

doi:10.3969/j.issn.1672-5565.2014.04.12

生物信息学数据库研究文献引文与热点分析

王蕊¹,胡德华^{1,2*}

(1.中南大学湘雅医学院,长沙 410013;

2.湖南省高等学校医学信息研究重点实验室,长沙 410013)

摘要:以 Web of Science 为数据源,简要概括生物信息学数据库研究的发展趋势。利用 CiteSpace 可视化工具展现生物信息学数据库研究的知识基础和研究热点图谱,为开展生物信息学数据库领域相关的理论研究和实践活动提供借鉴,以便推动生物信息学数据库研究的发展。研究表明:1990年 Altschul SF 发表的“局部比对搜索工具——BLAST”是生物信息学数据库研究的重要知识来源文献;热点主题集中在序列库、基因组数据库、分类数据库、蛋白质数据库、数据库更新、集成系统等。

关键词:生物信息学数据库; CiteSpace; 研究热点; 可视化分析

中图分类号:R-1 文献标志码:A 文章编号:1672-5565(2014)-04-305-08

Analysis of citation features in bioinformatics database literature studies

WANG Rui¹, HU Dehua^{1,2*}

(1.XiangYa School of Medicine, Central South University, Changsha 410013, China;

2.Key Laboratory of Medical Information Research, College of Hunan Province, Changsha 410013, China)

Abstract: Using literature data from Web of Science as the research sample, we briefly summarize the trend of bioinformatics database research to provide reference for theory research and practice activities and to promote the development of bioinformatics database research. Then we use information visualization tools named CiteSpace to display the knowledge sources and research focuses. The results show that the earliest literature was Basic local alignment search tool by Altschul SF in 1990; the research focuses include sequence databases, genome database, classification database, protein database, database updates, integration system, etc.

Keywords: Bioinformatics database; CiteSpace; Research focus; Visualization analysis

生物信息学(Bioinformatics)是在生命科学的研究中,生物学与计算机科学及应用数学等多学科相互交叉而形成的一门新兴的综合性学科,包含生物信息的获取、加工、存储、分配、分析、解释等诸多方面^[1]。生物信息学数据库是生物信息学的重要内容之一。1960年左右, Margaret Dayhoff 创立了国际蛋白质序列数据库(PSD)。1982年, Los Alamos 建立了第一个核酸序列数据库 GenBank。随着生物信息学的不断发展,生物信息学数据库层出不穷,覆盖了多个领域,已成为分子生物学、生物医药工业、农业科学和环境科学等学科的必不可少的重要成分^[2]。生物信息学数据库种类繁多,分类方法各有

不同。依据数据类型可分为:核酸序列数据库、蛋白质序列数据库、三维分子结构数据库。按数据层次可分为:基本数据库、复合数据库、二次数据库。按其来源可分为原始数据库、衍生数据库、集成数据库和知识数据库。

近年来,生物信息学数据库的研究得到了广泛关注。姜鑫研究了生物信息学数据库及其利用方法,介绍了世界上主要的生物信息学数据库及其分类和特点以及如何利用生物信息学数据库^[3]。万跃华等研究了网上生物信息学数据库资源,对主要的热门生物信息学数据库站点和资源进行了评价^[4]。胡德华等进行了生物信息学数据库调查分

收稿日期:2014-06-16;修回日期:2014-11-12。

基金项目:国家社会科学基金项目(No.11BTQ044)。

作者简介:王蕊,女,硕士研究生,研究方向:情报学;E-mail:843681179@qq.com。

*通信作者:胡德华,男,博士,教授,博士生导师,研究方向:信息检索;E-mail:hudehua2008@qq.com。

析及其利用研究^[5]。杨长平等对 Web of Science 收录生物信息学数据库研究文献进行了分析^[6],但其主要采用计量学方法对生物信息学数据库研究文献的年代分布、语种、期刊分布、作者、文献类型、主题分布等进行了数量上的统计,尚未从深层次上探讨。国外学者 Whitfield EJ 进行了生物信息学数据库基础设施的研究,提出其不同的类型,即核酸序列和蛋白质序列的基本类型数据库、蛋白质结构数据库、蛋白质鉴定的专业型数据库和大型数据集成数据库^[7]。此外,其他学者多是对单一数据库的原理、应用与更新进行了分析,如 Mavridis L 介绍了参数免聚类算法——PFClust^[8]; Wattam AR 分析了细菌生物信息学数据库——PATRIC 及其应用^[9];以及家庭蛋白数据库的更新发展变化^[10-12]。

目前,国内外学者基于知识图谱对生物信息学数据库研究文献引文与热点进行可视化分析的文献非常少。本文主要运用 CiteSpace 软件对生物信息学数据库研究文献引文与热点进行分析,希望能为开展生物信息学数据库领域相关的理论研究和实践活动提供借鉴,以便推动生物信息学数据库研究的发展。

