选择你想要操作的数据库。

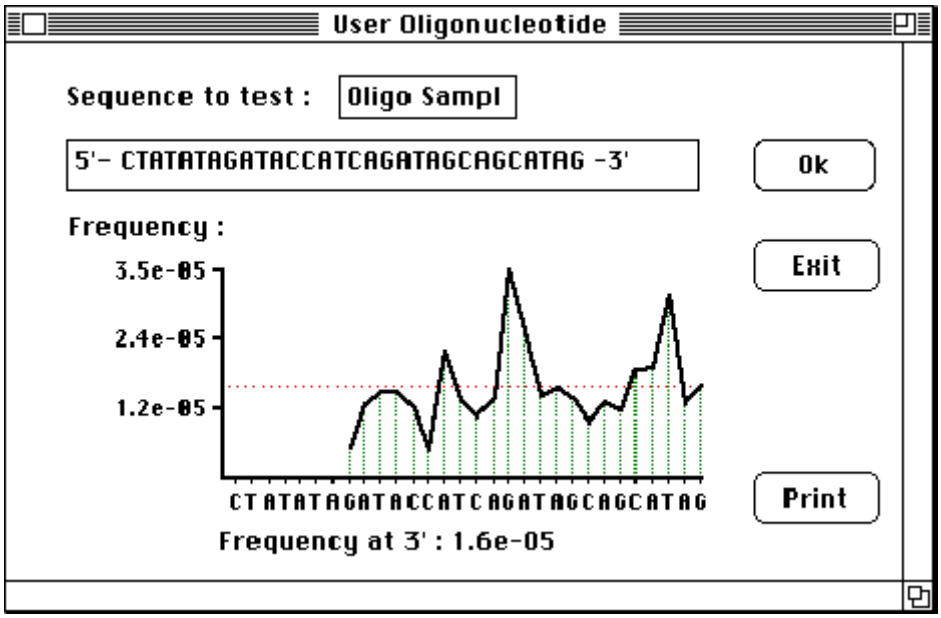


例如：你想检测一个人病毒的序列

- 选择 DNA context 中的 HUMAN！：
  - –单击 DNA context。
  - –选择种（HUMAN）
  - –单击 OK

User Oligonucleotide 窗口 （Sequence 菜单 - User Oligonucleotide sub 菜单）.

当选择一个数据库后你就可以访问这个窗口。它是用来显示所有短序列中所有八聚体的频率曲线的。



- **Sequence to test**：你想要显示的八聚体的名字。Replace the entry under "Name" with the name of your oligonucleotide. The sequence to be tested appears within the frame. As example the sequence is ATGGCC....GAACA. Replace this sequence with the one you wish to test. This line accepts only letters a, c, g, and t wish you type on the computer keyboard. The editing area is standard.
- **OK**：used to display your frequency curve. A same sequence gives a different curve for each database （DNA context）.
- **Exit**：关闭此窗口.
- **Print**：打印获得的曲线（你可以让两个八聚体打印在同一页上）。

