

我国园艺学领域研究热点与前沿可视化分析

孙会军

(中国农业大学 图书馆, 北京 100193)

摘 要:以 2003~2012 年我国 10 a 间的园艺学领域的 SCI 论文为对象,以信息分析领域中影响力较大的信息可视化软件 CiteSpace 为工具,采用科学知识图谱法,对我国园艺学领域的研究力量、研究热点和研究前沿进行可视化分析。结果表明:近 10 a 来我国园艺学的研究主要集中在中国科学院、华中农业大学、中国农业科学院、中国农业大学、南京农业大学、浙江大学;中国农业科学院与其他机构的合作最多;根据主题词出现的频次进行分析可知,分子标记、遗传学研究、数量性状位点、育种、抗氧化酶、表型变化研究、基因精细定位 7 个方面代表了园艺学领域的研究热点;以中间中心性为指标进行园艺学领域关键点的测度显示,图位基因克隆、简单重复序列、遗传距离、QTL 图谱定位、氨基酸类为园艺学领域的关键节点,是园艺学领域一个研究方向到另一个研究方向的转折点;通过探测突变专业术语(Burst terms),并进行共聚类分析可知,园艺学的研究前沿主要为数量性状位点、抗氧化酶(活性)以及植物再生等。

关键词:园艺;研究力量;研究热点;研究前沿;可视化;CiteSpace

中图分类号:S 6 **文献标识码:**A **文章编号:**1001-0009(2013)24-0215-05

信息可视化是常用的数据挖掘方法之一,它可以利用人类在可视化形式下对模型和结构的获取能力来解决科技文献数量过大、无法快速进行有效交流的问题。可视化数据挖掘可以观察、发现、筛选和理解信息,发现数据和信息背后所隐藏的含义^[1]。CiteSpace 是用来分析和可视化共引网络的 Java 应用程序,能够帮助分析知识领域中的新趋势^[2]。园艺是农业的一个重要组成部分,是指果树、蔬菜、花卉及观赏树木的栽培与繁育技术。园艺作物包含果树、蔬菜、观赏植物三大类作物群。中国是世界园艺大国,蔬菜、果树和花卉的种植面积均居世界第一^[3-4]。现以 CiteSpace 为工具,以 SCI 数据库中收录的我国园艺学科文献数据为样本,采用科学计量学的最新研究手段—科学知识图谱,对我国园艺学领域的研究力量、研究热点和研究前沿进行可视化分析,并推测其发展趋势。

1 数据来源

数据来源于美国《科学引文索引》(Science Citation Index,简称 SCI)数据库,SCI 数据库收录全球自然科学、工程技术等领域内 6 300 多种最具影响力的学术刊物。目前全球许多研究机构及研究型的大学都是该数据库

的用户,以此数据库为统计源进行调研,结果较具有说服力^[4]。该研究数据采集时间为 2013 年 4 月 8 日。数据年份限定为 2003~2012 年,共检索到 10 a 内我国园艺学科 SCI 论文总数为 2 731 篇。

2 研究方法

运用计量学的方法,通过美国 Drexel 大学陈超美团队开发的 CiteSpace 软件绘制科学图谱来展示我国园艺学研究热点及前沿趋势等。CiteSpace 是一款在科学文献中识别与可视化新趋势与新动态的 Java 应用程序,已成为信息分析领域中影响力较大的信息可视化软件^[5-6]。

3 园艺学科研究力量分布

图 1 为利用 CiteSpace 生成的机构图谱,网络节点选择为机构(Institution),时区为 2003~2012 年,时间跨度设置为 1,数据抽取对象为 Top 30,选择合适的阈值,共生成 101 个节点,63 条连线。图 1 中的圆形节点代表机构,圆形的大小代表该机构的发文频次,圆点越大代表发文频次越高,连线代表机构之间的合作,连线的粗细代表合作程度的大小,线越粗代表 2 个机构之间的合作越多。

由图 1 可知,园艺学科的研究主要集中在 6 个机构,分别是中国科学院(Chinese Acad Sci)、华中农业大学(Huazhong Agr Univ)、中国农业科学院(Chinese Acad Agr Sci)、中国农业大学(China Agr Univ)、南京农业大学(Nanjing Agr Univ)、浙江大学(Zhejiang Univ),除此 6 个机构以外,西北农林科技大学(NW A&F Univ)、华南

作者简介:孙会军(1976-),女,河北抚宁人,博士,馆员,现主要从事情报研究与信息咨询和科学计量分析等研究工作。E-mail: tsg003@cau.edu.cn.