1 材料与方法

1.1 数据来源

本研究以 Web of Science 为数据源,检索策略为:出版物名称=(nucleic acids research);文献类型=(ARTICLE OR REVIEW);索引=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CCR-EXPANDED, IC;时间跨度=2004~2014年;分别选择 Database Issues。检索时间为2014年4月29日,共检索到有效文献1856篇。

本研究选择《Nucleic Acids Research》期刊为研究对象。原因有以下几点:(1)《Nucleic Acids Research》是国际权威杂志之一,其2014年的影响因子为8.278,主要刊登核酸和核酸代谢以及相互作用有关蛋白质方面的物理、化学、生化和生物领域中领先研究的重要成果;(2)《Nucleic Acids Research》自2004年起,每年单独出版“Database Issue”一期,重点探讨生物信息学数据库研究的发展,尚无其他期刊可比拟,从而确定本研究的专业性和准确性;(3)已有研究证明,《Nucleic Acids Research》是生物信息学数据库研究的重要期刊^[6],因此证明了此期刊的权威性。

1.2 分析方法

(1)文献计量学方法是常用的统计学方法,其能描述或解释文献的数据特征和变化规律。本文采

用文献计量学方法对生物信息学数据库研究的年文献量、新建与更新数量(每年)进行统计,探讨生物信息学数据库的总体发展趋势及更新趋势。

(2)CiteSpace 是用来分析和可视共引网络的 Java 应用程序,能够显示一个学科一定时期发展的趋势与动向,形成若干研究前沿领域的演进历程^[13]。已成为信息分析领域中影响力较大的信息可视化软件。

(3)知识基础分析有利于研究某领域的前沿和本质,由共引聚类来表示,共引聚类阈值越低,其共引聚类中的文献越多,从而知识基础越大并趋于稳定状态。早期奠基性文献、共被引频次和中心性较高的文献是该领域的重要知识基础。本文利用 CiteSpace 软件的引文共被引分析功能绘制引文共被引聚类图谱,探索生物信息学数据库研究的知识基础。

(4)生物信息学数据库研究热点与前沿分析:绘制关键词共现的知识图谱,识别出生物信息学数据库研究的热点主题。

2 结果与分析

2.1 发展阶段分析

生物信息学数据库研究的总体趋势趋于稳定,呈小幅度的增长趋势。可划分三个阶段:2004~2008年,文献量逐年缓慢上升,这说明《Nucleic Acids Research》期刊重视生物信息学数据库的研究,且关注加大;2009~2010年,文献量显著减少,说明这一阶段生物信息学数据库研究遇到了瓶颈期,导致成果产出减少,当然也不能排除选择单一年刊因素的影响;2011~2014年,文献量波动期,说明生物信息学数据库研究达到了稳定状态(见图1)。因此,可预测未来几年的生物信息学数据库研究的文献量将保持稳定的状态。