printer setup 窗口 （**sequence** 菜单）

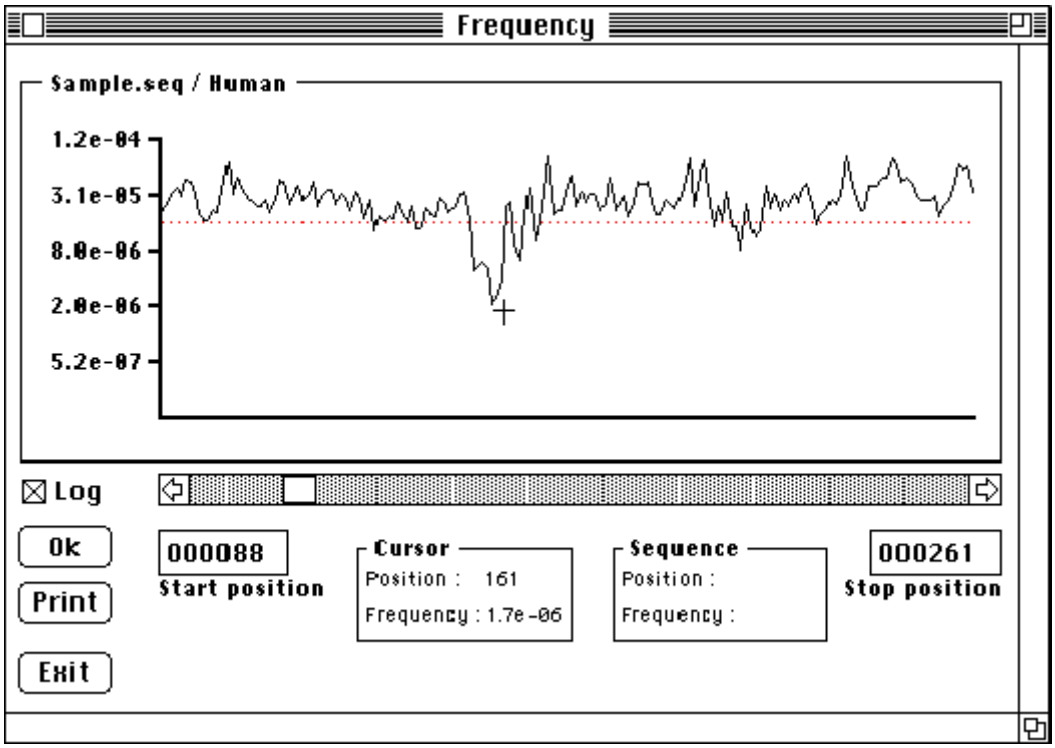
这是一个打印机设置的标准对话框。

重要提示：
Each time you use **PC-Rare**, remember to adjust your printer. **PC-Rare** prints documents in A4 format （210 by 297 mm） in the portrait orientation.

Frequency 窗口 （**sequence** 菜单, **Frequency** 窗口子菜单） 编辑研究序列的频率曲线。

左边的 **frequency** 坐标：

1e-5 说明此八聚体在 10000 个八聚体中有一个。一共有 48（65536）个不同的八聚体，所以在一个随机的序列中，一个八聚体的预期频率是 1/65536，或者说 1,53e-5。频率用水平点线标记。



在右边：

- **Log**：当此选项被选中时，频率坐标以对数单位显示。这对分析曲线上低频率的片段是很有用的。这个选项利用了数学协处理器。
- **OK**：用于在参数改变后编辑曲线。
- **Exit** 用来退出此窗口。
- **Print**：用来打印显示的曲线。（你可以测试并美化打印效果）

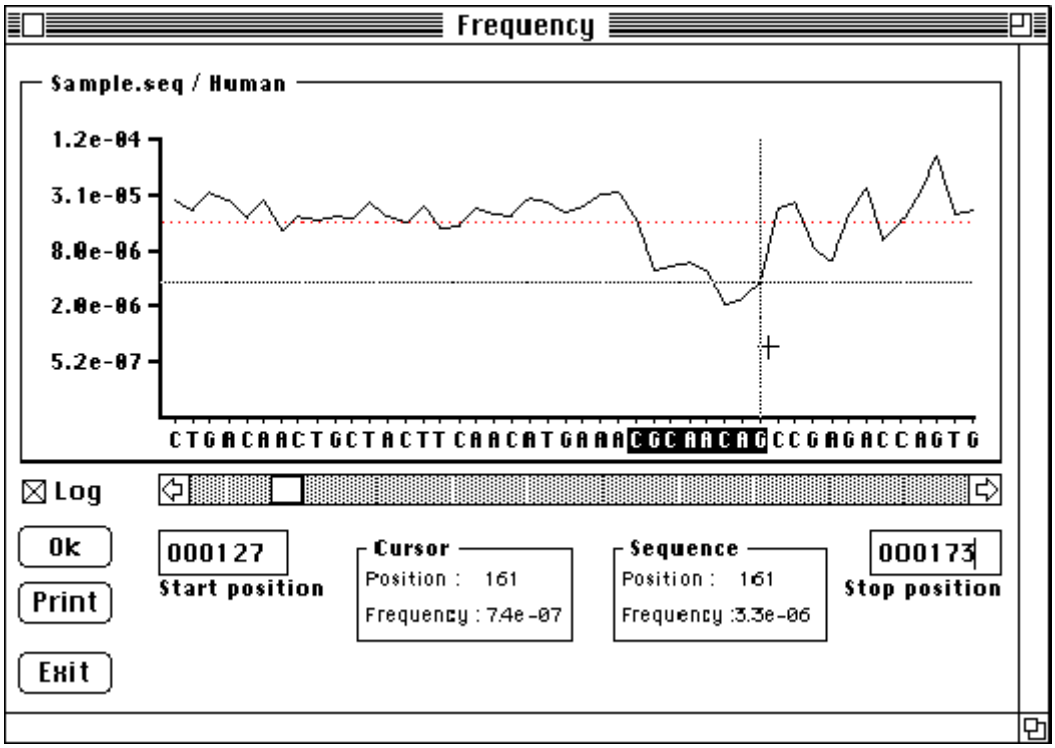
在图表区域下面

- **Start position**：开始描绘序列的位置
- **Last position**：描绘序列的终点

Last 和 Start 之间应该相差 40 到 2000 之间。如果只显示很少的几个核苷酸，那么这个序列显示在曲线底下。（试一下 Start position = 1 且 Last position = 45）。
水平滚动条是标准的，用来直接移动序列。
在 Cursor 框中显示了两种信息：