收稿日期:2013-09-16

中农业大学、中国农业科学院、中国农业大学、南京农业大学、浙江大学、西北农林科技大学、华南农业大学与其他机构的合作较多,尤其是中国农业科学院,与其他机构的合作最多。

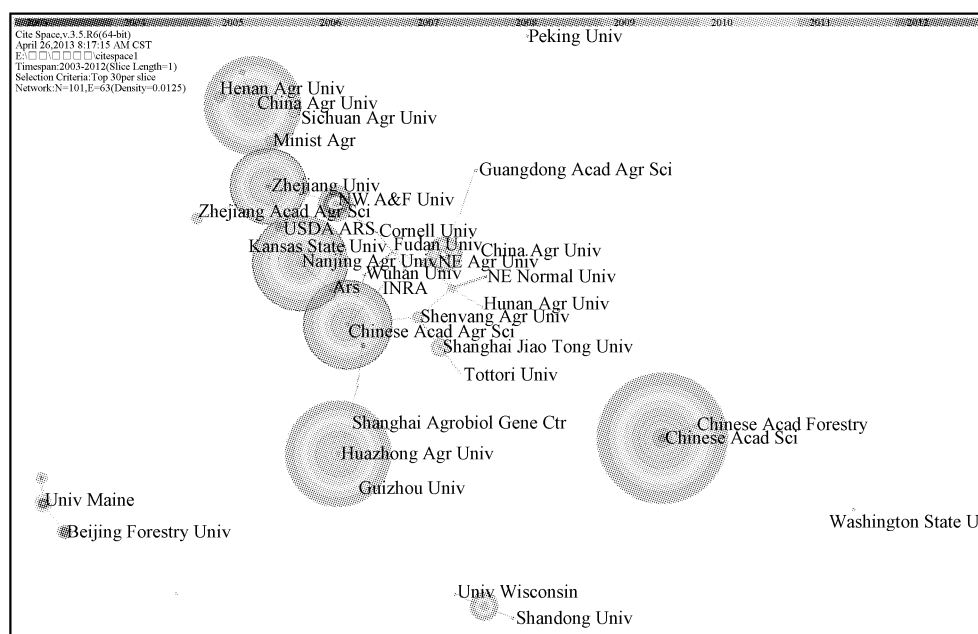


图1 园艺学研究力量分布知识图谱

4 园艺学领域研究热点及前沿分析

4.1 园艺学领域研究热点

研究热点是某一时期内,有内在联系的、数量相对较多的一组文献共同探讨的科学问题或专题^[1]。从文献计量学的角度看,在某学科领域内被引频次最高的研究型文献通常是该领域研究热点的集中体现。主题词

是一篇文章的核心和精髓所在,是对文章主题的高度概括和精炼,是规范化的语言。对文章的主题词进行分析,频次较高的主题词在一定程度上可以看作是该领域的研究热点^[7]。该研究将主题词(Terms)作为热点词汇来源,选择合适的阈值,运用 CiteSpace 对文献中的主题词进行可视化分析,通过显示高频主题词来确定园艺学

