[基因组学98(2011)137–144](http://dx.doi.org/10.1016/j.ygeno.2011.04.009)

目录列表可从以下网址获得[科学指导](http://www.sciencedirect.com/science/journal/08887543)

基因组学

期刊主页:www.elsevier. com/ locate/ yge no

方法

用于聚合酶链反应、电子聚合酶链反应以及寡核苷酸组装和分析的Java网络工具

鲁斯兰·卡伦达尔[a](#_bookmark0),[b](#_bookmark1),[⁎](#_bookmark4)，李尚义[c](#_bookmark2)艾伦·舒尔曼[a](#_bookmark0),[d](#_bookmark3)

芬兰赫尔辛基大学生物技术研究所MTT/BI植物基因组学实验室，邮政信箱65号，芬兰赫尔辛基-00014

芬兰赫尔辛基芬兰-00790第一数码有限公司

英国剑桥大学亨廷顿路国家农业植物研究所约翰·宾汉姆实验室

d生物技术和食品研究，芬兰农业食品研究公司，芬兰约基奥宁，FIN-31600

# 一辆卡车

*文章历史:*

2010年12月31日收到

2011年4月25日接受

2011年5月3日在线提供

*关键词:*

聚合酶链反应引物设计

初级语言复杂性序列汇编软件

探针设计

连接酶链式反应

# a b s t r a c t

聚合酶链反应是分子生物学的基础，是研究实验室最重要的实用分子技术。我们已经开发并测试了用于聚合酶链反应引物和探针设计的有效工具，这些工具也基于对聚合酶链反应效率的实验研究来预测寡核苷酸的特性。这些工具为大多数聚合酶链反应应用及其组合设计引物提供了全面的工具，包括标准、多重、长距离、反向、实时、独特、组特异性、亚硫酸氢盐修饰分析、重叠-延伸聚合酶链反应多片段组装，以及通过连接酶链式反应设计用于长序列组装的寡核苷酸组的程序。电子聚合酶链反应引物或探针搜索包括对单个引物和引物对的综合分析。它计算标准和简并寡核苷酸(包括LNA和其他修饰)的解链温度，为一组引物提供分析，预测寡核苷酸特性、二聚体和G-四链体检测、语言复杂性，并提供稀释和再悬浮计算器。

2011爱思唯尔公司版权所有。

1. 介绍

聚合酶链反应是分子生物学的基础，也是研究实验室最重要的实用分子技术。然而，该方法的实用性依赖于鉴定独特的引物序列和设计高效的引物。引物设计是所有类型PCR方法中的关键步骤，以确保靶序列的特异性和高效扩增[[1–10]](#_bookmark13)。尽管目前有许多在线的和商业的生物信息学工具，用于聚合酶链反应的引物设计仍然不如常规使用的那样方便和实用。聚合酶链反应适应不同的应用使得有必要为聚合酶链反应引物和探针设计开发新的标准，以涵盖诸如逆转录聚合酶链反应、定量聚合酶链反应、群体特异性[[2,6–8,11]](#_bookmark14) 独特的聚合酶链反应，多重聚合酶链反应中多引物的组合，简单序列重复的发现及其作为诊断标记的扩增，IRAP/REMAP[[12]](#_bookmark17)、TaqMan以及分子信标和微阵列寡核苷酸。

在开发Java网络工具([Table 1](#_bookmark5))，我们的目标是为大多数聚合酶链反应应用程序的常规操作和序列分析创建实用且易于使用的软件。所采用的参数基于我们用于高效聚合酶链反应的实验数据，并被转化为算法，以设计最佳扩增的引物对组合。Java网络工具基于我们的

\*通讯作者在:芬兰赫尔辛基FIN-00014邮政信箱65号赫尔辛基大学生物技术研究所MTT/BI植物基因组学实验室。

*电子邮件地址:*[ruslan.kalendar @赫尔辛基. fi](mailto:ruslan.kalendar@helsinki.fi)(R. Kalendar)，[david.lee@niab.com](mailto:david.lee@niab.com)(李德全)，[alan.schulman @赫尔辛基. fi](mailto:alan.schulman@helsinki.fi)(舒尔曼)。

视窗快速聚合酶链反应软件[[2]](#_bookmark14) 并且一起成功地在整个科学界广泛应用于与聚合酶链反应相关的应用中。

1. 结果
   1. *聚合酶链反应引物设计概述*

引物设计是聚合酶链反应成功的关键步骤之一。对于聚合酶链反应应用，引物长度通常为18-35个碱基，设计时应确保它们与待扩增的目标片段具有完全的序列同一性。这些参数可由用户控制或自动控制，包括引物长度(12–500 nt)、由最近邻热力学参数计算的短引物的解链温度、理论引物聚合酶链反应效率(质量%)值、引物CG含量、3’末端末端强制、退化公式中优选的3’末端核苷酸序列组成以及5’末端添加的序列标签。用于引物选择的其他主要参数是:引物的一般核苷酸结构，如语言复杂性(核苷酸排列和组成)；特异性；整个底漆的熔化温度和3′和5′端的熔化温度；自我互补；次级(非特异性)结合。

该软件可以动态优化输入参数的最佳引物长度。所有PCR引物(探针)的设计参数都是灵活的，并可根据所分析序列和任务的具体情况而变化。分析引物对的交叉杂交、两个引物的特异性，并任选地选择相似的

0888-7543/$–参见前面的事项2011爱思唯尔公司。保留所有权利。doi:[10.1016/j.ygeno.2011.04.009](http://dx.doi.org/10.1016/j.ygeno.2011.04.009)

表1

用于聚合酶链反应、电子聚合酶链反应、寡核苷酸组装和分析的Java网络工具概述。

特征

聚合酶链反应工具为大多数聚合酶链反应应用及其组合设计引物提供了全面的工具:

标准、多重、长距离、反向、实时、独特或群体特异性、后期聚合酶链反应、亚硫酸氢盐修饰分析、聚合酶延伸聚合酶链反应多片段组装克隆；

*电子(虚拟)聚合酶链反应或多重引物或探针搜索，预测可能的聚合酶链反应产物，并搜索指定引物或探针的潜在不匹配位置；*

通过连接酶链式反应(LCR)和聚合酶链反应设计用于长序列组装的寡核苷酸组；

测试单个引物，计算标准和简并寡核苷酸的解链温度，包括LNA和其他修饰；

PCR效率、语言复杂性、二聚体和G-四聚体检测、稀释和再悬浮计算器；

同时分析多个引物的特征，包括Tm、GC含量、语言复杂性、二聚体形成；最优Ta；

简单序列重复位点的鉴定；

序列的模式分析包括(G-C)/(G+C)、(A-T)/(A+T)、(S-W)/(S+W)和嘌呤-嘧啶(R-Y)/(R+Y)偏斜、CG%含量、引物质量、Tm和语言序列复杂性