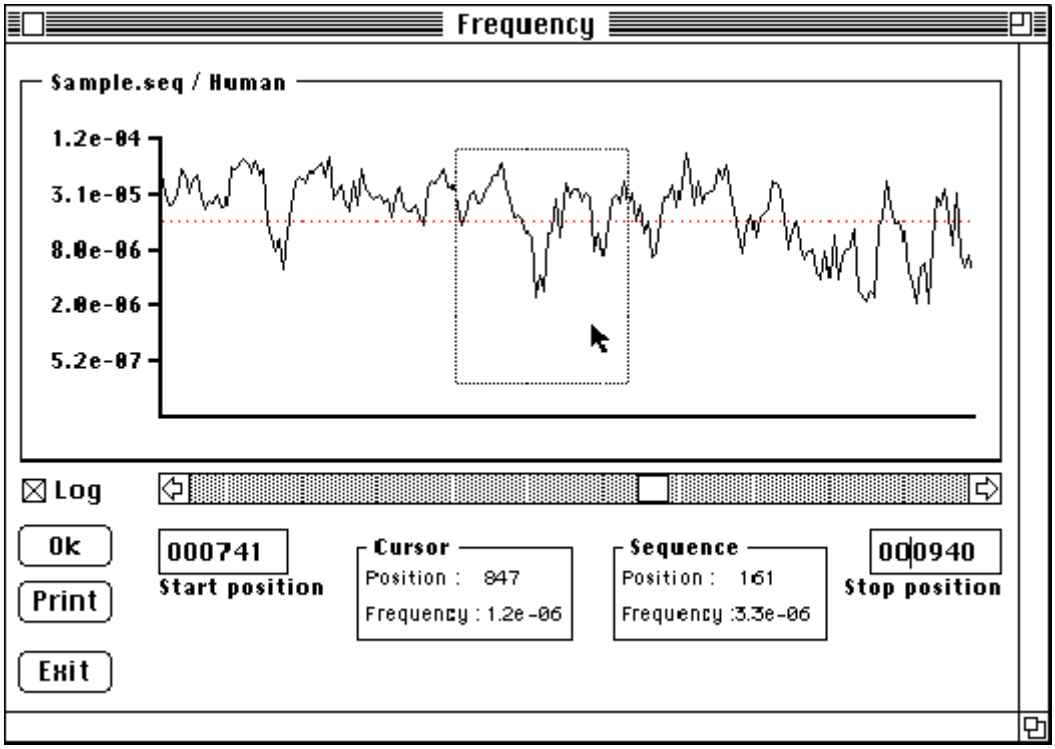
- 位置：显示鼠标指针在序列上的位置 （X 坐标）
- 频率：显示了鼠标所在的频率（Y 坐标）

