

## 基于 Scopus 的植物表型组学研究进展\*

唐惠燕<sup>1</sup>, 倪峰<sup>1</sup>, 李小涛<sup>3</sup>, 陈蓉蓉<sup>1</sup>, 袁曦临<sup>4</sup>, 朱艳<sup>2</sup>, 周济<sup>2,5,6\*</sup>, 丁艳锋<sup>2\*</sup>

(1. 南京农业大学图书馆 江苏 南京 210095; 2. 南京农业大学植物表型组学研究中心, 江苏 南京 210095; 3. 南京航空航天大学图书馆 江苏 南京 211106; 4. 东南大学图书馆 江苏 南京 210096; 5. Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK; 6. University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK)

**摘要:** 文献计量方法可以用于反映和预测科学技术发展的历史和趋势。本文基于科学大数据的计量方法探讨植物表型组学的研究现状, 为植物表型组学的发展提供参考。基于 Scopus 数据库中植物表型组学及其相关学术产出的文献数量、引用次数、合作单位、研究方向、学术机构和科研团队等信息, 利用 SciVal 和 CiteSpace5.0 等统计分析工具, 运用可视化数据方法, 分析植物表型组学研究领域的发展特点和趋势。研究显示, 植物表型组学已逐渐为许多研究团队所接受。作为一个新兴的研究方向, 学术产出以研究论文为主, 还包括图书、会议论文等。研究主题体现出农业、栽培、育种、生物学与计算机科学等多学科交叉发展的特性。高通量图像及相关数据分析是现阶段植物表型组学的重要研究方向, 主题显著度指数达到 98.8%, 受关注度极高。植物表型组学作为一个重要的组学研究方法, 目前主要研究对象包括拟南芥、水稻、小麦和玉米等。美国、德国、中国、英国和法国等国家的研究实力较强, 中国的主要研究机构为中国科学院、中国农业科学院以及一些高等院校。中国在植物表型组学的应用上已走在世界的前列。特别需要指出的是, 多学科合作提升了学术产出的水平, 尤其是国际间的跨学科合作。为推动植物表型组学领域的快速发展, 本文最后提出了增加植物表型组学相关的出版载体和不同学科间合作交流手段的展望。

**关键词:** 植物表型组学; 文献计量学; Scopus; 可视化; 主题显著度; CiteSpace  
**中图分类号:** G353.1; S501

### Advances in plant phenomics research based on Scopus analysis

#### tools

TANG Huiyan<sup>1</sup>, NI Feng<sup>1</sup>, LI Xiaotao<sup>3</sup>, CHEN Rongrong<sup>1</sup>, YUAN Xilin<sup>4</sup>, ZHU Yan<sup>2</sup>, ZHOU Ji<sup>2,5,6\*</sup>, DING Yanfeng<sup>2\*</sup>

1. Library of Nanjing Agricultural University, Nanjing 210095, China;
2. Plant Phenomics Research Center, Nanjing Agricultural University, Nanjing 210095, China;
3. Library of Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
4. Library of Southeast University, Nanjing 210096, China;
5. Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK;
6. University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK)

**Abstract:** Bibliometric analyses are capable of demonstrating the history and the

收稿日期: 2018-09-10

基金项目: “作物表型组学研究科学中心”项目 (YDZX20183200002058); 江苏省高校哲学社会科学研究重点项目 (ZDIXM018)

作者简介: 唐惠燕, 副研究馆员, 研究方向为学科分析、图书馆管理, E-mail: hytang@njau.edu.cn。\*通信作者: 丁艳锋, 教授, 研究方向为植物表型组学、作物栽培学, E-mail: dingyf@njau.edu.cn; 周济, 教授, 研究方向为植物表型组学、机器学习、图像分析、小麦育种, E-mail: [ji.zhou@njau.edu.cn](mailto:ji.zhou@njau.edu.cn), [ji.zhou@earlham.ac.uk](mailto:ji.zhou@earlham.ac.uk)。