生成移液表，用于建立聚合酶链反应或定量聚合酶链反应。

熔化温度。具有平衡熔化温度(彼此在1–6℃范围内)的底漆是可取的，但不是强制性的。默认的底漆设计选择标准如所示[Table 2](#_bookmark6)。可以使用预先设计的引物或探针，或者，预先设计的引物可以作为设计新引物的参考。该程序接受预先设计的寡核苷酸序列列表，并检查每个引物与新设计的引物或探针的相容性。

该程序能够产生长寡聚物或聚合酶链反应引物，用于扩增用户定义长度的基因特异性脱氧核糖核酸片段。到目前为止，已经开发了几个公开可用的引物/寡核苷酸设计程序[[3–8,11]](#_bookmark15)。它们都专门用于设计聚合酶链反应引物或寡聚体；其中一些基于Primer3代码[[3]](#_bookmark15)。jPCR基于我们独特的FastPCR软件及其快速高效的代码；它为许多应用提供了设计引物的更灵活的方法。它将检查引物或探针在输入序列中是否有可能产生额外的聚合酶链反应产物的二级结合位点。长寡聚物设计的最佳靶区的选择以与聚合酶链反应引物相同的方式进行。引物设计中的基本参数也被用作低聚物质量的量度，并评价了3’和5’末端碱基的热力学稳定性。引物对的建议和最佳对的选择是可能的。用户可以改变整个序列的产品大小或设计引物对，而无需使用默认或预先设计的参数来指定参数。预先设计的参数是针对不同的情况指定的:例如，针对低CG含量或长距离PCR的序列，或退化序列，或针对手动输入。结果显示了最佳候选引物列表和所有对聚合酶链反应最佳的相容引物对。用户可以指定，

表2

默认引物设计选择标准。

|  |  |  |
| --- | --- | --- |
| 标准 | 默认 | 理想的 |
| 长度(nt) | 19–21 | N 21 |
| Tm范围[(°C)a](#_bookmark7) | 52–68 | 60–68 |
| [Tma](#_bookmark7) 3’端有12个碱基 | 34–48 | 41–47 |
| GC (%) | 45–65 | 50 |
| 3′-末端组成(5′-nnn-3′) | sww、sws、ssw、wss | ssa、sws、wss |
| 序列语言复杂性(LC，%)[b](#_bookmark8) | N 80 | N 95 |
| 序列质量(PQ，%) | N 80 | N 95 |

最近邻热力学参数(11)。

使用字母容量L-gram方法进行序列语言复杂性测量。

对于每个序列，每个序列内正向和反向引物设计都有多个位置，而聚合酶链反应设计将针对不同的靶独立进行。多重聚合酶链反应可以在具有多个扩增子的单个序列中同时进行，也可以针对不同的序列或两者的组合进行。用户可以用类似的方式指定聚合酶链反应产物的大小。

* 1. *熔化温度(Tm)计算*

Tm定义为一半DNA链处于双螺旋状态，一半处于“随机卷曲”状态的温度。具有正常或简并(混合或“摆动”)核苷酸组合的短寡核苷酸的Tm在默认设置中使用最近邻热力学参数计算[[13–17]](#_bookmark17)。寡核苷酸的CG含量是影响Tm值的最重要因素。混合碱的熔化温度是通过平均每个混合点的最近邻热力学参数(焓和熵值)来计算的；类似地，消光系数是通过对混合地点的最近邻值取平均值来预测的[[5]](#_bookmark16)。表S1显示了最近邻热力学参数，正常和混合核苷酸的焓和熵值。5′n1 N2中的第一个核苷酸显示在水平栏中，第二个核苷酸5′n1 N2显示在垂直栏中。

* 1. *序列的语言复杂性；核苷酸偏差分析*

序列复杂性计算方法可用于搜索比较序列之间的保守区域，以检测低复杂性区域，包括简单序列重复、不完全直接或反向重复、多嘌呤和多嘧啶三链脱氧核糖核酸结构和四链结构(如G-四链体)。语言复杂性测量使用字母容量L-gram方法进行[[18,19]](#_bookmark17)沿着整个序列长度，计算为序列中从1到L大小的字的观察范围(xi)的总和除以预期的总和

该序列长度的值。富含G(和富含C)的核酸序列可以折叠成四链DNA结构，该结构包含G-四分体的堆叠(详见<http://www.quadruplex.org/>].这些四链体可以通过两个或四个DNA分子的分子间结合，含有两个G-碱基的序列的二聚化，或者通过含有四个鸟嘌呤嵌段的单链的分子间折叠形成[[20–26]](#_bookmark17)；这些很容易从引物设计中消除，因为它们的语言复杂性低，LC=32% for (TTAGGG)4。

该计划包括各种生物信息学工具，用于对具有气相色谱偏斜、(G-C)/(G+C)、AT偏斜、(A-T)/(A+T)、CG-AT偏斜、(S-W)/(S+W)或嘌呤-嘧啶(R-Y)/(R+Y)偏斜的序列进行模式分析，涉及CG含量和熔化温度，并考虑语言序列复杂性特征。例如，根据公式(G-C)/(G+C)，n个碱基的滑动窗口中的GC偏斜以一个碱基的步长计算，其中G是鸟嘌呤的总数，C是窗口中所有序列的胞嘧啶总数[[27]](#_bookmark18)。正的气相色谱偏移值表示过量的G碱基，而负的气相色谱偏移值表示过量的C碱基。同样，序列中也会计算其他偏斜。

* 1. *引物质量(虚拟聚合酶链反应效率)测定*

实验数据表明，引物核苷酸组成和引物3’端12个碱基的解链温度是影响聚合酶链反应效率的重要因素。12碱基3’末端的熔化温度优选通过最近邻热力学参数计算[[15]](#_bookmark17)。3’末端序列的组成很重要；为了提高聚合酶链反应效率，建议使用带有两个末端碳/庚碱基的引物[[28–30]](#_bookmark19)。核苷酸残基C和G形成强配对结构

