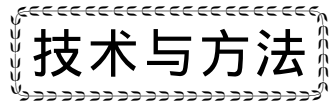


# BioSun2.0 : 一个综合性的辅助分子生物学实验 设计软件



技术与方法

查磊, 应晓敏, 曹源, 李华, 李伍举\*

(军事医学科学院基础医学研究所计算生物学中心, 北京 100850)

[摘要] 我们曾于2004年推出了计算机辅助分子生物学实验设计的软件系统 BioSun 1.0, 该系统提供了较为全面的数据处理与分析功能。为了更好地服务于生物医学工作者, 我们对该软件系统进行了升级, 推出了2.0版本, 新增的功能主要有: 基于 Blast 的多种形式的序列比对、基于 ClustalW 的多序列比对与进化树构建、蛋白质三维结构展示、基于 RNAfold 的 RNA 二级结构预测和序列格式转换等。通过与商业化综合性的生物信息学软件系统 DNASIS MAX 2.05、DNASStar 5.0、Vector NTI 9.1 和 BioEdit 7.0 的比较发现, BioSun2.0 具有操作简便、功能众多和性价比高等特点, 能够满足生物医学实验室的常规需求。

[关键词] BioSun; 生物信息学; 计算机辅助设计

[中图分类号] Q811.4

[文献标识码] A

[文章编号] 1000-5501(2006)05-0461-04

## BioSun2.0 : a comprehensive software for molecular biology experiments

ZHA Lei, YING Xiao-Min, CAO Yuan, LI Hua, LI Wu-Ju\*

(Center of Computational Biology, Institute of Basic Medical Sciences, Academy of Military Medical Sciences, Beijing 100850, China)

[Abstract] BioSun1.0 was released by in 2004. It is a comprehensive bioinformatics software of computer-aided design of molecular biology experiments. In order to provide better tools for biomedical scientists, the BioSun have been updated recently to version 2.0. The functions such as Blast-based comparison, multiple sequence alignments, construction of phylogenetic tree, 3D representation of a protein, RNA secondary structures prediction, sequence format conversion and others were added in the new version. Finally, a detailed comparison of BioSun2.0 with the commercial software, which includes DNASIS MAX 2.05, DNASStar 5.0, Vector NTI 9.1 and BioEdit 7.0 was provided. The results show that BioSun2.0 has the characteristics such as easy operation, multiple functions, and high quality/price ratio and can fulfill the common requirements of biomedical Labs.

[Key words] BioSun; bioinformatics; computer-aided design

随着生物信息学的不断发展,特别是人类基因组与模式生物基因组计划的实施,无论是从数量上还是质量上都极大地丰富了生物科学的数据资源,迅速增长的数据越来越需要计算机快速的处理能力,因此,各种辅助实验设计的软件系统也不断推出,以帮助分子生物学工作者解决他们在实验中所遇到的各种问题。

我们曾于1993年在国内首次推出了辅助实验设计的核酸和蛋白质序列分析系统 Goldkey<sup>[1]</sup>,到目前为止,已推广到上百家科研院所。通过查询清华同方全文数据库,发现 Goldkey 软件已被引用 400 多次。2004 年,我们又对 Goldkey 软件进行了全面改进,推出了 BioSun 软件系统<sup>[2]</sup>。为了进一步完善 BioSun 系统,满足分子生物学实验人员的需求,对 BioSun 系统进行了升级,推出了 BioSun2.0 版本,新增了基于 Blast 的多功能序列比对、多序列比对与进化树构建、蛋白质三维结构展示、RNAfold 二级结构预测和序列格式转换等

功能。在本文中,我们首先简要描述 BioSun2.0 系统的新增功能,然后与常用的生物信息学软件进行了比较,介绍了 BioSun2.0 系统的主要特色( BioSun 软件的详细情况见 <http://www.biosun.org.cn>)。

### 1 BioSun2.0 系统的新增功能

为了更好地为分子生物学工作者服务,我们根据分子生物学实验的实际需求和用户反馈,对 BioSun 系统进行了升级,增加了下列新的功能。

[收稿日期] 2006-03-20

[基金项目] 国家自然科学基金项目(30470411)和北京市自然科学基金项目(5042021)