tendency of scientific and technological development. This article utilizes big scientific datasets to explore the present status of plant phenomics, so that sound recommendations could be provided for this emerging research domain. Based on academic outputs such as publications, citations, collaborations, research areas, academic organizations, and authors retrieved from the Scopus database, statistical analysis tools such as SciVal and CiteSpace 5.0 were applied to visualize the development and tendency of plant phenomics related research, in a quantitative manner. Literature-based studies indicate that plant phenomics have been adopted by many research groups around the world. As an important field, academic outputs are mainly based on research articles, books and conference papers. The field requires interdisciplinary efforts to integrate agriculture, cultivation, breeding, and other plant research topics with computing sciences. Presently, high-throughput image analysis and related analytical solutions are important research themes in the field. The topical saliency index reaches 98.8%, a high relevance score. Arabidopsis, rice, wheat and corn are the main model crops. Research groups in the United States of America, Germany, China, Britain and France demonstrate strong research and development capabilities. From a methodological perspective, plant phenomics have been widely used by Chinese Academy of Sciences, Chinese Academy of Agricultural Sciences, and leading universities and academic organizations in China, which indicate that China is at the forefront of this emerging field. In particular, multidisciplinary and cross-country collaborations are required to raise the level of academic outputs. Finally, the prospect of increasing the publication channels and cross-disciplinary collaborations is put forward in order to promote the future development of this research domain.

**Keywords:** plant phenomics; bibliometrics; Scopus; visualization; topic prominence index; CiteSpace

表型组学是一门研究生物个体或群体,在特定条件下(如各类环境和生长阶段等)所表现出的可观察的形态特征及其变化规律的学科<sup>[1-2]</sup>。该研究领域近年来发展较快,特别是针对人、动物和植物单一或系列表型的研究已经受到学术界和业界的广泛关注<sup>[3-5]</sup>。目前,表型组学在作物育种和农业研究中的作用已获得普遍认可,植物表型组学正在成为植物基因组功能分析、分子育种和农业应用中的重要技术支撑<sup>[6-11]</sup>。

科技文献是科学研究成果的主要表现形式和载体。科技文献数量和结构变化以及相互引用、相互促进的规律反映了相应科技领域的发展特征<sup>[12]</sup>。因此,通过文献计量方法可以反映和预测科学技术发展的历史和趋势。目前,文献计量学的理论和方法已经被广泛应用于科研绩效评价、学科评估等领域<sup>[13-15]</sup>。

高水平学术成果的数量和质量是衡量该领域科学技术研究水准的重要指标。为了量化分析植物表型组学研究的现状,本文基于 Scopus 数据库对植物表型组学领域学术产出的文献数量、引用次数、合作单位、研究方向、学术机构和科研团队等信息进行大数据量化分析,通过可视化展示植物表型组学领域的研究现状,为植物表型组学的发展方向提供参考。

## 1 数据来源与计量评价指标

### 1.1 数据来源及分析工具

Scopus 是爱思唯尔出版社于 2004 年推出的文摘引文数据库<sup>[16]</sup>,收录全球 5000 余家出版社的 23 000 余种同行评议期刊、75 000 册图书以及 500 万篇会议论文,也

包括一些非英语国家和地区的文献资源。文献引文可追溯至 20 世纪 70 年代，并且每日更新。爱思唯尔出版社提供的 SciVal 研究分析工具可以从跨学科的角度揭示科研绩效的产出，通过共引聚类原理和可视化数据分析，量化比较全球 8 500 个研究机构和 220 个国家的 Scopus 数据库中近 6 年数据。

本文基于 Scopus 数据库检索植物表型组学相关的研究成果，主要包括学术论文、专著、专利等学术产出，利用 SciVal 平台重点分析 2013 年—2018 年 9 月植物表型组学应用及相关方法研究的论文及专著等学术产出类型，主要包括学术论文产出量、合作和影响力、重要研究机构和学者、热点主题等内容<sup>[17]</sup>。科研团队合作情况则利用 Scopus 数据输出功能，导入到 CiteSpace 5.0 工具进行分析<sup>[18]</sup>。由于专利成果的复杂性，针对专利的详细分析将作为本研究的后续研究。

与多名植物表型组学领域专家讨论，并经过多轮检索结果比对，最后确定检索用的关键词。检索用的一级关键词包括：phenomics, phenotyping, phenotype, image analysis；二级关键词包括：plants, crops，检索时间：2018 年 9 月。