序列框中显示了两种类型的信息。要得到这个信息，按住 SHIFT 键，并将鼠标移至曲线区域。则鼠标所指的地方的位置和频率都会显示在图表中。

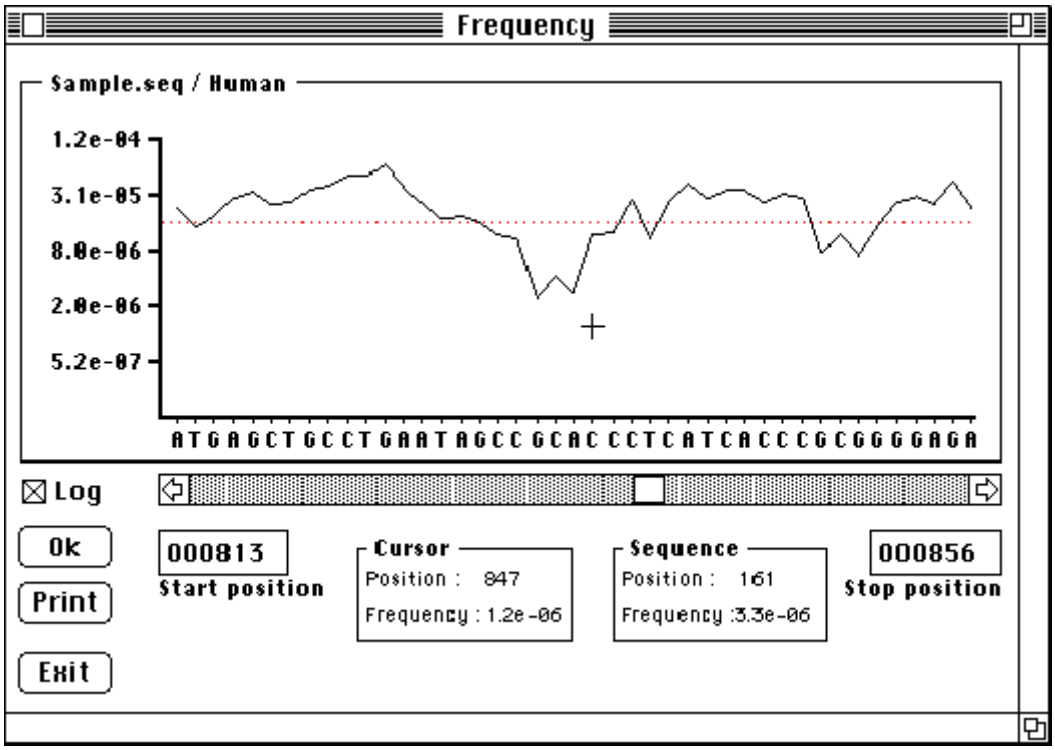


在这个窗口中还有另外值得注意的功能：

- 缩放：在图表区域按住鼠标键移动（Windows 版本为左键）。在屏幕上会显示出一个框。将指针移到这个矩形框中单击，则矩形范围内的曲线就会被显示出来。



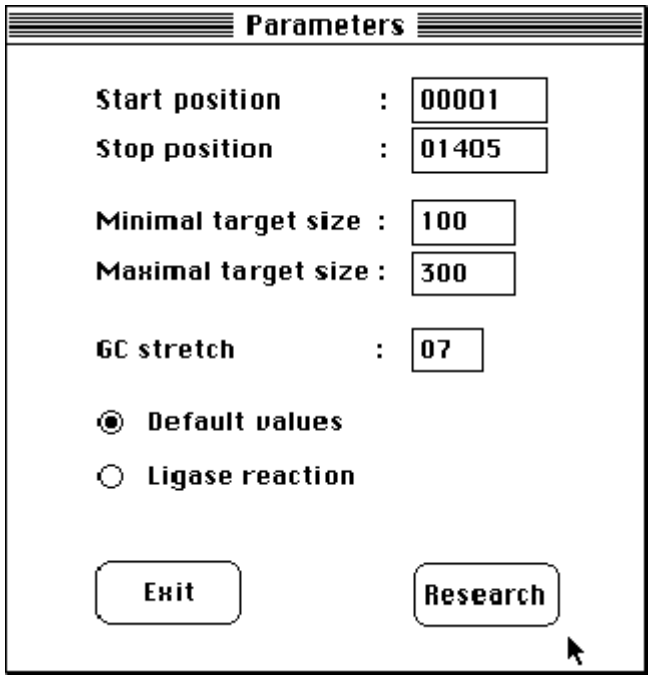
Direct Zoom in



- 移动：双击鼠标键（Windows 版本为左键），将定义一个 400 为基数的曲线。这是顺着序列移动的另一种途径。

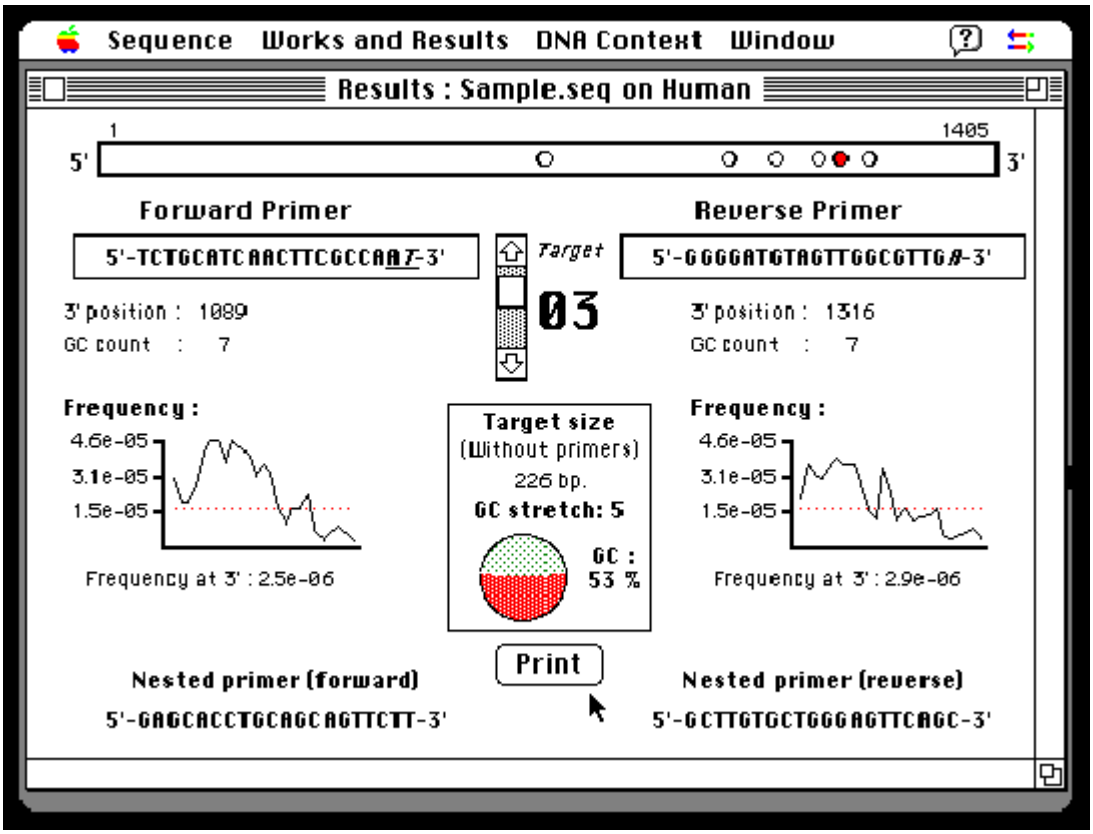
parameters 窗口 （**Works and results** 菜单，**Research** 子菜单）