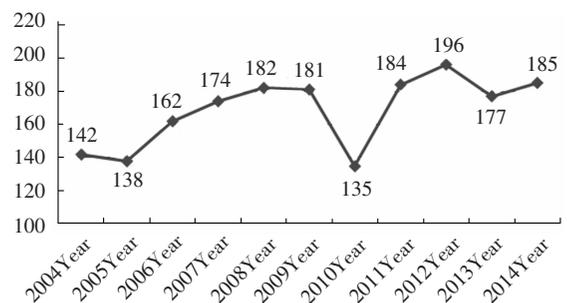


图1 生物信息学数据库研究的发展趋势

Fig.1 The development trend of bioinformatics database research

2.2 更新趋势分析

图 2 显示的是 2006 ~ 2014 年《Nucleic Acids Research》中数据库的年更新趋势。

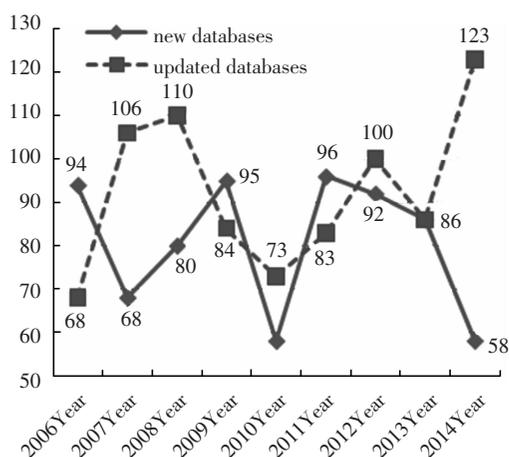


图 2 《Nucleic Acids Research》中数据库的更新趋势

Fig.2 The updated trend of database in the Nucleic Acids Research

由于《Nucleic Acids Research》始于 2004 年出版“Database Issue”一期,因此 2006 年开始对该杂志

的数据库更新数量进行对比。实线表示每年新建生物信息学数据库的数量,虚线表示每年对原有生物信息学数据库的更新数量。总体上,原有数据库更新的数量超过新建数据库的数量。原因是 2006 年以前已建立数百种生物信息学数据库,而生物信息学数据库的研究势必建立在前人的基础上,以拓展新数据库的研发。新建数据库的数量和更新数据库的数量变化每年不定,没有形成规律。这说明数据库的研究仍处在发展阶段,尚未达到顶峰。此外发现,新建数据库所占比例与更新数据库所占比例相差不大,这说明随着生物信息学的发展,生物信息学数据库等工具的研究已成为研究热点,研究者智力研究新型数据库以适应生物信息学的发展与需求。

2.3 引文分析

2.3.1 早期奠基性文献

网络节点确定为被引参考文献,时间分割设为 1 年,引文筛选标准设置为 2004 ~ 2013 年,每年文献引文集合中的 Top 20% 高被引文献,选择时间轴视图,得到生物信息学数据库研究的早期奠基性文献时间序列知识图谱,见图 3。

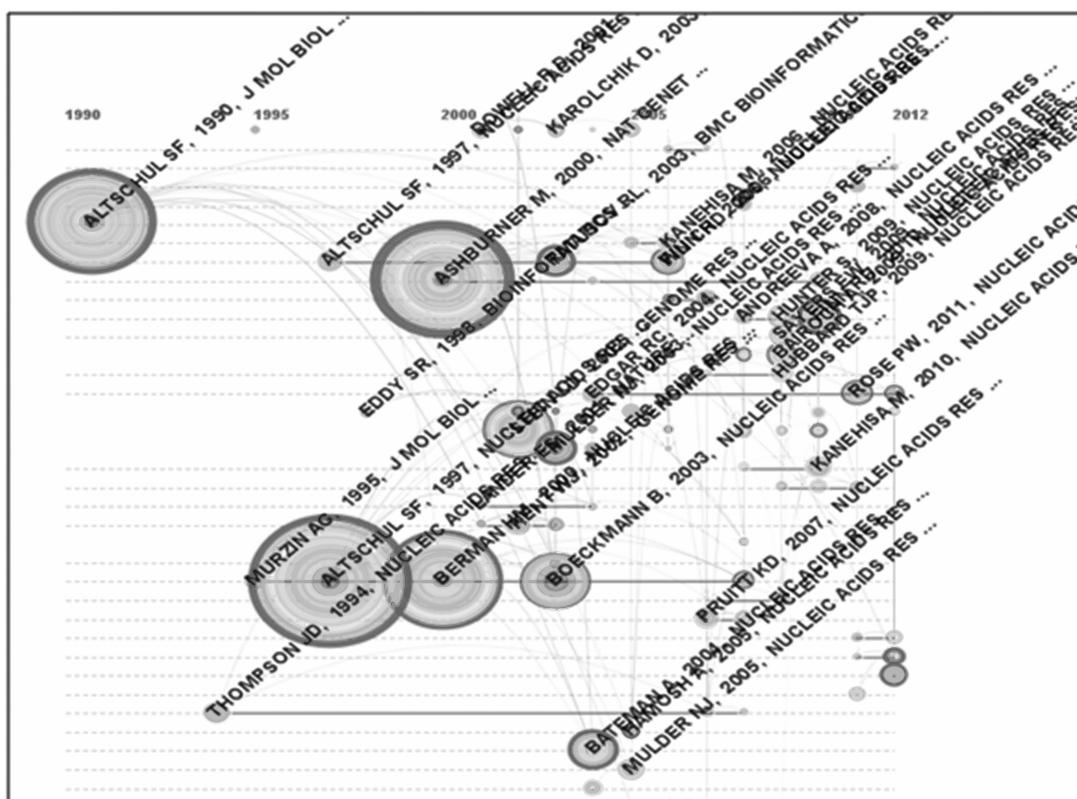


图 3 生物信息学数据库研究早期奠基性文献时间序列知识图谱

Fig.3 The time sequence knowledge map of early founding documents in bioinformatics database study

图谱中越靠近左边的节点,说明文献发表的年代越早。以下列举的是起源较早的三篇高被引文献,第一篇奠基性文献是 Altschul SF 于 1990 年在 J Mol

Biol 上发表的 Basic local alignment search tool,首次提出快速对比序列的新方法——BLAST 是局部比对搜索工具,其结合数学算法(MSP)的优点,提高了搜索