[作者简介] 查磊(1983-),男,云南省大理市人,研究实习员。

\* 通讯联系人, Tel: 010-66931324; E-mail: liwj@nic.bmi.ac.cn

### 1.1 基于 Blast 的多功能序列比对

Blast 序列比对是分子生物学实验中常用的一个功能,但传统的 Blast 程序是基于命令行模式的,命令及参数多而复杂,要完成一次比对要输入比较复杂的指令,对于实验人员来说使用非常不方便。因此,我们采用图形用户界面集成了 Blast 程序,用户只需在窗口上选择比对数据及各种参数,即可进行比对,并且,在此基础上,我们实现了多种方式的序列比对:第一工作区序列与第二工作区序列比对,第一工作区序列(或第二工作区序列)与数据库比对,数据库与数据库的比对。

### 1.2 多序列比对及进化树构建

在多序列比对及进化树构建方面,BioSun2.0 提供了一套完整的流程,用户只需输入一个包含多个序列的 FASTA 格式的文本文件并选择参数后,即可进行多序列的比对分析(图1),在比对完成后,点击“View Tree”按钮,即可以构建对应的进化树(图2)。

### 1.3 蛋白质三维结构展示

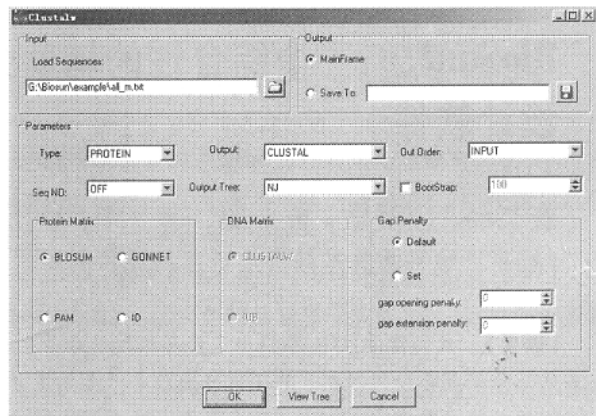


图1 BioSun2.0 多序列比对与进化树构建界面

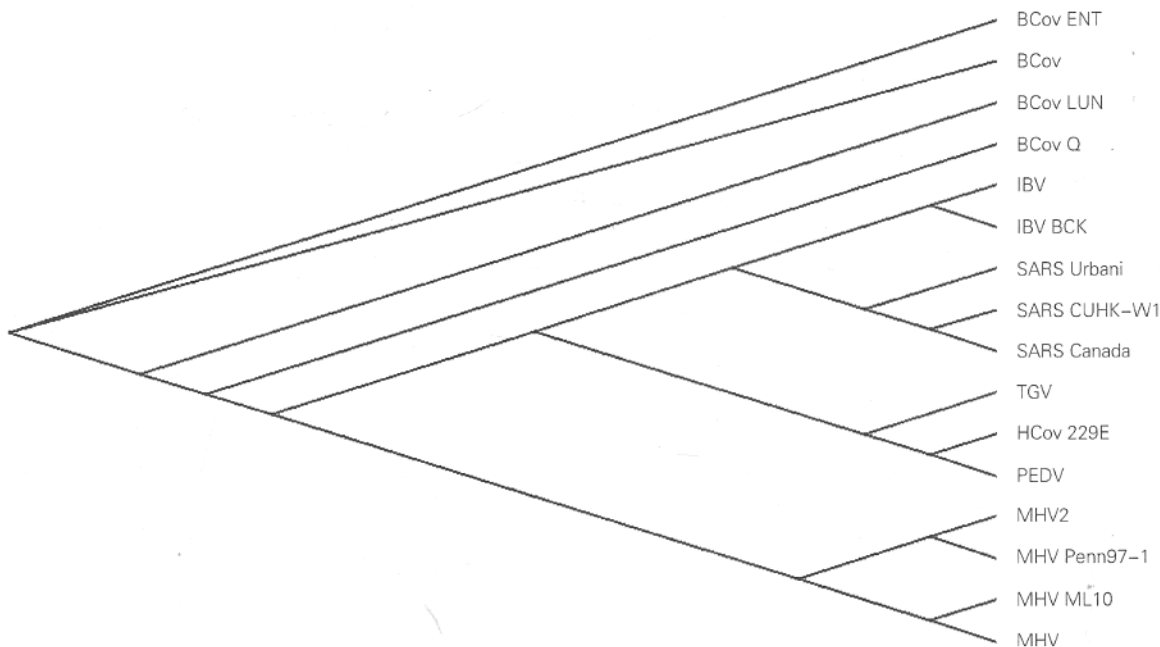


图2 SARS 冠状病毒的进化树

对于蛋白质 PDB 三维结构文件,BioSun2.0 可以对其进行三维结构展示,同时可以进行旋转,缩放,选择,发送图像至主窗口,保存图像等操作,方便实验人员观察。

### 1.4 RNAfold 二级结构预测

在 RNA 二级结构预测方面,在原有的 RNA 二级结构预测方法的基础上,增加了 Vienna RNAfold<sup>[3-6]</sup> 二级结构预测算法,并加入了螺旋区配对文件—CT 文件的自动生成功能。

### 1.5 序列格式转换

BioSun2.0 提供了比较全面的序列管理功能,可以接受 GenBank,EMBL 和 FASTA 等序列格式,并可以对这些格式进行批量处理,存入 BioSun 数据库中,同时,也提供了从 BioSun 数据库向 GenBank,EMBL,FASTA,SwissProt 格式的批量转换。