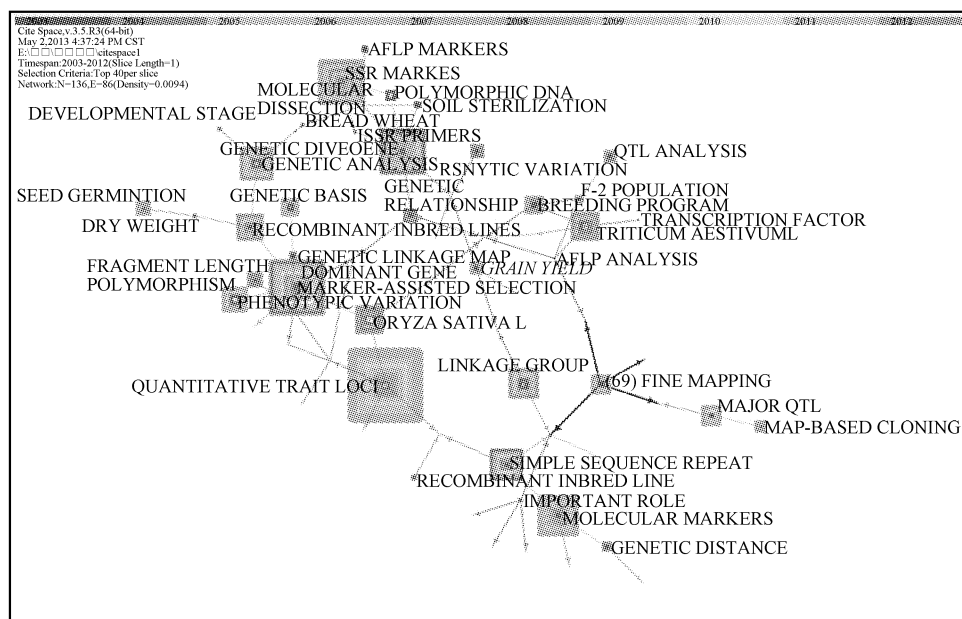


图2 园艺学领域研究热点知识图谱

领域的研究热点。

图2中的节点代表主题词,节点越大代表主题词出现的频次越高。从图2可以清晰的看到园艺学研究的热点领域,主要有数量性状位点(Quantitative trait loci)、标记辅助选择(Marker-assisted selection)、遗传多样性(Genetic diversity)、SSR标记(SSR markers)等。从Citespace导出的热点词信息中选取热点词中频次排名较高的词,可以得到该领域热点词的统计(表1)。

将表1中意思相同或相近的词进行合并,例如将

表1

园艺学热点词统计

热点词	频次	热点词	频次
quantitative trait loci(数量性状位点)	208	major QTL(主效 QTL)	70
marker-assisted selection(标记辅助选择)	166	genetic basis(遗传基础)	69
genetic diversity(遗传多样性)	147	fine mapping(精细定位)	69
SSR markers(SSR 标记)	153	breeding program(育种程序)	64
molecular markers(分子标记)	129	ascorbate peroxidase(抗坏血酸过氧化物酶)	64
genetic analysis(遗传分析)	109	fragment length polymorphism(片段长度多态性)	60
recombinant inbred lines(重组自交系)	101	genetic relationship(遗传关系)	60
linkage group(连锁群)	99	genetic variation(遗传变异)	57
superoxide dismutase(超氧化物歧化酶)	86	genetic distance(遗传距离)	55
phenotypic variation(表型变化)	81	QTL analysis(QTL 分析)	54

点的中心性是一个用以量化点在网络中地位重要性的图论概念^[2]。中间中心性是常用来进行中心性测度的指标^[8]。中间中心性高的点往往位于连接2个不同聚类的路径上,是不同研究领域或方向的连接点,从文献的角度来说,就是被不同的研究者共同引用次数较多的文献。CiteSpace 为了使用户很容易的看到具有较高中间中心性的节点,更容易的辨认关键点,在可

quantitative trait loci 与 major QTL 合并,genetic diversity 与 genetic analysis、genetic relationship、genetic basis、genetic variation 等合并,并参考原文献进行概括总结,主要有分子标记(包括发分子标记的开发、分子标记辅助选择)、遗传学研究(包括遗传基础、遗传多样性、遗传变异、遗传关系等)、数量性状位点、育种、抗氧化酶、表型变化研究和基因精细定位等7方面内容。这7个方面代表了园艺学领域的研究热点。