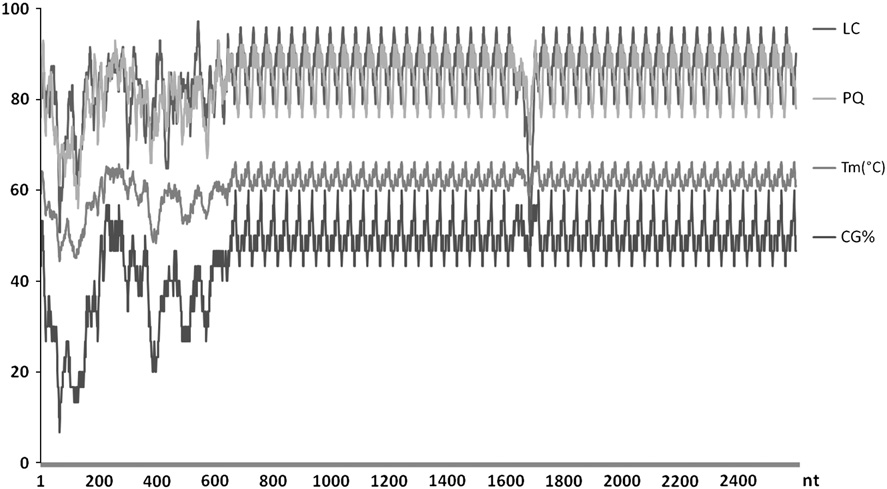


图1。不同引物特征的图谱。滑动窗口大小为30 nt的图谱，同时分析Tm、CG%、语言复杂性和引物质量的序列。

在双链DNA链中。引物模板复合物3’端的稳定性将提高聚合效率。

我们指定了一个称为引物质量(PQ)的抽象参数，它可以帮助估计聚合酶链反应引物的效率。PQ是根据以下参数通过点的连续求和来计算的:总序列和嘌呤-嘧啶序列的复杂性，整个引物和末端3’和5’12个碱基的解链温度。导致可能的二聚体和发夹结构的自互补降低了最终值。PQ试图描述每个引物PCR成功的可能性；该值从最佳引物的100变化到最差引物的0([Fig. 1](#_bookmark9)).

为了满足多路复用需求，在程序中可以选择具有最佳温度范围的最佳引物，允许为任何CG和重复内容的任何目标序列设计合格的引物或探针。PQ值为80或更高允许快速选择最佳的聚合酶链反应引物对组合。由于反应缓冲液的改变、选择的热稳定聚合酶或退火温度的变化，没有观察到对使用高PQ引物的聚合酶链反应扩增的再现性的不利影响。

* 1. *发夹(环)和二聚体的形成*

涉及一个或两个序列的引物二聚体可能出现在聚合酶链反应中。jPCR工具在生成引物列表和候选引物对之前，消除了寡核苷酸内部和之间的反应。避免稳定和抑制性二聚体的产生对聚合酶的效率非常重要，特别是避免聚合酶延伸的引物3’端的互补性。稳定的引物二聚体形成在抑制PCR方面非常有效，因为形成的二聚体被有效地扩增并与预期的靶竞争([Fig. 2](#_bookmark10)).

引物二聚体的预测基于对非间隙局部比对和引物3’端和中心部分稳定性的分析。当引物有可能形成稳定的二聚体时，它们将被拒绝，该二聚体在3’端至少有5个碱基，或在中心部分有7个碱基。工具使用为DNA/DNA双链体提供的平均最近邻热力学参数计算引物二聚体的Tm，所述引物二聚体具有纯的、混合的或修饰的(肌苷、尿苷或锁定核酸)碱基的错配[[13–16,31,32]](#_bookmark17)。除了沃森-克里克碱基配对，还有其他各种氢键构型Crick base pairing, there is a variety of other hydrogen bonding[[20–26]](#_bookmark17) 例如G-四链体([Fig. 3](#_bookmark11)).默认情况下，软件预测引物序列中存在假定的G-quadra-plex。

* 1. *最佳退火温度的计算*

最佳退火温度是指在没有非特异性产物的情况下，聚合酶链反应扩增效率最高的温度范围。估算Ta最重要的值是引物质量、引物的Tm和聚合酶链反应片段的长度。与低Tm(b50c)引物相比，高Tm(n60c)引物可用于Ta范围较宽的PCR。聚合酶链反应的最佳退火温度直接计算为Tm (Tmin)最低的引物的值。然而，聚合酶链反应可以在比引物Tm高10℃的温度下工作，尤其是当反应含有高引物浓度(0.6-1.0微米)以利于引物靶双链体形成时:

m

ta = Tmin+ln(s)；

m

其中s是聚合酶链反应片段的长度。

根据我们的经验以及jPCR算法所基于的400多名快速聚合酶链反应用户的经验(<http://www.primerdigital.com/fastpcr/citations.html>)，几乎设计了所有优质引物

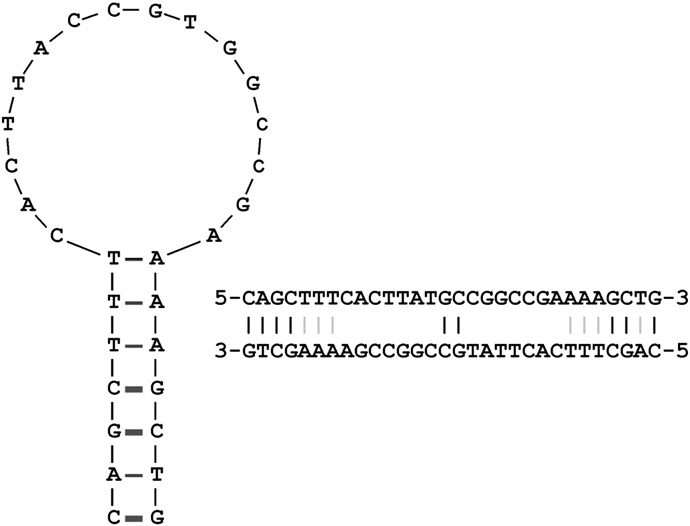


图2。由具有反向重复序列的脱氧核糖核酸分子形成的替代结构(例如在分子信标qPCR探针或LUX引物中)。分子内相互作用将产生发夹(左)，而分子间杂交将产生二聚体(右)。

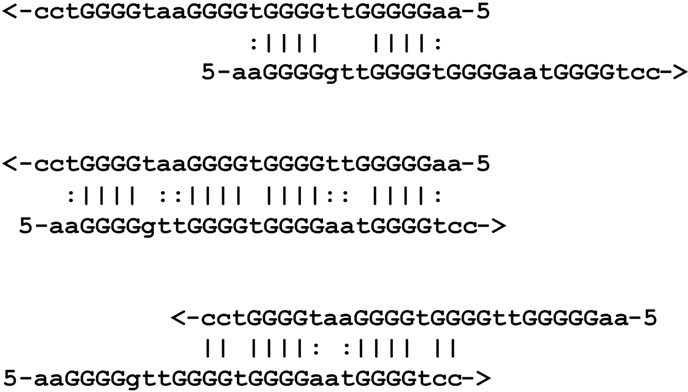


图3。四链体可以由一条、两条或四条单独的脱氧核糖核酸链形成，例如，对称的G-track四链体二聚体可以以相反的方向在两条链之间形成(5′-aagggtgggtggggaatggggtcc)。