## 2 BioSun2.0 与常用综合性分析软件比较

我们选取了目前比较常用的综合性分析软件:DNASIS MAX 2.05, DNASStar 5.0, Vector NTI 9.1, BioEdit 7.0<sup>[7]</sup> 作为代表,与 BioSun2.0 进行了比较。

### 2.1 常用软件简介

DNASIS MAX 是日立软件公司(Hitachi Software Engineering Co, Ltd.)开发的一套综合性分析软件,界面友好,简单易用,同时对 Perl 提供了良好的支持,DOS 时代就已经比较流行。

DNASStar 是 DNASStar 公司(DNASStar Inc.)于 1989 年推出的一套综合性分析软件,经过不断改进,现在最新版本由 EditSeq, GeneQuest, MapDraw, MegAlign, PrimerSelect, Protean, SeqMan 7 个部分组成,分别进行序列编辑、序列特性分

析、酶切图制作、序列比对与进化树构建、引物与探针设计、蛋白质性质分析与预测、序列装配与编辑输出。特别值得一提的是,它的序列装配程序能够对自动测序仪产生的数据进行评价,自动完成装配。

BioEdit 是一个直观的、使用简单的综合性分析软件,它是由北卡罗来纳州大学微生物系的汤姆·霍尔博士开发的,其特有的 BioEdit 文件格式可以快速处理包含大量序列的文件。

Vector NTI 是由 Invitrogen 公司所属的专门从事生物信息学相关开发的 InforMax 公司推出的一个综合性分析软件包,功能强大,但相对的操作也比较复杂。

## 2.2 与常用软件比较分析

选取了生物信息学软件常用的 3 个主要部分:序列编辑

与管理功能、核酸相关分析和蛋白质相关分析进行比较,比较结果见表 1。从表 1 中可以看出,与常用的国外综合性分析软件相比,BioSun2.0 不仅涵盖了其他软件的绝大部分功能,还包含了下列新的功能:DNA 和蛋白质序列的 k-tuple 分析、信号肽预测、外源基因高效表达设计、寡核苷酸芯片的探针设计、cDNA 微阵列的引物设计和基于基因表达谱的差异基因识别等功能;其次,BioSun2.0 的界面友好,操作简单,便于生物医学工作者使用。以常用的多序列比对与进化树构建为例,BioSun2.0 将多序列比对与进化树构建集成到了一个界面上,操作起来更加便捷;又如,在预测 RNA 二级结构时,用户可以选择自动给出螺旋区配对的 CT 文件,方便了用户的使用。