用来寻找在 3’末端含有低频率八聚体的引物。



- **Start position**：你想开始研究序列的位置
- **Stop position**：停止研究的序列位置
- **Minimal target size**：扩增片段的最小值（！切记，这个大小不包括你所要使用的引物）。如果目标太小，你在分析扩增产物时可能存在问题。如果你使用 20bp 的引物，那么此值应 A value of 100 will yield 140 bp amplification product if you use 20 bp primers. This is the lower limit for analysis on agarose gel.
- **Maximal target size**：扩增片段的最大值（！不包括你所要使用的引物）。太长的目标可能会被错误的扩增。这个参数的最佳值应在 300 左右，在某些情况下它需要作一些改变。
- **GC stretch**：扩增目标允许的连续 G 或 C 的最大数目。大量的相邻的 G 或 C 会将低扩增产量。此参数的最佳值为 7。如果你在此值时没有发现目标（因为可能会遇 GC 非常丰富的序列），请增到 8 或 9。
- **Default values**：还原各个参数的默认值。它们是：
  - Start position =
  - Stop position = End of sequence
  - Minimal target size = 100
  - Maximal target size = 300
  - GC stretch = 7
- **Ligase reaction**：corresponds with Minimal target size = 0 and Maximal target size = 0.
- **Exit**：当不需要再做任何研究时
- **Research**：开始研究

Results 窗口（Work and Results 菜单，Results 子菜单）



- **The rectangular frame**（有一个圆形的地方）显示建议目标的中心。单击每个圆形以显示相应的目标。
- **The start search position** 显示在框的左上方。
- **The end search position** 显示在框的右上方。

建议引物的序列显示在 Forward primer 和 Reverse primer 框中。

> 重要提示：
> 3’末端的这些引物是必须的（框的右边），这是一个低频率末端。

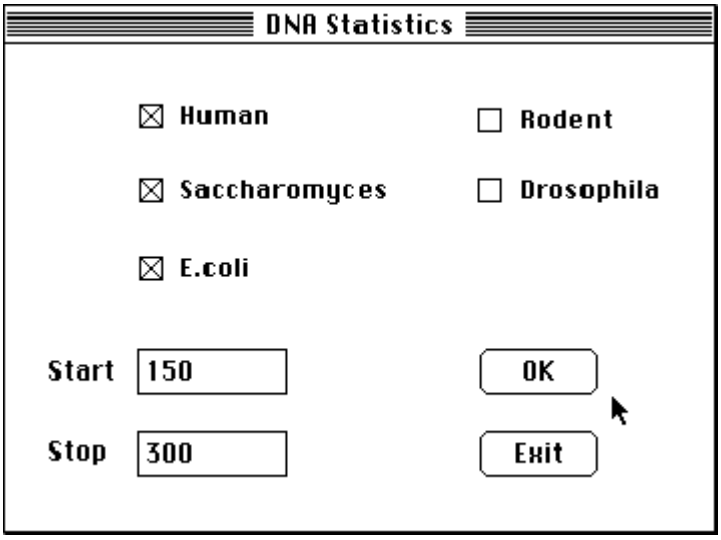
在这个引物的 3’末端，**斜体显示的碱基**表示这两个引物是相互退火的。带下划线的碱基表明引物自身是退火的。

- **3' position**：给出了引物 3’末端在序列的位置。
- **GC count**：显示引物 3’末端最后 16 个碱基的 G 和 C 的数目
- **Target N¡**：一个识别目标的数字，没有特别的顺序。
- **The vertical scrollbar** 是发现目标的另一个途径。
- **Target size**：扩增目标的大小（不包括引物）。

- **GC stretch**：在扩增目标中发现的连续 CG 对的数目。
- **%GC**：扩增目标中发现的 GC 百分比。
- **Frequency**：显示每个引物的频率曲线。
- **Frequency at 3'**：显示八聚体形成引物 3’末端的频率。
- **Nested Primer**：与扩增对象相对应的内部区域的序列。这些引物可以特别用于使用第一 PCR 产物作为模板（Chenal and Griffais 1994）制造用于 PCR 的高度特异性的非放射性探针（Griffais *et al*, 1990 ）。
- **Print**：用来打印结果（图象 A4 格式）
- **The Save Results 窗口**：一个保存结果文件的标准对话框。
- **The Load Results 窗口**：一个载入结果文件的标准对话框。