## 1.2 计量评估指标

从学术角度分析植物表型组学的研究现状，所选用的主要评估指标见表 1。

表 1 主要评估指标及说明

Table 1 Main assessment indicators indices and explanations		
编号 Number	指标类型 Index type	说明及用途 Explanation and use
1	学术产出类型和数量	主要包括学术产出类型比例及不同类型的数量，如文献数量、专利数量。
2	学术产出引用情况	包括总被引数、单篇被引数、TOP1%高被引论文比例、TOP10%高被引论文比例等。
3	归一化文献引用次数	通过归一化消除不同学科差异性，用于评价学术产出质量情况，国际平均值为 1。
4	H 指数 H-index	用于评价学者发文量与引用情况的指标，包括 H 指数、H5 指数等。
5	合作情况	用于评价国家、机构、学者层面合作的类型，包括国际合作比例、国内合作比例、机构内合作比例等。
6	参与学者数	产出的学术成果中参与的学者数量。
7	主题显著性指数	用于评价主题被关注程度。爱思唯尔采用直接引用聚类方法将整个科学领域的主题一次识别，共得到 9.6 万个科学主题，每个主题都计算出主题显著性指数。
8	不同学科领域占比	用于评价领域的多学科性特征，Scopus 共包含 27 个学科领域和 334 个子领域。
9	引用分 citescore	期刊引用情况的指标，类似于期刊影响因子。

## 2 统计数据及分析

### 2.1 学术产出及分布情况

基于 Scopus 共检索到与植物表型组学研究和应用相关的文献共 51 618 篇，其中研究型期刊论文 44 950 篇，占 87.1%；综述型期刊论文 2 709 篇，占 5.2%；还包括会议论文、图书等，时间跨度为 1935 年—2018 年 9 月。学科范围涵盖了 Scopus 中 27 个研究领域中的 26 个领域，体现出植物表型组学广泛的应用领域和多学科交叉发展的特性，其中农业与生物科学领域 31 642 篇，占 61.3%；生物化学、遗传学和分子生物学领域 31 154 篇，占 60.3%；计算机科学领域 1 867 篇，占 3.6%。

本文根据不同时期相关文献的数量情况,把植物表型组学研究的发展大致分成4个阶段。第1阶段1935—1968年,共50篇文献,年发文量<10篇,年均1.5篇;第2阶段1969—1984年,共635篇文献,年发文量<100篇,年均39.7篇;第3阶段1985—2000年,共10 062篇文献,年发文量<1 000篇,年均628.9篇;第4阶段2001年以后,年发文量>1 000篇,2001年—2018年9月共发文36 538篇,年均2 029.9篇。本文对2013年—2018年9月植物表型组学学术产出情况进行了详细的统计分析,结果见表2。

表2 2013年—2018年9月植物表型组学相关学术产出情况

Table 2 Statistics of plant phenomics related academic outputs between 2013 and 2018-09

指标类型	2013	2014	2015	2016	2017	2018-09	总计 Total
文献数量 Number of publications	3 562	3 746	3 821	4 076	3 937	1 811	20 953
总被引数 Total citations	76 263	59 243	43 206	27 550	9 945	898	217 105
篇均被引数 Average citations per paper	21.4	15.8	11.3	6.8	2.5	0.5	10.4
归一化文献引用次数 field-weighted citation impact (FWCI)	1.5	1.49	1.42	1.46	1.33	1.43	1.44
TOP1%高被引论文比例 The proportion of TOP1% of highly cited publications (TOP1%HCP)	2.9	2.1	2.0	1.8	1.3	1.7	2.0
TOP10%高被引论文比例 The proportion of TOP 10% of highly cited publications (TOP10%HCP)	30.2	27.8	25.4	25.1	23.7	30.4	26.7
国际合作比例 The proportion of international collaborations	34.4	33.8	35.3	34.9	34.2	34.3	34.5
参与学者数 The number of involved scholars	17 881	19 286	20 002	21 795	21 499	9 889	80 330
专利数量 Number of patents	349	344	260	276	190	31	1 733