视化的网络中用紫色外圈来突出显示。从图3可以看出,目前园艺学研究领域的关键点主要有图位基因克隆(Map-based clone)、简单重复序列(Simple sequence repeat)、遗传距离(Genetic distance)、QTL图谱定位(QTL-mapping)、氨基酸类(Amino acids)。这些领域在园艺学发展过程中起着承上启下的作用,是一个研究方向到另一个研究方向转变过程中的关键点,即转折点。

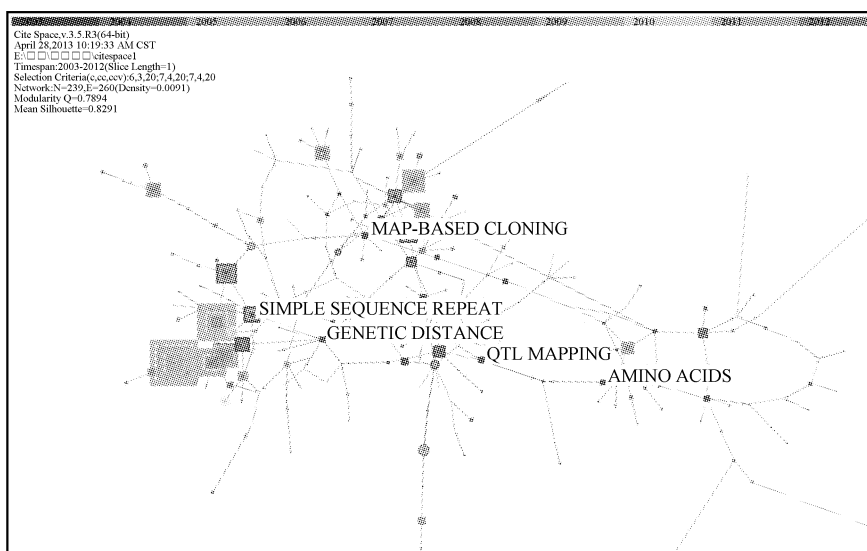


图3 园艺学领域关键节点知识图谱

4.2 园艺学领域研究前沿及发展趋势

研究前沿(Research front)指临时形成的研究课题以及其基础研究中问题的概念集合,也是正在骤变兴起或突现的理论研究趋势或新主题,是一个研究领域现状

的代表。研究前沿伴随着研究领域内新旧文献的交替而变动,体现着每个研究阶段中的最高研究水平^[9]。在CiteSpace中,研究前沿是基于从题目、摘要、系索词(Descriptors,指标引文献主题的单元词或词组)和文献

记录的标识符中提取出的突变专业术语(Burst terms)而确定的^[2]。CiteSpace通过考察词频变化(不单是统计频次的高低,而是检测词频的变动频次和变动趋势),将某段时间词频变化率最高的词从大量的主题中探测出来,根据这些突现词的共现聚类分析,得到“研究前沿术语的共现网络”。将数据导入到 CiteSpace,设置合适的阈值,运行 CiteSpace 提取突变专业术语,生成被引文献和突现词混合网络,共形成 28 个聚类,并通过时间线

(Timeline)方式显示,见图 4。

图 4 揭示了随时间形成的显著聚类,每个聚类对应一个研究方向。黑色的圈对应的突变专业术语具有很高的突增性,代表该领域的研究前沿,黑色的横线代表研究前沿知识基础跨越的年代。从图 4 可以看出,园艺学的研究前沿专业术语主要有数量性状位点(QTL)、抗氧化酶(活性)(Antioxidant enzyme(activity))以及植物再生(Plant regeneration)等。