速度,同时把数据库搜索建立在严格的统计学基础之上^[14]。它可以在各种语境下以多种方式实现简单的DNA和蛋白质序列数据库搜索、主题搜索、基因识别搜索和多个区域相似DNA序列的分析等。此数据库也是目前最常用的同源检索工具。第二篇对生物信息学数据库研究产生重大影响的奠基性文献是THOMPSON JD于1994年在Nucleic Acids Res上发表的CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice,指出CLUSTAL W软件能够通过序列加权、位置特异空位罚分和权重矩阵,提高多重序列对比的敏感性^[15]。第三篇对生物信息学数据库研究产生重大影响的奠基性文献是MURZIN AG于1995年在J Mol Biol上发表的SCOP: a structural classifica-

tion of proteins database for the investigation of sequences and structures,该文介绍了蛋白质结构分类数据库——SCOP^[16]。该数据库详细描述了已知蛋白质结构之间的关系,并且提供坐标链接、结构图像、交互式查看器、序列数据和文献引用等功能。该数据库提供了两个搜索功能:同源性搜索允许用户输入序列和获得相似序列的结构列表;关键字搜索将用户输入的单词与scop数据库的文本和Brookhaven蛋白质数据库结构的文本进行匹配。

2.3.2 高被引文献

网络节点确定为被引参考文献,时间分割设为1年,引文筛选标准设置为2004~2013年,每年文献引文集集中的Top 10%高被引文献,在生成图谱的可视化窗格中将节点标签选择为按被引频次显示,可得到生物信息学数据库研究的高被引文献知识图谱,见图4。

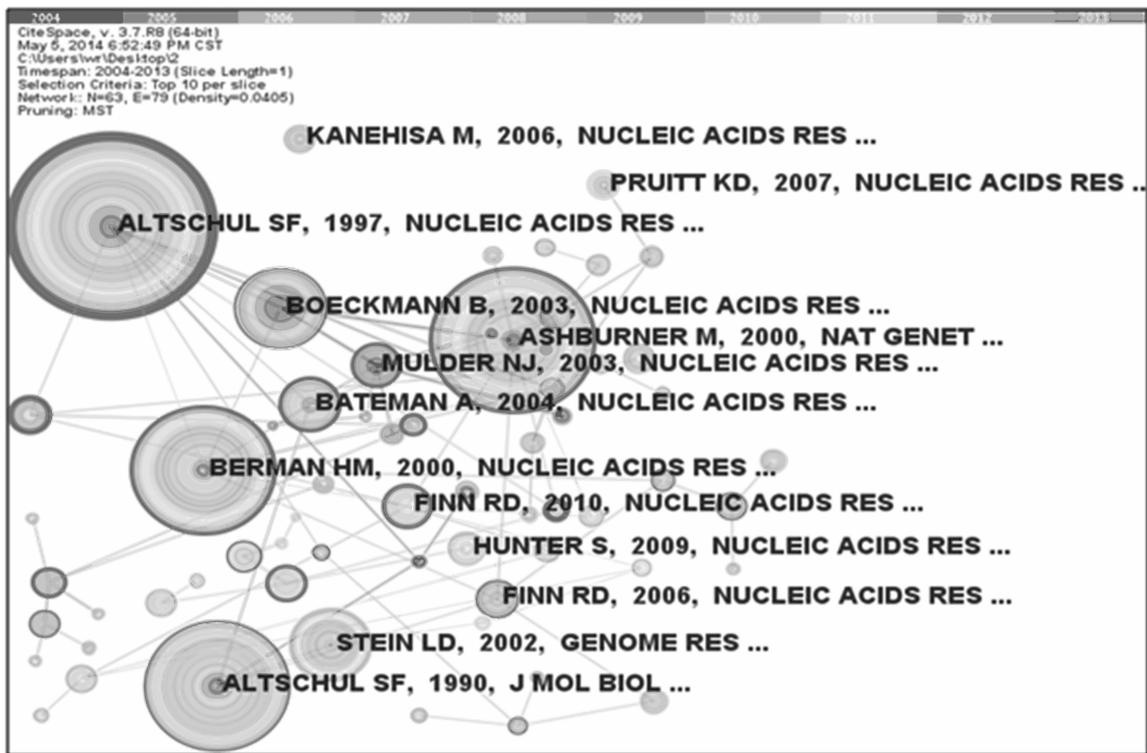


图4 生物信息学数据库研究高被引文献知识图谱

Fig.4 The knowledge map of highly cited documents in bioinformatics database study