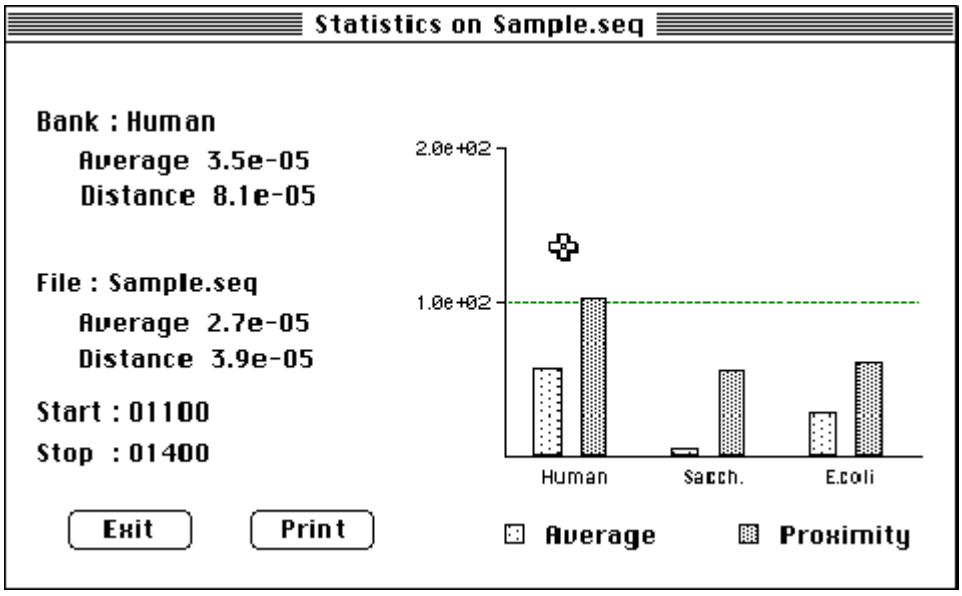
DNA statistics 窗口（-仅限于 Mac-）DNA context 菜单 under statistics 菜单）.

这个窗口通过显示与一个或多个数据库相关的参数用来进行部分序列的统计学研究。每个数据库都必须由用户考虑决定（默认的有 Human，Sacharomyces 和 E. coli）。



- **Start**：开始研究的序列位置。
- **Stop**：研究结束的序列位置。
- **OK** ：开始研究。
- **Exit**：关闭这个窗口。

statistics on 窗口 （仅限于 Mac 版）



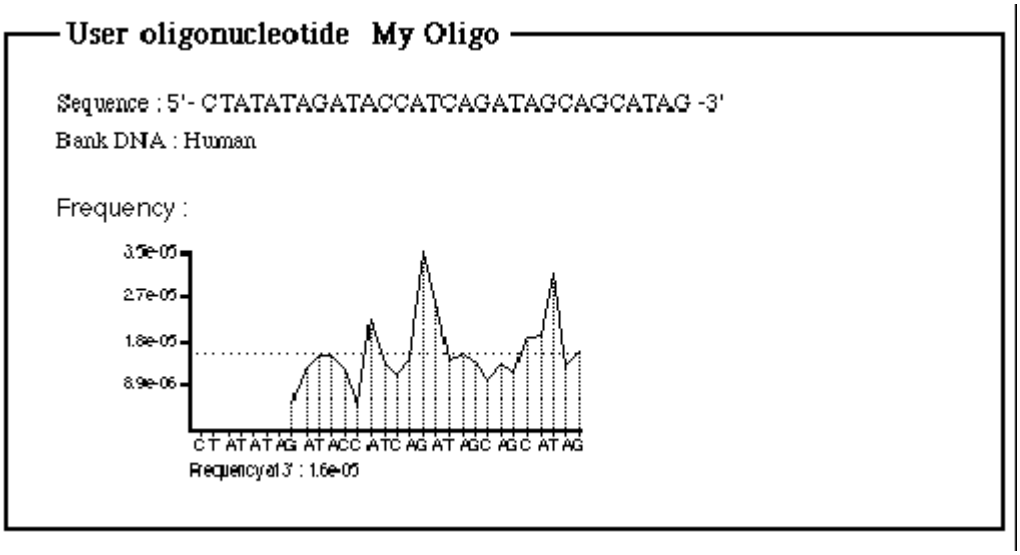
- **Bank**：用来建立结果显示值的数据库
- **average** 在 Bank 中的总八聚体的平均观察值
- **distance**：bank 自身平均值的差异
- 显示的研究的文件序列
- **average** 与 bank 相关的序列的平均观察值
- **distance**：在 bank 中的序列的平均观察值的差异
- 显示研究时序列的起始位置。
- 研究结束时序列的结束位置。