默认或“最佳”模式下的jPCR在60至72℃的退火温度下提供扩增而不损失PCR效率，并且在不同的PCR退火温度下以及使用不同的DNA聚合酶和缓冲液时显示出良好的扩增。已经发布了jPCR与用于PCR引物设计和寡核苷酸分析的流行工具的比较，包括最流行的在线引物设计和分析包<http://primerdigital.com/tools/soft.html>。

* 1. *初级分析*

使用jPCR、PrimerAna- lyser或PrimersList软件评估单个引物和引物组。他们使用正常和简并核苷酸组合、CG含量、消光系数、单位转换(nmol)、质量(μg/OD)、分子量、语言复杂性和引物聚合酶链反应效率的默认或其他公式计算引物Tm。用户可以选择带有正常或简并寡核苷酸或带有不同标记(例如肌苷、尿苷或荧光染料)的修饰的脱氧核糖核酸或核糖核酸引物。工具允许选择其他最近邻热力学参数或非热力学Tm计算公式。例如，对于寡核苷酸的非热力学Tm计算，我们建议使用简单的公式:

Tm = 2(A + T + U) + 4(G + C)(对于寡b 7 nt)或

TM = 77.1+11.7 log k++ 0.41(GC %)-528/长度[33]。

对于锁定核酸(LNA)修饰，使用四个符号:dA=E，dC=F，dG=J，dT=L。这两个程序都执行on-type分析，这允许用户在屏幕上立即看到结果。他们还可以根据已知的干寡核苷酸的质量(毫克)、外径或摩尔数来计算获得特定浓度所需的溶剂体积。

分析所有引物的引物内和引物间相互作用，形成二聚体。引物可以利用序列的5’端或中间有效杂交。尽管这种相互作用没有被脱氧核糖核酸聚合酶有效地延伸，但是它们的形成降低了可用于结合靶的有效引物浓度，并且它们的存在可以强烈抑制聚合酶链反应，因为高浓度的双链脱氧核糖核酸是脱氧核糖核酸聚合酶的强抑制剂。

* 1. *二级非特异性结合试验；替代放大*

寡核苷酸的特异性是良好聚合酶链反应的最重要因素之一；最佳引物应该只与目标杂交

序列，特别是当复杂的基因组DNA用作模板时。扩增问题可能是由于引物退火到重复序列(反转座子、脱氧核糖核酸转座子或串联重复序列)而引起的。当引物与反向重复序列互补并产生多个条带时，也可以发生替代产物扩增。当引物是使用特定的脱氧核糖核酸序列设计的(独特的聚合酶链反应)时，这是不可能的。然而，反向重复序列的产生在两种常见的通用DNA指纹方法中被利用——RAPD和聚合酶链反应[[34,35]](#_bookmark17)。因为在这些聚合酶链反应中只使用了一个引物，所以产物的末端必须是反向互补的，因此可以形成茎环。

反转座子间扩增多态性技术、反转座子-微卫星扩增多态性技术、MITE间扩增技术[[12,36,37]](#_bookmark17)，和Alu-重复多形[[38,39]](#_bookmark17) 利用这些高度丰富的分散重复序列作为标记。然而，与重复DNA互补的引物可能在单引物扩增中产生许多非特异性条带，并损害独特的多克隆抗体的性能。引物序列的同源性搜索，例如使用“blastn”对GenBank或EMBL-Bank中的所有序列进行同源性搜索，将确定引物是否可能与分散的重复序列相互作用。或者，人们可以基于Repbase中的重复序列创建一个小型的局部专门文库[[40]](#_bookmark17) 或者TREP [<http://wheat.pw.usda.gov/ITMI/Repeats/>].

默认情况下，jPCR对每个给定序列执行非特异性结合测试。此外，软件允许对照一个或多个参考序列(如BAC、YAC)或自己的数据库进行该测试。与当前序列上一个以上位置结合的引物将被拒绝。即使非特异性引物结合测试作为所有引物的默认测试进行，用户也可以取消操作。包括错配杂交在内的二级结合位点的鉴定通常通过考虑引物与沿着整个引物序列的靶的相似性来进行。一个隐含的假设是引物与模板的稳定杂交是由脱氧核糖核酸聚合酶引发的先决条件。jPCR特别关注引物的3’末端部分，并计算引物3’末端与目标的相似性(长度由用户选择)，以确定3’末端的稳定性。二级非特异性引物结合试验是基于参考序列和输入序列之间快速、无间隙的局部比对(允许9聚体的散列指数中有一个错配)。

* 1. *多重简并引物设计*

多重聚合酶链反应是一种常用于在单个反应中扩增多个脱氧核糖核酸靶区的方法。许多靶标的同时扩增减少了需要进行的反应数量；多重聚合酶链反应因此提高了生产效率。多重聚合酶链反应检测的设计可能是困难的，因为它涉及对交叉相互作用的引物对的大量计算分析。为了实现靶的均匀扩增，引物必须设计成以相同的效率与其靶结合。jPCR可以为所有输入序列和/或每个序列内的多重靶标快速设计一套多重PCR引物。PCR条件可能需要调整；例如，提高或降低退火温度，使得所有的产物被同等有效地放大。为了实现这一点，大多数现有的多重引物设计包使用引物熔化温度。

实际上，几乎相同的Ta和Tm的设计非常重要。聚合酶链反应产物的熔化温度也很重要，因为它们与退火温度值有关。聚合酶链反应产物的Tm取决于其气相色谱含量和长度；短产物在低PCR退火温度(100 bp，55℃)下比长产物(3000 bp，n60℃)更有效地扩增。对于大多数多重聚合酶链反应，在所有引物对的最佳Ta和聚合酶链反应产物之间通常有一个小的变化(高达5℃)。退火温度必须是最佳的，以便最大化

扩增目标基因组序列的可能性，同时将非特异性扩增的风险降至最低。进一步的改进可以通过选择最大化普通引物范围的最佳引物组来实现。一旦出现提示，jPCR将为给定的目标序列计算多重聚合酶链反应引物对。计算速度取决于涉及的靶序列和引物对的数量。