图4中显示的节点是被引频次45次以上的高被引文献。图谱中的圆圈大小表示引文的被引频次,圆圈之间连线的粗细表示两个引文的共被引频次。其中被引频次大于100的共有6篇,由高到低排列依次为:(1)1997年ALTSCHUL SF发表的“新一代的蛋白质数据库搜索程序——Gapped BLAST和PSI-BLAST”^[17],指出Gapped BLAST允许在对准的序列中引入空位,即在比较两个相关序列时不会中断现象,从而反映相关序列的类似程度,PSI-

BLAST是特殊位置重复BLAST,它提供了自动、易用的概貌检索,是查找序列同源的有效方法,其被引频次为254次。(2)2000年ASHBURNER M发表的名为“Gene ontology: tool for the unification of biology. The Gene Ontology Consortium (GO——生物学的统一化工具)”的文章^[18],介绍了GO数据库主要包括3大独立的本体:生物过程、分子功能及细胞组分,其被引频次为207次。(3)1990年Altschul SF发表的名为“Basic local alignment search tool

(BLAST——局部比对搜索工具)”的文章,被引频次为173次。(4)2000年BERMAN HM发表的蛋白质数据库(被引频次为164次)^[19]。(5)2002年STEIN LD发表的使用GBrowse构建生物模型系统数据库(被引频次为119次)^[20]。(6)2003年BOECKMANN B发表的SWISS-PROT蛋白质知识库和2003年TrEMBL的扩充(被引频次为104次)^[21]。与图2对比发现,Altschul SF发表的关于BLAST的文献不仅是早期奠基性文献,也是高被引频次文献,因此可推断其是生物信息学数据库研究的重要知识来源文献。

2.3.3 高中心性文献

网络节点的中心性指网络中经过某点并连接这两点的最短路径占这两点之间的最短路径线总数之比,是用来描述网络节点中心性测度的指标,中心性测量为发现学科研究或主题演进网络中的支点提供了计算依据^[22]。

在生成图4时发现,生物信息学数据库研究领域高中心性文献有10篇,按照文献节点的中心性大小进行排列,选取中心性>0.2,详见表1。由表1和图4对比发现,“Gapped BLAST and PSI-BLAST: a new generation of protein database search programs”、“Gene ontology: tool for the unification of biology. The Gene Ontology Consortium”既是高被引文献,又是高中心性文献。因此,这些文献也是生物信息学数据库研究领域的核心知识来源。

此外发现,Batema A、Finn RD和Punta M分别在2004年、2008年、2010年、2012年发表同名为“The Pfam protein families database”的文章都是高中心性文献。这说明生物信息学领域对蛋白质家族数据库——Pfam的研究关注较大,并且更新速度较快。这也许与近年来蛋白质家族发现的日益增多有关,由于蛋白质的种类逐渐增多,成分逐渐细化,这就要求其数据库的功能不断增强。

表1 生物信息学数据库研究高中心性文献

Table 1 The literatures with high centrality in bioinformatics database study

序号	被引频次	中心性	年代	作者	文献名称
1	254	0.87	1997	Altschul SF	Gapped BLAST and PSI-BLAST: a new generation of protein database search programs
2	68	0.45	2004	Batema A	The Pfam protein families database
3	207	0.40	2000	Ashburne M	Gene ontology: tool for the unification of biology. The Gene Ontology Consortium
4	17	0.37	2008	Finn RD	The Pfam protein families database
5	34	0.35	2013	Apweiler R	Gene Ontology annotations and resources.
6	43	0.33	2009	Bairoch	The Universal Protein Resource (UniProt) 2009.
7	57	0.31	2010	Finn RD	The Pfam protein families database
8	44	0.30	2011	Rose PW	The RCSB Protein Data Bank: redesigned web site and web services
9	164	0.29	2000	Berman HM	The Protein Data Bank.
10	54	0.29	2003	Mulder NJ	The InterPro Database, 2003 brings increased coverage and new features
11	55	0.25	2012	Punta M	The Pfam protein families database

2.4 生物信息学数据库研究的热点与前沿分析

关键词是文献内容的精炼,高频关键词常被用来确定一个研究领域的热点问题。中心性高的关键词在一定程度上也代表研究领域的热点内容。中心性较高的词体现在整个网络中的重要连接作用,即两个节点之间的最短路径,关键词占据这样的位置越多,说明其在整个网络中的连接作用越强。通过对文献题录中的关键词进行分析,运行CiteSpace软件,对生物信息学数据库研究领域的高频次、高中心性关键词进行分析,绘制生物信息学数据库研究热点知识图谱,见图5。

图5显示,生物信息学数据库研究较高频次的关键词为:序列库、染色体识别、基因组数据库、分类数据库、蛋白质数据库、数据库更新、集成系统等。按照中心性高低排列,主要有:基因组数据库、序列