在 PC-Rare 的打印结果中你可以获得什么？

在 **sequence** 窗口中

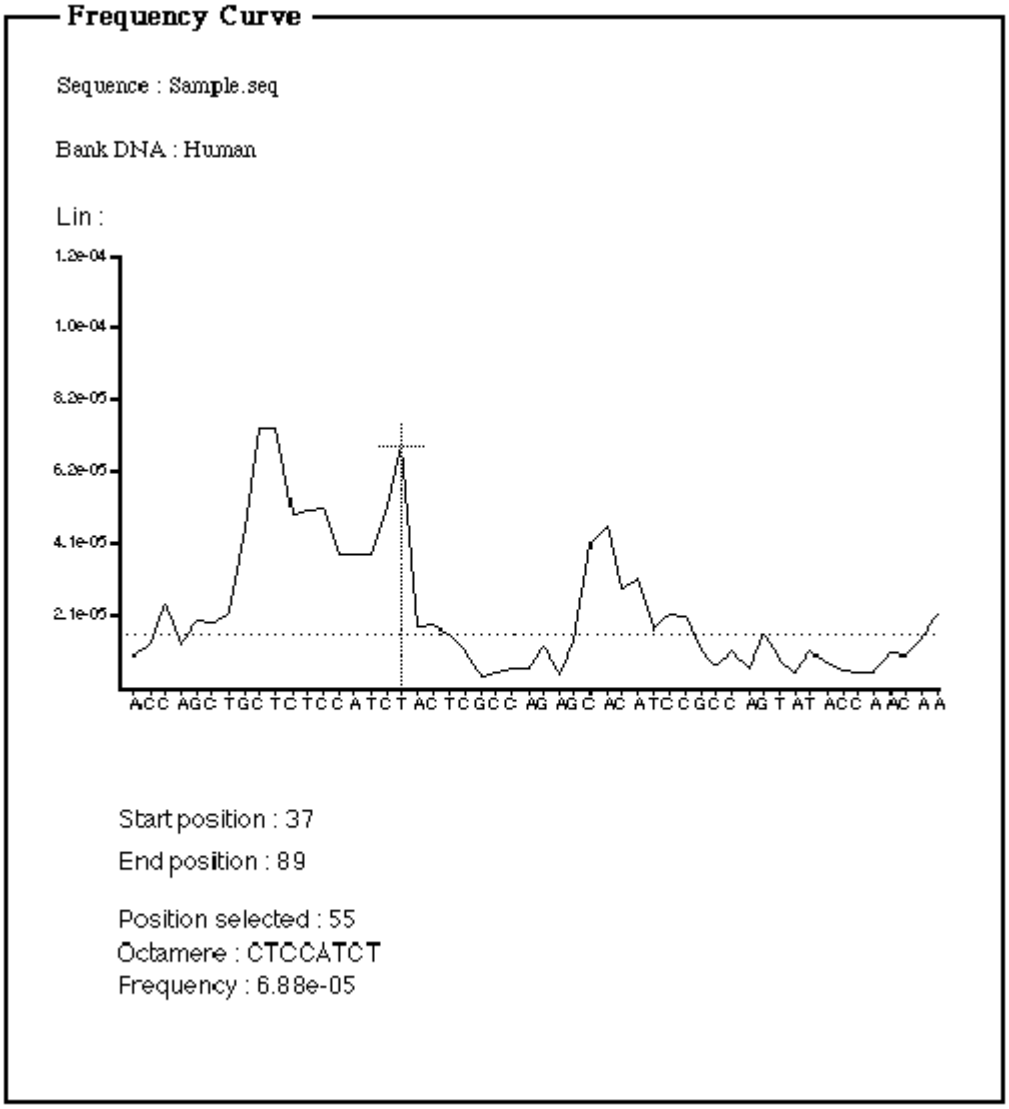
你可以在序列中选择你所操作的一部分或它的头部进行打印。

在 **User Oligonucleotide** 窗口中



你可以打开印一个通过你的计算机键盘编辑的一个短的序列频率曲线。每页可以打印一至两个序列。

在 **frequency** 窗口中



你可以打印在屏幕上显示的频率曲线。If a specie portion of the sequence has been selected on the curve （在移动鼠标的过程中始终按住 SHIFT 键）, 相关的八聚体将出现在打印的页面上。

Result windows 窗口中显示的结果

PC_Rare 04/28/96

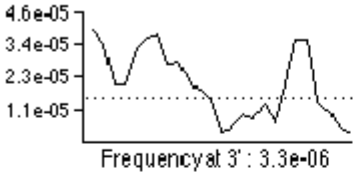
File : Sample.seq 1405 bp.
Bank DNA : Human
Search : Start at : 1
Stop at : 1405
GC Stretch : 7