设计相容的多重聚合酶链反应引物对的另一种方法是使用预先设计的引物作为设计新引物的参考。用户还可以选择聚合酶链反应产物的输入选项，例如扩增子之间的最小产物大小差异。可以为每个给定序列单独设置引物设计条件，或者使用共同的值。与常规设置相比，单独设置对聚合酶链反应引物或探针设计具有更高的优先级。结果包括单个序列的引物、相容的引物、产物大小和退火温度。因为产物的明确区分依赖于在单个反应中使用相容的引物对，该程序回收引物组合的所有潜在变体，用于分析所选的DNA区域，并以表格形式提供它们与信息的相容性，包括引物二聚体、交叉杂交、产物大小重叠和基于Tm的类似替代引物对。用户可以选择那些提供所需产品尺寸的替代相容引物对组合。使用该程序，研究人员可以通过改变上述过滤条件，从目标中选择预先设计的引物对进行所需类型的聚合酶链反应。例如，传统的多重聚合酶链反应需要一组目标基因的不同大小(至少10 bp)的扩增子，因此可以选择聚合酶链反应产物之间最小大小差异的值。

除了需要避免相同大小的扩增子之外，多重聚合酶链反应

还必须最小化引物二聚体和次级产物的产生，随着反应中引物数量的增加，这变得更加困难。为了避免非特异性扩增的问题，jPCR允许通过选择避免重复或其他基序的序列来选择最有可能只产生靶序列扩增子的引物对。该程序还允许用户不仅设计相容的引物对，而且为不同的靶或序列设计相容的单一引物。

* 1. *群体特异性聚合酶链反应引物*

群体特异性扩增，也称为家族特异性和序列特异性扩增，是相关基因、序列和基因组比较研究的重要工具，可应用于进化研究，特别是基因家族和克隆新的相关序列。特定的目标，如抗病类似物(NBS- profiling)或转座因子可以被扩增，以揭示与这些序列相关的脱氧核糖核酸多态性。设计组特异性聚合酶链反应引物的总体策略是使用9聚体的散列指数来识别目标序列中的公共区域，遵循当前序列的标准聚合酶链反应设计，然后测试这些引物与其他序列的互补性。与Primaclade软件相比[[7]](#_bookmark17)Primique[[8]](#_bookmark17)，UniPrime[[6]](#_bookmark17) 还有GeneFisher[[11]](#_bookmark17)，jPCR不使用序列比对，使其可以灵活地使用不同的策略进行引物设计。jPCR不设计简并PCR引物来扩增所有相关序列的保守或多态性区域。

jPCR包为每个给定序列设计大组通用引物对，鉴定保守区域，并为所有给定靶产生合适的引物。算法的步骤是自动执行的，用户可以影响底漆设计选项的一般选项。只要有可能在集合中找到短的(至少12 nT)共有序列，jPCR将与任何序列源一起工作。引物设计的质量取决于序列关系、系统发生的相似性以及共有序列对设计好引物的适用性。

该软件能够独立地为每组序列生成适用于所有序列的组特异性引物。组特异性聚合酶链反应引物的引物比对参数类似于用于电子聚合酶链反应的参数。该软件已经在群体特异性聚合酶链反应中进行了广泛的实验测试。

* 1. *用于片段组装的聚合酶链反应延伸*

不依赖序列的克隆，包括不依赖连接的克隆(LIC)[[41]](#_bookmark17) 需要在载体和插入片段中产生互补的单链突出端。类似地，在PCR中可以使用重叠引物以有序的方式连接或连接多个片段[[42]](#_bookmark17)。引物重叠中不同靶之间互补区域的退火允许聚合酶在热循环过程中合成包含靶序列的连续片段，这一过程称为“重叠延伸聚合酶链反应”(OE-PCR)[[43]](#_bookmark17)。效率取决于Tm以及重叠的长度和唯一性。为了实现这一点，该程序在每个片段的末端设计相容的正向和反向引物，然后使用来自最终产物中相邻片段引物的序列延伸引物的5’末端。该程序选择重叠区域，使得来自重叠片段的引物在大小和最佳退火温度上相似。该程序添加了所需的碱基，使得重叠的Tm与初始引物的Tm相似或更高。在合适的引物对中测试引物的二聚体。

* 1. *用于装配长序列的寡核苷酸设计*

基于寡核苷酸长度和Tm阈值的说明，已经开发了几个程序来自动设计用于长序列合成的寡核苷酸[[44–46]](#_bookmark17)。只有程序TmPrimer[[44]](#_bookmark17)提供寡核苷酸之间潜在相互作用的预测；然而，该程序不考虑非特异性杂交。我们的算法能够为含有重复序列的长序列设计寡核苷酸，并在聚合酶链反应的3’末端延伸过程中最小化它们潜在的非特异性杂交。

* 1. *电子聚合酶链反应*

模拟引物与目标退火位点的杂交是预测PCR产物的唯一方法[[47–52]](#_bookmark17)。引物3’端的最后10–12个碱基对结合稳定性很重要；单个错配会降低PCR效率，这种影响随着接近3’端而增加。jPCR允许同时测试单个引物或为多重靶序列设计的一组引物。它进行快速、无间隙的比对，以测试引物与靶序列的互补性。可以设置参数以允许引物3’端有不同程度的错配。该程序还可以处理简并引物或探针，包括具有5’或3’尾序列的引物或探针。使用标准或反向聚合酶链反应以及多重聚合酶链反应可以发现线性和圆形模板的可能聚合酶链反应产物。这种电子工具可用于快速分析引物或探针的靶序列，确定引物的位置、方向、结合效率和计算它们的Tm。

* 1. *简单序列重复基因座搜索*

简单序列重复(SSRs或微卫星)是一个或多个碱基的短串联重复。微卫星无处不在地分布在真核生物基因组中，通常长度高度多态，因此是群体遗传研究的一类重要标记。我们的SSR搜索方法是使用语言序列复杂度来分析低复杂度区域。这种方法允许检测具有单个、最多10个碱基重复基序的完美和不完美的SSR。对每个条目序列进行处理，以识别SSR和SSR