库、蛋白质数据库、结构数据库,其中心性分别是0.27、0.26、0.23、0.15。对比高频次关键词和高中心性关键词发现,高中心性关键词大都是高频关键词。

研究发现,近年来生物信息学数据库研究主要集中在基因组生物信息学数据库(人类基因组数据库GDB、线虫基因组数据库AceDB、美国基因组研究所的数据库TIGR、美国国家基因组资源中心基因组序列库GSDB及其他数据库)、生物信息学序列数据库(核酸序列数据库、EMBL和GenBank数据库格式、蛋白质序列数据库及其他蛋白质序列数据库)、生物信息学结构数据库(蛋白质三维结构数据库PDB、蛋白质三维结构分类数据库SCOP和CATH)、生物信息学二次数据库(基因组生物信息学二次数据库、蛋白质序列二次数据库、蛋白质结构二次数据库)和生物信息学数据库的集成系统等^[3-4,23]。

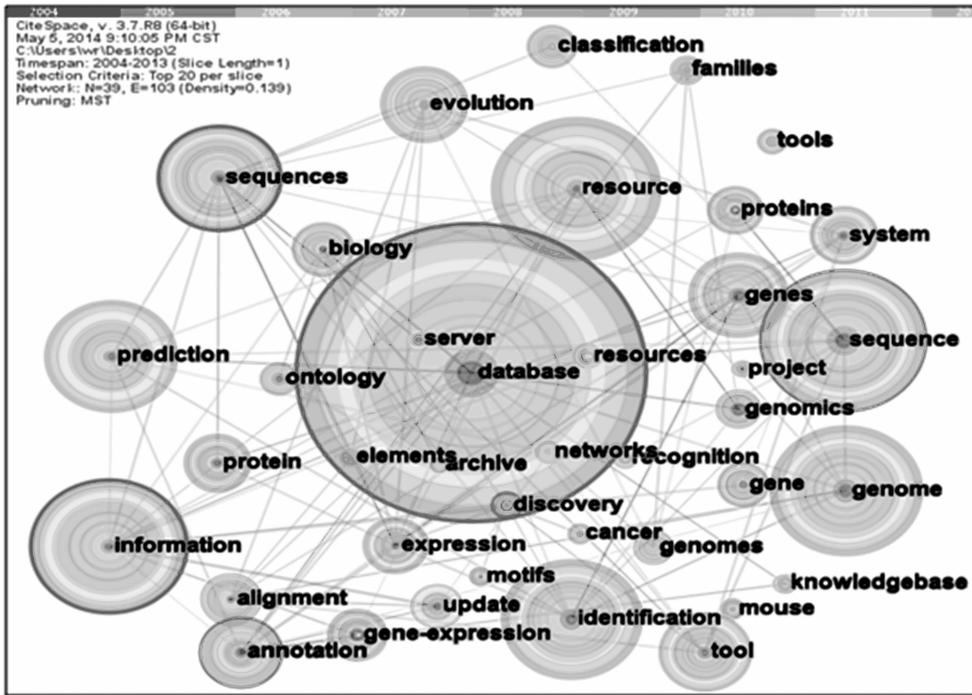


图5 生物信息学数据库研究热点分析

Fig.5 The hot-spot analysis of the bioinformatics database study

通过与上述相关研究的对比发现,本研究存在大部分相同或相似的研究热点,如序列库、基因组数据库、分类数据库、蛋白质数据库、数据库更新、集成系统等,这也在一定程度上说明了本研究的相对合理性。

图6是关键词时间线性图谱,展示生物信息学数据库近十年研究主题的演化过程。可知,在2004~2005

年该领域的研究主题集中在基因表达和序列数据库,2006~2008年倾向于蛋白质数据库的探析,2009~2010年集中在本体和结构域的研究,2011~2012年研究者主要探索基因、蛋白质等在疾病、药物设计中的应用,2013~2014年关注其分子间相互作用,从而探索其应用价值。