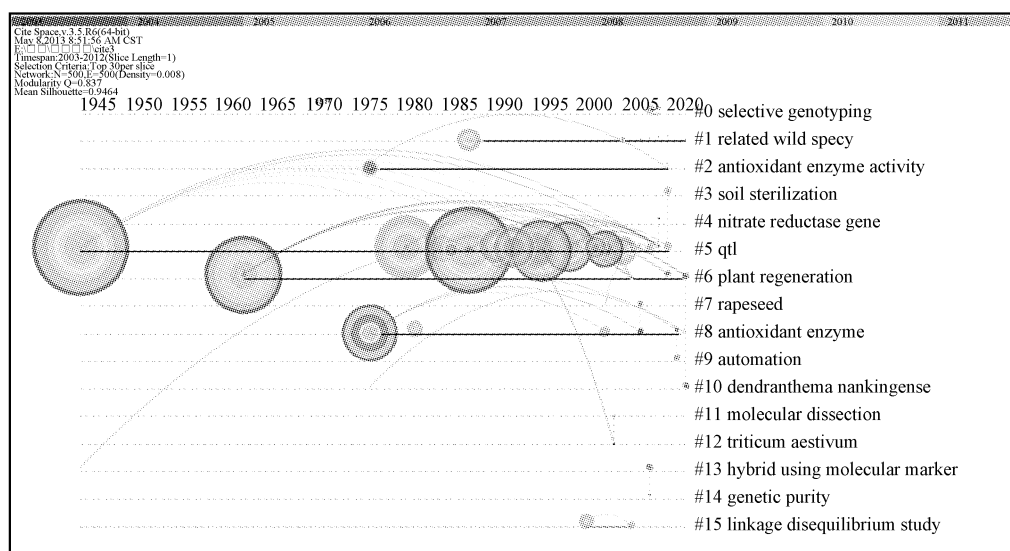


图 4 园艺学领域研究前沿术语知识图谱

将知识图谱以时区(Timezone)方式显示,由图 5 可以看出园艺学的发展趋势。从图 5 可以看出,近 3 a 的研究前沿术语主要有植物再生(Plant regeneration)、抗氧化酶(Antioxidant enzymes)和体细胞胚胎发生(Somatic embryogenesis)以及相关基因(Related-genes)或主效基因(Major genes)等,2007~2008 年的研究前沿为候

选基因(Candidate gene)、基因连锁图谱(Linkage map)等方面,2006 年较注重水果贮藏温度方面的研究,而 2005 年之前基本是遗传分析(Genetic analysis)、遗传相似性(Genetic similarity)和植物品种(Plant species)等方面的研究。通过阅读分析文献并结合图 4 可知,关于数量性状(QTL)的研究基本贯穿于近 10 a 当中。概括来说,近