Results

Amplification Target N° : 11

Forward Primer

5'- GGATGTGAAGAACTCTCGCAGG TCCATCGG -3'



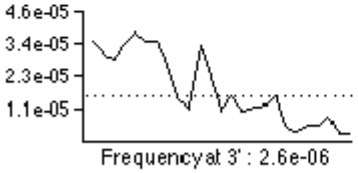
3' Position : 1016

3' Match
-with hitself : 0
-with reverse primer : 0

Nested : 5'-CCTCCTGGACATCTTTGGGT -3'

Reverse Primer

5'- GTTCTTGGGGGGGATGTAGTTG GCGTTGAG -3'



3' Position : 1315

3' Match
-with hitself : 0
-with reverse primer : 0

Nested : 5'-CTTGTGCTGGGAGTTCAGCT -3'

- **File**：将要执行查找的文件名
- **Bank DNA**：DNA context 用来得到它的结果
- **Search**：用户定义的查找标准
  在 results 框中：
- **Amplification target N**º 目标的数目，一个帮助读者辨别目标的经验值。

- **-166bp.#-**：扩增产物的大小，要想得到扩增片断的大小，请添加每个引物的大小。
- （**56% gc**）：扩增片段中的 GC 百分比
- 序列的形象展示。
- 扩增起点和终点
- **Forward Primer**（正向引物）：正向引物的序列，给出了 30-mer。

> 重要提示：
> 大部分时候 20-mer 就足够了，取 3’末端的 20 个核苷酸。

-
  - 正向引物的频率曲线
  - **3' Frequency**：这是引物 3’末端的八聚体频率。（results sheet 表单上的方框里）
  - **3' Position**：序列内引物 3’ 端位置。
  - **3' Match**：自身和反向引物的互补碱基对数目。
  - **Nested**：给出了嵌套引物的序列。这个序列可以用来在嵌套 PCR 和/或从一个扩增产物合成嵌套 PCR 的合成中作为内部填充物。
- **Reverse primer**：细节参见 Forward primer （上方）。引物序列以 DNA 负链的形式，即常规的 5'->3'方向显示，此序列就是你需要定制的序列。

> 重要提示：
> 绝对要将八聚体放在引物的 3’末端 （方框中的八聚体）。

感谢

感谢 Dr. Pierre SOUBIRAN，Dr. Michel DARMON，Viviane CHENAL，Philippe SOUQUE，Andrea MARKOVITS

以及 EUROGENTEC：...

参考书目

- Chenal, V., Souque P., Markovits A. and Griffais R.
  Choosing highly specific primers for the polymerase chain reaction

using the octomer frequency disparity method：application to *Chlamydia trachomatis*. *Gene* （*in-press*）

- Chenal, V. and Griffais, R.
  Chemioluminescent and colorimetric detection of a fluorescein-labelled probe and a digoxigenin-labelled probe after a single hybridization step. *Mol. Cell. Probes* 8 （1994） 401-407.
- Fach, P., Gibert, M., Griffais, R., Guillou, J.P. and Popoff M.
  PCR and Gene Probe Identification of Botulinum Neurotoxin A-, B-, E-, F-, and G-Producing Clostridium spp. and Evaluation in Food Samples. *Appl. Environ. Microbiol.* 61 （1995） 389-392.
- Gabellec, MM., Griffais, R., Fillion, G. and Haour, F.
  Expression of interleukin 1a, interleukin 1b and interleukin 1 receptor antagonist mRNA in mouse brain：regulation by bacterial lipopolysaccharide （LPS） treatment. *Molecular Brain Research* 31 （1995） 122-130.
- Grau, O., Kovacic, R., Griffais, R., Launay, V. and Montagnier, L.
  Development of PCR-based assays for the detection of two human mollicite species, Mycoplasma penetrans and M. hominis. *Mol. Cell. Probes 8* （1994） 139-148.
- Grau, O. and Griffais, R.
  Diagnosis of mutations by the PCR double RFLP method （PCR-dRFLP）. *Nucleic Acids Res.* 22 （1994） 5773-5774.
- Griffais, R., Andre, P.M. and Thibon, M.
  Synthesis of digoxigenin-labelled DNA probe by polymerase chain reaction：Application to Epstein-Barr virus and *Chlamydia trachomatis*. *Res. Virol.* 141 （1990） 331-335.
- Griffais, R., Andre, P.M. and Thibon, M.
  K-tuples frequency in the human genome and the polymerase chain reaction. *Nucleic Acids Res*. 19 （1991） 3887-3891.
- Saiki, R.K., Sharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlish, H.A. and Arnheim, N.
  Enzymatic amplification of b-globulin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230 （1985） 1350-1354.
- Saiki, R.K., Gelfand, D.H. and Stoffel, S.
  Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239 （1988） 487-491.

注释：Rémy Griffais