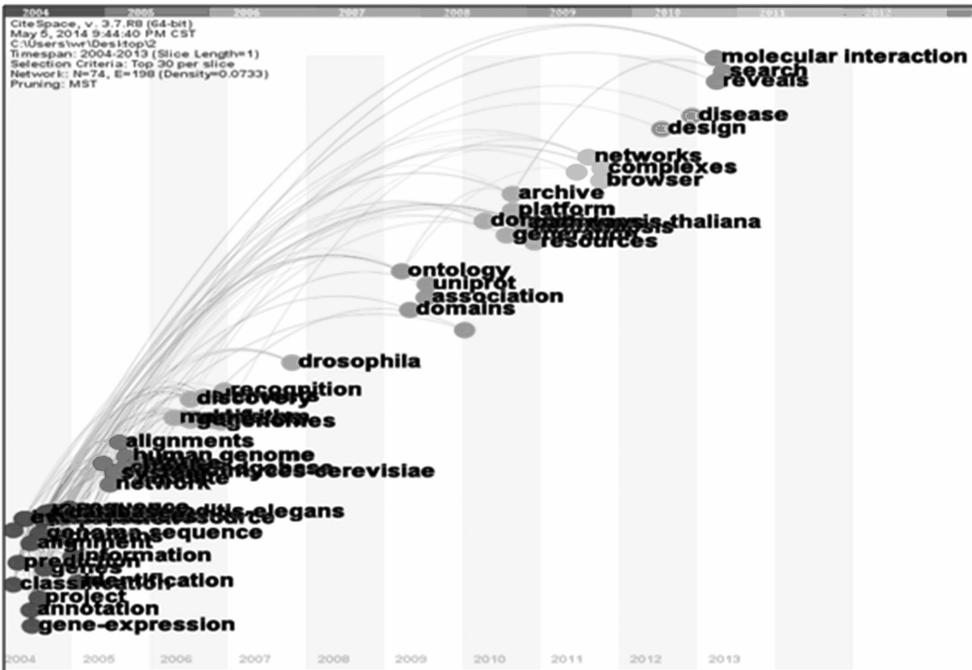


图6 生物信息学数据库研究前沿分析

Fig.6 The frontal analysis of the bioinformatics database study

由此可见,近几年生物信息学数据库研究逐渐走向关注其实际效益与意义的应用阶段,集中于生物信息学在疾病中的应用、肿瘤基因、分子间相互作用等,如王海波利用生物信息学数据库 NCBI 与 EMBL,获得植物细胞色素 C 家族的成员蛋白,绘制系统进化树及同源聚类分析^[24]。张文形等利用生物信息学工具,对现有的 H3N2 亚型人甲型流感病毒全球分离株的 H3A1 抗原序列进化规律进行分析研究^[25]。这说明,随着生物信息学的不断发展,其前沿扩展到人类疾病的预测和防治中,同时也验证了本研究的准确性。

3 结 论

(1)生物信息学数据库研究的总体趋势趋于稳定,呈小幅度的增长趋势,预测未来几年的生物信息学数据库研究的文献量将会保持稳定的状态。《Nucleic Acids Research》发表的新建数据库的数量和更新数据库的数量变化每年不定,没有形成规律。这说明数据库的研究仍处在发展阶段。

(2)生物信息学数据库研究早期重要奠基性文献是于1990年 Altschul SF 发表的 Basic local alignment search tool,首次提出快速对比序列的新方法——BLAST——局部比对搜索工具。被引频次最高的是1991年 KUHLTHAU CC 发表的新一代的蛋白质数据库搜索程序——Gapped BLAST 和 PSI-BLAST,同时也是中心性最高的文献。此外发现,“Gene ontology: tool for the unification of biology. The Gene Ontology Consortium”既是高被引文献,又是高中心性文献。这些文献是生物信息学数据库研究领域的核心知识来源。

(3)生物信息学数据库研究的热点集中在序列库、基因组数据库、分类数据库、蛋白质数据库、数据库更新、集成系统等。近几年生物信息学数据库研究逐渐走向关注其实际效益与意义的应用阶段,集中在生物信息学在疾病中的应用、肿瘤基因、分子间相互作用等。

参考文献(References)