表 1 BioSun 与常见综合性分析软件比较

功 能	软 件 名 称					
	DNASIS	DNASar	BioEdit	Vector NTI	BioSun	
序列管理	常见编辑功能	√	√	√	√	√
	数据库管理	×	×	×	√	√
核酸	组成分析	√	√	√	√	√
	k-tuple 分析	×	×	×	×	√
	互补链/反向链	√	√	√	√	√
	翻译	√	√	√	√	√
	ORF 查找	√	√	√	√	√
	重复序列搜索	√	√	×	√	√
	两两比对	√	√	√	√	√
	多序列比对	√	√	√	√	√
	进化树构建	√	√	√	√	√
	RNA 二级结构预测	√	√	×	√	√
	酶切分析	√	√	√	√	√
	质粒绘图	√	√	√	√	×
	引物设计	√	√	√	√	√
	转录因子结合位点预测	×	√	×	×	√
	Motif 检索	√	√	√	√	√
	序列拼接	√	√	×	√	×
	模拟电泳	×	√	×	√	×
	外源基因高效表达设计	×	×	×	×	√
	寡核苷酸芯片探针设计	×	×	×	×	√
	cDNA 芯片引物设计	×	×	×	×	√
基于基因表达谱的差异基因识别	×	×	×	×	√	
蛋白质	组成分析	√	√	√	√	√
	k-tuple 分析	×	×	×	×	√
	两两比对	√	√	√	√	√
	多序列比对	√	√	√	√	√
	进化树构建	√	√	√	√	√
	亲水/疏水性分析	√	√	√	√	√
	信号肽预测	×	×	×	×	√
	抗原表位预测	×	√	×	×	√
	功能位点分析	×	×	×	√	√
	三维结构展示	×	×	×	√	√
统计(√)	20	23	17	24	30	

注:√代表有此功能,×代表无此功能

### 2.3 BioSun 的特色

与其他常用的软件相比, BioSun2.0 系统不仅实现了分子生物学中常见的算法, 也包含了我们自己在长期研究积累过程中提出的一些算法。例如蛋白质功能位点分析<sup>[8]</sup>、基于螺旋区分布的 RNA 二级结构预测<sup>[9]</sup>、外源基因高效表达设计<sup>[10, 11]</sup>、辅助寡核苷酸微阵列探针设计<sup>[12]</sup>和辅助 cDNA 微阵列引物设计等, 其中有些方法已经在实验中得到了多例验证。以 RNA 二级结构预测为例, 与其他软件系统相比, 在 BioSun2.0 中不仅集成了常见的 Vienna RNAfold 二级结构预测算法, 也包含有我们自己的基于螺旋区分布的 RNA 二级结构预测算法<sup>[9]</sup>, 此算法在实际序列的 RNA 二级结构预测中, 特别是短序列的预测中, 有效性得到了验证。又如抗原表位预测功能, 在 BioSun2.0 中, 对于抗原表位预测, 综合性的考虑了各种因素, 通过对 5 个蛋白的测试表明, 预测准确率达到了 45.8%<sup>[2]</sup>。另外, 外源基因高效表达设计等功能, 在多例实验中也得到了有效的验证<sup>[13-19]</sup>。

### 3 总结

通过与常用综合性分析软件比较可以看出, 升级以后的 BioSun2.0 已经涵盖了分子生物学实验所需的各项常用功能, 基本满足了分子生物学实验人员的常规需求, 与国外的综合性分析软件相比, 在实现同类软件常见功能的基础上, 也突出了自己的特色, 是我们长期研究工作的一个积累与体现。同时, 多例实验证明, 采用生物信息学手段对实验进行分析与辅助设计, 在现代分子生物学研究中显得尤为重要, 两者之间的关系也变得越来越紧密, 这是生物学研究的一个发展趋势。今后, 我们将基于现有的基础, 为分子生物学实验提供更强有力的帮助。

#### [ 参考文献 ]

- [ 1 ] 吴加金, 李伍举, 雷红星, 等. 核酸和蛋白质序列分析的软件系统——Goldkey[ J ]. 生物技术通讯, 1994, 5( 4 ): 189 - 193.
- [ 2 ] 李伍举, 应晓敏. BioSun 计算机辅助分子生物学实验设计的软件系统[ J ]. 军事医学科学院院刊, 2004, 28( 5 ): 401 - 404.
- [ 3 ] Hofacker IL, Fontana W, Stadler PF, et al. Fast folding and comparison of RNA secondary structures[ J ]. Monatsh Chem, 1994, 125( 2 ): 167 - 188.
- [ 4 ] Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information[ J ]. Nu-

cleic Acids Res, 1981, 9( 1 ): 133 - 148.