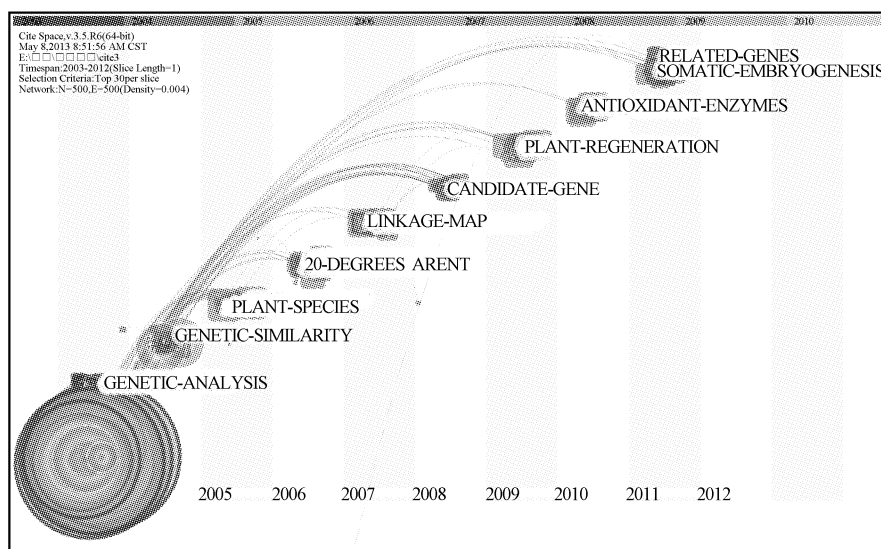


图 5 园艺学发展趋势知识图谱

10 a 来,园艺学的研究是从基础的遗传学方向不断向分子领域转变的过程。

5 结论与讨论

该研究结果表明,园艺学的研究主要集中在中国科学院、华中农业大学、中国农业科学院、中国农业大学、南京农业大学、浙江大学,除此 6 个机构以外,西北农林科技大学、华南农业大学、山东农业大学、北京林业大学的研究实力也较强,其中中国农业科学院与其他机构的合作最多;根据主题词出现的频次进行分析可知,分子标记、遗传学研究、数量性状位点、育种、抗氧化酶、表型变化研究、基因精细定位 7 个方面代表了园艺学领域的研究热点;以中间中心性为指标进行园艺学领域关键点的测度显示,图位基因克隆、简单重复序列、遗传距离、QTL 图谱定位、氨基酸类为园艺学领域的关键节点,是园艺学领域一个研究方向到另一个研究方向的转折点;通过探测突变专业术语(Burst terms),并进行共聚类分析可知,园艺学的研究前沿主要为数量性状位点、抗氧化酶(活性)以及植物再生等。通过时区(Timezone)方式显示突变专业术语聚类可以发现,近 10 a 来,园艺学的研究是从基础的遗传学方向不断向分子领域转变的过程。该研究的局限性在于由于学科的交叉性以及 SCI 数据库学科划分的局限性,园艺学 SCI 论文的检索结果

可能存在误检或漏检的情况,而且 SCI 数据库中的园艺学论文只是园艺学领域研究的表现形式之一,SCI 数据库中未收录的论文及其它形式的园艺学论文有待进一步分析,该研究结论仅能反映我国园艺学领域的部分研究热点、研究前沿及发展趋势。

参考文献

- [1] 徐灿,陈晨.基于 citespace 的学科领域研究热点与前沿可视化分析-以无线传感器网络领域为例[J].信息资源管理学报,2012(4):69-75,87.
- [2] 陈超美,陈悦,侯剑华,等. Citespace II:科学文献中新趋势与新动态的识别与可视化[J].情报学报,2009,28(3):401-421.
- [3] 霍尚一.中国水果出口贸易影响因素的实证分析[D].杭州:浙江大学,2008.
- [4] 孙会军.我国高校园艺学科 SCI 论文产出力与影响力比较研究[J].北方园艺,2012(20):194-197.
- [5] 肖明,陈嘉勇,李国俊.基于 citespace 研究科学知识图谱的可视化分析[J].图书情报工作,2011,55(6):91-95.
- [6] Chen C M. Citespace II: detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the American Society for Information Science and Technology, 2006,57(3):359-377.
- [7] 赵蓉英,许丽敏.文献计量学发展演进与研究前沿的知识图谱探析[J].中国图书馆学报,2010(5):60-68.
- [8] Freeman L C. Centrality in social networks: conceptual clarification[J]. Social Networks, 1979(1):215-239.
- [9] 王磊.国内图书馆学科服务现状可视化分析[J].图书情报工作,2013(2):136-142.

Visualization Analysis of Hot Research Topics and Fronts of Horticultural Science in China

SUN Hui-jun

(Library, China Agricultural University, Beijing 100193)

Abstract: Taking ten years' SCI papers of horticultural science published by researchers in China from 2003~2012 as objects, with influential information visualization software CiteSpace as tool, the research strength, research topics and fronts in horticulture were visual analyzed using scientific knowledge mapping method. The results showed that the horticultural research in nearly 10 years were mainly in Chinese Academy of Sciences, Huazhong Agricultural University, Chinese Academy of Agricultural Sciences, China Agricultural University, Nanjing Agricultural University and Zhejiang University. Among them, Chinese Academy of Agricultural Sciences had the most cooperation with other organizations. The research topics were molecular markers, genetic study, quantitative trait loci, breeding, antioxidant enzymes, phenotypic change, gene mapping according to the analysis of keywords frequency. The measure of horticultural science key points showed that figure-bit clone, simple sequence repeats, genetic distance, QTL mapping, amino acids were the key nodes, and it was a turning point from a research direction to another. By detecting mutations jargon and copolymerization analysis, the research fronts were quantitative trait loci, antioxidant enzyme and plant regeneration.

Key words: horticultural science; research strength; hot research topics; research fronts; visualization; CiteSpace