表3

引物设计和寡核苷酸分析工具的比较。

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 特征 |  | 底漆-BLAST(底漆3) |  | IDT科学工具:PrimerQuest，寡核苷酸分析仪3.1 |  | PerlPrimer |  | 双搜索网络服务器 |  | 网络工具 |  |
| 输入(FASTA或原始序列) |  | + |  | + |  | + |  | + |  | +，制表符-表格 |
|  |  |  | 特征 |  |  |  |  |  |  |  |  |  |
|  | 引物或探针设计，长度(nT) |  | 15–30 |  | 16–35 |  | 12–30 |  | 10–35 |  | 12–500 |  |
|  | 序列长度限制(nT) |  | 50,000 |  | 无限制 |  | 无限制 |  | 5000 |  | 无限制 |  |
|  | 相对计算速度 |  | 快的 |  | 慢的 |  | 慢的 |  | 非常慢 |  | 非常快 |  |
|  | 多个模板(序列或引物)和 |  | − |  | − |  | − |  | − |  | + |  |
|  | 每个序列中有多个目标 |  |  |  |  |  |  |  |  |  |  |  |
|  | 每个序列的单独聚合酶链反应选项 |  | + |  | + |  | + |  | + |  | + |  |
|  | 所有操作中的简并核苷酸 |  | − |  | + |  | − |  | + |  | + |  |
|  | (Tm计算、搜索和探测， |  |  |  |  |  |  |  |  |  |  |  |
|  | 底漆设计等。) |  |  |  |  |  |  |  |  |  |  |  |
| LNA和其他核苷酸修饰 | | − | | + | | − | | − | | + | | |
| 启用高吞吐量运行 | | − | | − | | − | | − | | + | | |
|  | | LC = 88.6±4.8% | | LC = 88.2±7.4% | |  | | LC = 79.1±10.2% | | LC = 82.4±7.1% | | |
|  | | (2000本初级读本) | | (407个引物) | |  | | (525个引物) | | (2000本初级读本) | | |
| 最佳退火温度的计算 | | − | | − | | − | | − | | + | | |
| 引物的3’端交叉和自身二聚体 | | − | | + | | + | | + | | + | | |
| g-四链体检测 | | − | | − | | − | | − | | + | | |
|  | | 替代放大 | |  | |  | |  | |  | | |
| BLAST搜索 | | + | | − | | + | | + | | − | | |
| 内部序列测试 | | − | | − | | − | | − | | + | | |
| 外部(特定库)测试 | | + | | − | | + | | + | | + | | |
|  | | 聚合酶链反应应用 | |  | |  | |  | |  | | |
| 用成对引物和/或单引物进行多重聚合酶链反应 | | − | | − | | − | | − | | + | | |
|  | 火帽 |  |  |  |  |  |  |  |  |  |  |  |
| *多序列和引物的电子聚合酶链反应* | | − | | − | | − | | + | | + | | |
| 通用和独特的聚合酶链反应 | | − | | − | | − | | − | | + | | |
| 反向聚合酶链反应和循环序列 | | − | | − | | − | | − | | + | | |
| 亚硫酸氢盐修饰聚合酶链反应分析和电子聚合酶链反应 | | − | | − | | + | | + | | + | | |
| 聚合酶链反应延伸多片段 | | − | | − | | − | | − | | + | | |
|  | 装配克隆 |  |  |  |  |  |  |  |  |  |  |  |
| 用于LCR和聚合酶链反应的寡核苷酸组装 | | − | | − | | − | | − | | + | | |
| 提供图形网络界面 | | + | | + | | + | | + | | + | | |
| 实验室验证的底漆设计参数 | | + | | + | | + | | + | | + | | |

支持+功能，不支持-功能。

侧翼用于设计相容的正向和反向引物，以便通过PCR进行扩增。

jPCR识别每个进入序列中的所有SSR，并为每个SSR位点设计相容的聚合酶链反应引物对。默认的聚合酶链反应引物设计参数是引物必须在鉴定的SSR两侧100个碱基以内。通常，SSR位点周围可用的序列不足以设计好的引物，用户可以增加或减少与任一侧的距离，以找到更有效和相容的引物对。全球许多实验室已经成功实施了jPCR计划，为SSR位点设计聚合酶链反应引物；一些最终的文章显示在我们的FastPCR网页上[<http://primerdigital.com/fastpcr/citations.html>].jPCR的能力使其成为一个完整的生物信息学工具，可以将微卫星用作标记，从发现到引物设计。

* 1. *与其他软件的比较*

Primer3及其衍生应用，如NCBI/Primer-BLAST [<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>]，IDT科学工具[<http://eu.idtdna.com/scitools/>，5]，BiSearch web服务器[<http://bisearch.enzim.hu/>，10]，Primo Pro 3.4 [<http://www.changbioscience.com/primo/>]和PerlPrimer [<http://perlprimer.sourceforge.net/>，4]是最常用的基于web的应用程序。[Table 3](#_bookmark12) 我们的网页[<http://primerdigital.com/tools/soft.html>]将我们的引物设计和寡核苷酸分析软件的特性与这些程序进行比较。例如，我们研究了在给定序列中进行非特异性扩增分析的能力。的测试结果

NCBI/Primer-BLAST和BiSearch网络服务器可作为补充材料(S2)；一般来说，其他工具不考虑序列内部的重复区域，可以在其中设计引物或探针，这可能导致非特异性扩增(S2)。

NCBI/Primer-BLAST为高CG含量(含65-70% CG)序列设计的引物在3’端的12个碱基末端有90-100% CG。双搜索网络服务器和IDT PrimerQuest在默认条件下正确设计了CG含量为50%的引物。我们发现引物扩增效率不仅取决于引物的Tm及其3’端序列，还取决于其语言复杂性。对于一个好的引物，语言复杂度值必须超过80%，最好的引物必须超过90%。BiSearch网络服务器引物的语言复杂度平均值为79.1±10.2%(测试了525个引物)；对于NCBI/引物-BLAST，82.4±7.1%(2000个引物)；对于IDT总理来说，

88.2±7.4%(407个引物)；对于我们的jPCR，88.6±4.8%(2000个引物)。例如，用NCBI/Primer-BLAST和BiSearch web服务器设计的引物5’tttttttttttttgggggggggag 3’(LC = 38%)、聚嘌呤5’ggagagagagagagaag 3’(LC = 33%)和聚嘧啶5’ctcttctctctccttctcc 3’(LC = 40%)存在形成稳定四链结构的风险。

为了比较二聚体检测的效率，我们为拟南芥叶绿体DNA设计了2000个引物(1–50，000个碱基；AP000423)。使用NCBI/引物-BLAST，我们发现105个引物形成不可接受的自身二聚体，如一些实施例所示(S3)。BiSearch web服务器是最慢的基于web的软件(默认条件下，5 kb的100个引物对需要10分钟)，但它设计的引物没有3’端二聚体，只有一个含有弱内部