- [ 5 ] McCaskill JS. The equilibrium partition function and base pair binding probabilities for RNA secondary structure[ J ]. Biopolymers, 1990, 29( 6 - 7 ): 1105 - 1119.
- [ 6 ] Hofacker IL. Vienna RNA secondary structure server[ J ]. Nucleic Acids Res, 2003, 31( 13 ): 3429 - 3431.
- [ 7 ] Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT[ J ]. Nucl Acids Symp Ser, 1999, 41: 95 - 98.
- [ 8 ] 李伍举, 吴加金. 蛋白质功能位点预测[ J ]. 生物化学与生物物理进展, 1993, 20( 1 ): 60 - 62.
- [ 9 ] Li WJ, Wu JJ. Prediction of RNA secondary structure based on helical regions distribution[ J ]. Bioinformatics, 1998, 14( 8 ): 700 - 706.
- [ 10 ] Li WJ, Lei HX, Pei WH, et al. GeneDn: for high level expression design of heterologous genes in a prokaryotic system[ J ]. Bioinformatics, 1998, 14( 10 ): 884 - 885.
- [ 11 ] 李伍举, 吴加金. pBV220 载体中外源基因表达水平定量分析[ J ]. 病毒学报, 1997, 13( 2 ): 126 - 133.
- [ 12 ] Li WJ, Huang J, Fan M, et al. MProbe: computer aided probe design for oligonucleotide microarray experiment[ J ]. Appl Bioinform, 2002, 1( 3 ): 163 - 166.
- [ 13 ] 裴武红, 沈倍奋, 李伍举, 等. 计算机辅助设计使重组 ricin2A 链在 *E. coli* 中的高效表达[ J ]. 细胞与分子免疫学杂志, 1998, 14( 1 ): 33 - 36.
- [ 14 ] 裴武红, 胡美茹, 李伍举, 等. 人 FKBP12 基因克隆及其表达产物的生物活性[ J ]. 中国生物化学与分子生物学报, 2000, 16( 3 ): 322 - 325.
- [ 15 ] 刘淑红, 孙长凯, 张玉梅, 等. 人 NR1 靶片段的原核表达[ J ]. 军事医学科学院院刊, 2002, 26( 3 ): 188 - 190.
- [ 16 ] 黄坚, 陈苏红, 佟丽, 等. 小鼠细胞因子相关基因表达检测寡核苷酸芯片的制备及分析[ J ]. 生物工程学报, 2002, 18( 4 ): 501 - 504.
- [ 17 ] 洪海燕, 赖春宁, 冯健男, 等. 影响人干细胞因子基因高效表达的因素探讨[ J ]. 免疫学杂志, 2002, 18( 3 ): 225 - 228.
- [ 18 ] 李泽松, 李德良, 黄坚, 等. 心血管相关基因芯片的制备及其在知母皂荚作用机理研究中的应用[ J ]. 药学报, 2003, 38( 7 ): 496 - 500.
- [ 19 ] 佟丽, 陈苏红, 马增春, 等. 用寡核苷酸芯片研究血虚小鼠造血相关细胞因子表达谱[ J ]. 中草药, 2003, 34( 7 ): 625 - 629.

( 本文编辑 孙承媛 )

· 读者 · 作者 · 编者 ·

## 离心机转速的写法及相对离心力的正确表示

有的作者将离心机转速用 rpm 表示, 这是非法定单位的写法。正确的写法是: 当低速离心时用 r/min 表示, 当高速或超高速离心时用相对离心力“ × g ”表示, g 是斜体, 前面是乘号。

( 本刊编辑部 )