从表2可以看出:植物表型组学及相关研究经历了长时间的积累,近年来的文献数量仍呈不断上升趋势,参与的学者数量也在继续增加,目前已有超过2万名科研人员参与植物表型组学研究。从归一化文献引用次数(field-weighted citation impact, FWCI)可以看出论文整体质量较高,2013年—2018年9月发文的FWCI达到1.44,远高于世界平均水平1, TOP1%高被引论文比例2.0%, TOP10%高被引论文比例达到26.7%,这些数据均说明植物表型组学研究具有较高的科研水平。另外,近5年植物

表型组学专利申请量与发表文献数量相比并未呈现同步增长的趋势，可以看出目前植物表型组学研究还处于基础和理论研究阶段。与植物表型组学领域相关的专利申请数量通过 orbit 专利数据平台获取<sup>[19]</sup>。

表 3 列出了植物表型组学相关学术产出 TOP10 的国家和地区分布情况。从学术产出归属国家和地区分布来看，欧美和东亚发文量和引用率都处于世界前列。其中美国、中国、德国是文献数量和总被引数均排前三的国家，相比较而言，美国发文数量和总被引数均排世界第一，中国发文数量比美国略少，但总被引情况还有待提升。从 FWCI 看，部分西欧国家（特别是英、法、德三国）以及日本、澳大利亚等国发表了很多高影响力的学术成果。在中国，北京产出的植物表型组学及相关研究文献最多，南京、武汉、杭州等地也是重要的产出区。

表 3 2013—2017 年植物表型组学相关学术产出被引总量 TOP10 国家  
Table 3 Top 10 countries ranked by total citations of plant phenomics related academic outputs between 2013 and 2017

国家 country	发文献 Number of publications	被引总量 Total citations	篇均被引 Average citations per paper	归一化文献 引用次数 FWCI	TOP10%高被引 论文比例 The proportion of TOP 10% HCP
美国 U. S. A	5 106	78 420	15.36	1.82	35.1
中国 China	4 658	49 247	10.57	1.36	25.9
德国 Germany	1 873	28 840	15.40	2.00	38.7
英国 Britain	1 326	23 312	17.59	2.07	41.4
法国 France	1 386	20 099	14.50	1.88	35.2
日本 Japan	1 236	16 470	13.33	1.49	28.5
澳大利亚 Australia	971	15 466	15.93	2.00	37.6
西班牙 Spain	949	13 416	14.14	1.81	32.3
加拿大 Canada	782	11 216	14.34	1.70	29.8
荷兰 Netherlands	469	9 034	19.26	2.27	42.0

从表 4 可见：在被引总量 TOP10 机构名单中，美国和法国有 3 个机构入选，中国有 2 个机构入选。其中中国科学院共发文 835 篇，被引 12 740 次，均位列世界第一，中国农业科学院共发文 695 篇，被引 8 195 次，位列世界第五，2 个机构的篇均被引和 TOP10%高被引论文比例均相对较高，可以看出在植物表型组学领域的应用和研究，中国已经具备巨大的潜力，走在了世界的前列。从表 5 可见：德国克斯·普朗克分子植物生理研究所 Alisdair Robert Fernie 发表 58 篇植物细胞表型论文，被引 1 246 次，在被检索的所有学者中排名第一，其 H 指数和 H5 指数均非常高，是一位高产的研究工作者。另外，因为使用了表型组学的方法开展研究，中国农业科学院的万建民、郭秀萍和中国科学院的李家洋都在检索中进入 TOP10 学者名单。

表 4 2013—2017 年植物表型组学相关论文被引总量 TOP10 机构

Table 4 TOP 10 research organizations ranked by total citations of plant phenomics related academic outputs between 2013 and 2017