- [1] 赵屹,谷瑞升,杜生明.生物信息学研究现状及发展趋势[J].医学信息学杂志,2012,2,33(5):2-6.
ZHAO Yi, GU Ruisheng, DU Shengming. The research status and development tendency of bioinformatics [J]. Journal of Medical Informatics, 2012, 2, 33(5): 2-6.
- [2] 孙清鹏,贾栋,万善霞.生物信息学应用教程[M].北京:中国林业出版社,2012.
- [3] SUN Qingpeng, JIA Dong, WAN Shanxia. Tutorial bioinformatics applications [M]. Beijing: China Forestry Publishing House, 2012.
- [3] 姜鑫.生物信息学数据库及其利用方法[J].现代情报,2005,25(6):185-187.
JIANG Xin. Bioinformatics databases and its using methods [J]. Modern Information, 2005, 25(6): 185-187.
- [4] 万跃华,何立民.网上生物信息学数据库资源[J].情报学报,2002,21(4):497-512.
WAN Yuehua, HE Limin. Bioinformatics Database Resources on Internet [J]. Journal of The China Society For Scientific and Technical Information, 2002, 21(4): 497-512.
- [5] 胡德华,张洁,方平,等.生物信息学数据库调查分析及其利用研究[J].生物信息学,2005,3(1):22-25.
HU Dehua, ZHANG Jie, FANG Ping, et al. Investigate and analyze the bioinformatics databases and their utilization [J]. China Journal of Bioinformatics, 2005, 3(1): 22-25.
- [6] 杨长平,吴登俊.Web of Science 收录生物信息学数据库研究文献的分析[J].农业图书情报学刊,2009,(1):65-67.
YANG Changping, WU Dengjun. Analysis of Bioinformatics Databases Research Literatures Based on "Web of Science" [J]. Journal of Library and Information Sciences in Agriculture, 2009, (1): 65-67.
- [7] WHITFIELD E, PRUESS M, APWEILER R. Bioinformatics database infrastructure for biotechnology research [J]. J Biotechnol, 2006, 124(4): 629-39.
- [8] MAVRIDIS L, NATH N, MITCHELL J. PFClust: a novel parameter free clustering algorithm [J]. BMC Bioinformatics, 2013, 3(14): 213.
- [9] WATTAM A, ABRAHAM D, DALAY O, et al. PATRIC, the bacterial bioinformatics database and analysis resource [J]. Nucleic Acids Res, 2014, 42 (Database issue): D581-91.
- [10] FINN R, TATE J, MISTRY J, et al. The Pfam protein families database [J]. Nucleic Acids Res, 2008, 36 (Database issue): D281-8.
- [11] FINN R, MISTRY J, TATE J, et al. The Pfam protein families database [J]. Nucleic Acids Res, 2010, 38 (Database issue): D211-22.
- [12] PUNTA M, COGILL P, EBERHARDT R, et al. The Pfam protein families database [J]. Nucleic Acids Res, 2012, 40 (Database issue): D290-301.
- [13] CHEN C, IBEKWE-SANJUAN F, HOU J. The structure and dynamics of co-citation clusters: a multiple-perspective co-citation analysis [J]. Journal of the American Society for Information Science and Technology, 2010, 61 (7): 1386-1409.
- [14] ALTSCHUL S, GISH W, MILLER W, et al. Basic local alignment search tool [J]. J Mol Biol, 1990, 215(3): 403-

- 410.
- [15] THOMPSON J, HIGGINS D, GIBSON T. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice [J]. *Nucleic Acids Res*, 1994, 22(22):4673-80.
- [16] MURZIN A, BRENNER S, HUBBARD T, et al. SCOP: a structural classification of proteins database for the investigation of sequences and structures [J]. *J Mol Biol*, 1995, 247(4):536-40.
- [17] ALSTCHUL S, MADDEN T, SCHAFFER A, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs [J]. *Nucleic Acids Res*, 1997, 25(17):3389-402.
- [18] ASHBURNER M, BALL C, BLAKE J, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium [J]. *Nat Genet*, 2000, 25(1):25-9.
- [19] BERMAN H, WESTBROOK J, FENG Z, et al. The Protein Data Bank [J]. *Nucleic Acids Res*, 2000, 28(1):235-42.
- [20] STEIN L, MUNGALL C, SHU S, et al. The generic genome browser: a building block for a model organism system database [J]. *Genome Res*, 2002, 12(10):1599-610.
- [21] BOECKMANN B, BAIROCH A, APWEILER R, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003 [J]. *Nucleic Acids Res*, 2003, 31(1):365-70.
- [22] 赵丽梅, 张庆普. 基于科学知识图谱的我国知识管理研究范式分析[J]. *情报学报*, 2012, 31(1):95-103.
ZHAO Limei, ZHANG Qingpu. Based on scientific knowledge map of knowledge management research paradigm analysis in our country [J]. *Journal of The China Society For Scientific and Technical Information*, 2012, 31(1):95-103.
- [23] 张晓东, 张传富, 彭科峰, 等. 生物信息学数据库研究进展[J]. *生物信息学*, 2006, 4(3):143-145.
ZHANG Xiaodong, ZHANG Chuanfu, PENG Kefeng, et al. Study advances of bioinformatics database [J]. *China Journal of Bioinformatics*, 2006, 4(3):143-145.
- [24] 王海波. 植物细胞色素 C 家族蛋白的同源聚类分析 [J]. *生物信息学*, 2010, 8(4):334-335, 340.
WANG Haibo. Homology analysis for cytochrome C family from different Plant [J]. *China Journal of Bioinformatics*, 2010, 8(4):334-335, 340.
- [25] 张文彤, 姜庆五. 全球历年人甲型流感病毒 H3A1 抗原的分子进化研究 [J]. *中华流行病学杂志*, 2005, 26(11):843-847.
ZHANG Wenxing, JIANG Qingwu. Phylogenetic analysis for H3A1 strain of all human influenza A virus [J]. *Chinese Journal of Epidemiology*, 2005, 26(11):843-847.