自二聚体。用IDT引物对(S3)设计的引物发现了几个内部自二聚体(21)和两个弱的3’端二聚体。来自IDT引物组的引物显示出良好的语言复杂性和引物质量值。

1. 讨论

我们开发了一种有效设计聚合酶链反应引物和探针、长序列组装寡核苷酸、电子分析以及预测和分析寡核苷酸特性的方法。该软件可以同时处理多个核酸序列和内部序列目标。我们从其他实验室收集和分析了自己的发现和数据，以更精确地预测聚合酶链反应效率的寡核苷酸质量。使用这些来设置寡核苷酸设计的参数，我们能够提高聚合酶链反应的成功率。与此处的工具相比，几乎所有可用于设计PCR引物的程序都不调查引物的序列，而仅使用引物的熔化温度进行选择。软件算法可以选择适用于大多数聚合酶链反应应用的高特异性寡核苷酸，具有宽的功能温度范围，并且可以消除较差的寡核苷酸-

对于通过聚合酶链反应和连接组装多个片段来说，末端并不重要。对于电子聚合酶链反应，检测参考序列上引物位置的快速比对是通过使用3-或7-聚体的散列指数(内部最多包含一个错配)对两条链进行分析并计算整个引物序列的局部相似性来进行的。可以设置快速比对的参数:搜索起始的最小基数为3，局部相似度为50%。

该程序从输入序列生成引物对(和探针)，并显示每个引物对的最佳退火温度和聚合酶链反应产物的大小以及每个设计引物的信息。结果由程序生成，以表格形式显示建议的引物和引物对，用于Excel或Open Office。电子表格显示以下属性:自动生成的引物名称、引物序列、序列位置、方向、长度、熔化温度、CG含量(%)、分子量、摩尔消光系数、语言复杂度(%)和PQ。对于相容的引物对，还提供了退火温度和聚合酶链反应产物大小。

语言复杂度值转换为百分比，100%为最高级别:

利用二级结构预测的核苷酸。

设计聚合酶链反应引物有几种公开可用的工具

100 ×

*L*

∑ xi

*E*

但是jPCR是唯一一个具有集成工具环境的，它为大多数PCR设计引物提供了全面的工具

*LC(%) =* *i =1*

应用，有助于设计同时具有多种核酸序列的引物以及用于不同PCR应用及其组合的内部序列靶标，并允许其应用

*L*

*e =∑e*

*i =1*

*s I+1；s b4i 1+I*

4i；s≥4i 1+I

（

广泛项目的组件化功能。这个软件是灵活的，并且允许在广泛的项目中应用它的组件化功能。它已经在世界各地的许多实验室和公司进行了测试，为一系列应用和任务设计了成千上万的PCR引物和探针[参见<http://primerdigital.com/fastpcr/citations.html>]，提供了宝贵的反馈，用于进一步发展方案。我们将继续通过实验室和提供反馈的合作者网络来评估这些工具。

1. 材料和方法

工具<http://primerdigital.com/tools/>]是用NetBeans IDE (Oracle)用Java编写的，需要计算机上的Java运行时环境(JRE)。它可以用于任何操作系统(64位操作系统更适合大型染色体文件)。Java应用程序要么采用单个序列，要么接受多个单独的脱氧核糖核酸序列，采用FASTA或表格格式(Excel表或Word表)，或者作为一个RTF文件。聚合酶链反应引物设计算法生成的引物组具有很高的成功可能性，可用于任何扩增方案。所有引物都可以用于聚合酶链反应或测序实验。

多重聚合酶链反应算法基于快速非递归方法，软件对所有引物的产品大小兼容性和交叉二聚体相互作用进行检查。对于长序列组装，寡核苷酸设计从给定序列的5’端开始；寡核苷酸长度动态变化，直到发现独特的3’末端，并且寡核苷酸的Tm达到Tm阈值。所有寡核苷酸的设计都没有间隙。另一条链用于重叠寡核苷酸的设计，其算法与上述相同，但重叠区域的Tm达到Tm-15c阈值。3’末端的序列组成很重要，因为寡核苷酸复合物中3’末端的稳定性将提高聚合酶延伸的特异性。为了减少非特异性聚合酶延伸和连接(LCR)，该算法只选择3’末端的独特序列。至少，3’末端的最后2个核苷酸不能与非特异性靶互补。除了3’以外的其他互补区域

其中s是序列的长度。

例如，序列5′-acacacac，16 nT，含有两个核苷酸(A，C)，但预期E =4个变体；两个双核苷酸变体(AC，CA)，但预期为E =(16-1)变体；三核苷酸的两个变体(ACA，CAC)和预期的E =(16-2)变体。这

复杂度值为LC = 100(2+2+2)/(4+161+162)= 18.2%。

分子间G-四链体形成序列根据公式…GM1 ngm 2…检测，其中m是每个G-道中G残基的数量(m1，m2≥3)；间隙Xn (n≤ 2最小值(m1:m2))可以是任何残基的组合，包括G[[25]](#_bookmark17)。间隙序列(Xn)可以具有不同的长度，并且相对稳定的四链体结构仍然可以形成超过7个碱基长的环，但是一般来说，增加间隙的长度会导致结构稳定性降低。当有N 6个碱基的长聚G束时，其中一个间隙的长度也可能为零。

与本文相关的补充材料可在doi:10.1016/j . ygeno . 2011 . 04 . 009在线查阅。

感谢

学术机构可以免费获得网络工具，只要它们仅用于非商业研究和教育。不得出于商业目的复制或分发。这项工作得到了PrimerDigital有限公司和寡聚体有限公司的部分支持

参考

1. X.杨，谢弗勒，韦斯顿，高等植物DNA多态性引物设计和mRNA分析的最新进展，植物方法2 (2006) 4。
2. R.用于聚合酶链反应引物和探针设计和重复搜索的快速聚合酶链反应软件，基因、基因组和基因组学3(2009)1–14。
3. 南Rozen，H.J. Skaletsky，WWW上面向普通用户和生物学家程序员的Primer3，载于:S. Krawetz，S. Misener (Eds。)，生物信息学方法和协议:分子生物学方法，Humana出版社，Totowa，NJ，美国，2000，第365-386页。
4. 《标准、亚硫酸氢盐和实时聚合酶链反应的跨平台图形化引物设计》，《生物信息学》20(2004)2471–2472。
5. R.Owczarzy，A.V. Tataurov，Y. Wu，J.A. Manthey，K.A. McQuisten，h . g . alma bragi，

《核酸寡聚物的分析和设计套件》，核酸研究1(2008)W163–W169。

1. 米（meter的缩写））《基于工作流的通用引物设计改进平台》，核酸研究36 (2008) e56。
2. M.D. Gadberry，S.T. Malcomber，A.N. Doust，E.A. Kellogg，Primaclade——一种跨多个物种寻找保守PCR引物的灵活工具，《生物信息学》21(2005)1263–1264。
3. J.《一个家族中每个序列的特异性聚合酶链反应引物的自动设计》，BMC生物信息学8 (2007) 369。
4. T.《生物信息学研究》7 (2006) 431。
5. 胡斯利，扎勒夫斯基，贝克特，文塔，影响1147对哺乳动物引物跨物种扩增成功的设计因素，BMC Genomics 7 (2006) 253。
6. R.基因检测的软件支持，Proc Int Conf Intelt Syst Mol Biol 4(1996)68–77。
7. R.基于反转座子的基因分型和指纹识别。协议。1 (2006) 2478–2484.
8. 脱氧核糖核酸中内部G-T错配的热力学和核磁共振，生物化学36(1997)10581–10594。
9. 名词（noun的缩写）内错配DNA序列的最近邻热力学和核磁共振，生物化学38(1999)3468–3477。
10. 聚合物、哑铃和寡核苷酸的最近邻热力学的统一观点。纳特。阿卡德。Sci。美国95(1998)1460–1465。
11. 名词（noun的缩写）改进的热力学参数和螺旋起始因子预测DNA双链体的稳定性，核酸研究24(1996)4501–4505。
12. 南《具有悬垂末端的DNA序列的热力学参数》，《核酸研究》第28期(2000)929-1934。
13. A.序列复杂性和DNA曲率，计算机与化学23(1999)263–274。
14. 《复杂性:一个分析DNA序列复杂性的互联网资源》，《核酸研究》第32期(2004)W628–W633页。
15. 何炳生，脱氧核糖核酸的非B-脱氧核糖核酸结构与Z-脱氧核糖核酸没有区别，程序。纳特。阿卡德。Sci。美国91(1994)9549–9553。
16. 寡核苷酸的四链复合物

四链体，俄罗斯分子生物学杂志26(1992)512–531。

1. D.鸟嘌呤四重奏结构，酶方法。211 (1992) 191–199.
2. 拉希瓦尔，福克斯，四链体熔化，方法43(2007)291–301。
3. 南伯格，帕金森，p .黑兹尔，A.K .托德，k .内德尔，四链体DNA:序列，拓扑和结构，核酸研究34(2006)5402–5415。
4. A.盖丁，格罗，阿尔贝提，梅格尼，多长才算长？环大小对G-四链体稳定性的影响，核酸研究38(2010)7858–7868。
5. O.《预测和理解G-四链体的稳定性》，《生物信息学》25(2009)i374–i382。
6. Y.《模板脱氧核糖核酸的区域化气相色谱含量作为聚合酶链反应成功的预测因子》，核酸研究31 (2003) e99。
7. R.《预测大基因组中聚合酶链反应的失败率》，《核酸研究》第36期(2008年)e66。
8. B.单核苷酸多态性和引物错配对定量聚合酶链反应影响的评估。9 (2009) 75.
9. 《计算结合亲和力的统计热力学基础:一个批判性的评论》，生物物理学。j . 72(1997)1047-1069。
10. 计算核酸双链体解链温度的免费工具，生物信息学17(2001)1226–1227。
11. 脱氧肌苷对在DNA双链体中的最近邻热力学，核酸研究33(2005)6258–6267。
12. 名词（noun的缩写）聚合酶链反应条件下的寡核苷酸熔化温度:Mg2+、脱氧核苷酸三磷酸和二甲基亚砜浓度的最近邻校正与替代经验公式的比较，Clin。化学。47 (2001) 1956–1961.
13. J.威尔士，麦克莱兰，使用任意引物的聚合酶链反应指纹基因组，核酸研究18(1990)7213–7218。
14. 由任意引物扩增的DNA多态性可用作遗传标记，核酸研究18(1990)6513–6535。
15. 张瑞林，欧多诺霍，环保局，基因多态性:一种基于转座子的高通量基因组作图和指纹分析方法，理论。应用基因。102 (2001) 773–781.
16. 《偷渡者:与单子叶和双子叶植物基因相关的反向重复元件的新家族》，植物细胞6(1994)907–916。
17. 纳尔逊，莱德贝特，科尔博，维多利亚，拉米雷斯-索利斯，韦伯斯特，

卡斯凯:一种从复杂的DNA来源中快速分离人类特异性DNA序列的方法。纳特。阿卡德。Sci。美国86(1989)6686–6690。

1. D.《利用Alu特异性引物通过聚合酶链式反应检测的人类DNA多态性》，基因组学7(1990)331–334。
2. J.Jurka，V.V. Kapitonov，A. Pavlicek，P. Klonowski，O. Kohany，J. Walichiewicz，Repbase update，真核重复元件数据库，Cytogentic和Genome RES . 110(2005)462–467。
3. 基因剪接和聚合酶链反应驱动的重叠延伸诱变。协议。2 (2007) 924–932.
4. K.序列标记位点的有序连锁和通过测序进行的多重单核苷酸多态性基因分型，核酸研究30 (2002) e11。
5. J.全，田俊杰，复杂基因文库和途径的环状聚合酶延伸克隆，公共科学图书馆综合4 (2009) e6441。
6. 米（meter的缩写））《用于基因合成的快速、灵活的寡核苷酸设计软件》，核酸研究37(2009)W214–W221。
7. Y.雷，，两步全基因合成法，核酸研究32 (2004) e59。
8. 《体外基因合成的寡核苷酸设计》，核酸研究32(2004)W176–W180。
9. A.尤耶夫，黄，波尔，帕奇，沃森，贝尔，唐纳森，菲利普斯，博伊斯-杰辛诺，用统计模型预测引物延伸基因分型分析的成功，核酸研究30 (2002) e131。
10. 《用于增强引物设计的转录组和基因组计算机聚合酶链反应》，生物信息学20(2004)2399–2400。
11. Y.曹立军，王立军，徐克军，寇春英，张光伟，何建军，王立军，赵立军，基于信息论的全基因组序列模板PCR产物计算机预测算法，BMC生物信息学6 (2005) 190。
12. E.使用复杂模板的聚合酶链反应的数学模型和计算机模拟，核酸研究24(1996)3538–3545。
13. 米（meter的缩写））莱克萨，g .瓦莱，PRIMEX:全基因组寡核苷酸匹配的快速鉴定，生物信息学19(2003)2486–2488。
14. K.全基因组序列对大肠杆菌随机聚合酶链反应产物序列的预测，核酸研究28(2000)1879–1884。