研究机构 research organizations	被引总数 Total citations	发文数 Number of publicati ons	归一化 文献引 用次数 FWCI	TOP10%高 被引论文比 例 The proportio n of TOP 10% HCP	国际合作比 例 The proportion of internatio nal collaborat ions
Chinese Academy of Sciences	13 137	835	1.75	36.8	39.8
INRA Institut National de La Recherche Agronomique	10 366	755	1.88	35.5	64.8
United States Department of Agriculture	9 943	778	1.70	30.7	49.4
CNRS	9 396	592	2.00	39.0	69.3
Chinese Academy of Agricultural Sciences	8 195	695	1.53	29.6	28.5
Cornell University	7 921	362	2.56	48.3	64.4
Spanish National Research Council	6 827	440	2.06	37.3	62.0
University of California at Davis	5 643	315	1.93	41.3	62.5
Universite Paris-Saclay	5 286	329	2.04	42.2	68.7
Wageningen University & Research	5 273	296	2.09	41.2	73.6

表 5 2013—2017 年植物表型组学相关学术产出较多的学者  
Table 5 Scholars with relatively high academic outputs in plant phenomics and related research between 2013 and 2017

学者姓名 Scholar name	归属机构 Affiliated organization(s)	被引总数 Total citations	发文数 Number of publications	归一化文献引用次数 FWCI	H5指数 H5 index	H指数 H index	TOP10%高被引论文比例 TOP 10% HCP
Alisdair Robert Femie	Max-Planck-Institut für molekulare Physiologie	1 246	58	2.37	32	97	58.6
Susan R. McCouch	Cornell University	1 230	24	5.19	18	74	83.3
Ed Buckler	Cornell University	1 189	33	3.46	23	66	75.8
Detlef Weigel	Max Planck Institute for Developmental Biology	999	25	5.09	28	101	76.0
WAN Jianmin	Chinese Academy of Agricultural Sciences	956	55	1.72	19	41	36.4
Li Jiayang	Institute of Genetics and Developmental Biology Chinese Academy of Sciences	869	16	6.24	22	51	100
Rajeev Kumar Varshney	International Crops Research Institute for the Semi-Arid Tropics	791	40	2.28	27	62	47.5
Gustavo De Los Campos	Michigan State University	789	12	3.98	13	28	75.0
Jean Luc Jannink	Cornell University	777	24	3.88	19	39	70.8
Guo Xiuping	Chinese Academy of Agricultural Sciences	745	31	2.11	17	27	48.4
Jiang Ling	Nanjing Agricultural University	721	32	1.96	16	32	40.6

## 2.2 学术产出合作情况

学术产出的合作包括国际合作、国内合作、机构内合作 3 种类型。2013—2018 年间植物表型组学研究这三种合作方式占 97.2%，只有 2.8% 属于无合作方式，由单个学者完成的学术成果，这说明合作是植物表型组学领域研究的主要方式。从图 1 可见：植物表型组学研究的国际合作比例远高于其他学科国际合作的平均比例，机构内合作比例却低于其他学科的平均比例。说明学术合作已逐渐成为全球植物表型组学领域的研究基础，而且国际间的合作也提升了该领域的学术产出水平（表 6）。

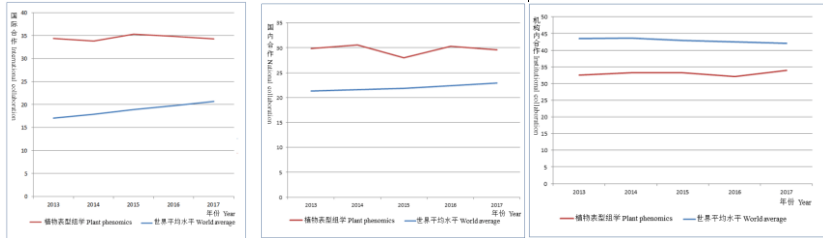


图1 植物表型组学研究的合作与学术产出国际合作平均的比较

Fig.1 The collaboration status between plant phenomics related research and other research domains

表6 2013—2017年植物表型组学相关研究论文合作情况

Table 6 Academic publications in plant phenomics related research based on collaborative research between 2013 and 2017

合作类型 Collaboration type	比例/% Ratio	发文数 Number of publications	总被引数 Total citations	篇均引用数 Average citations per paper	归一化文献引用次数 FWCI
国际合作	34.5	6 593	93 605	14.2	1.82
国内合作	29.6	5 651	58 839	10.4	1.35
机构内合作	33.1	6 320	52 762	8.3	1.14

学术产出的合作数据也可以反映不同科研团队的合作程度。图2是植物表型组学及其相关领域学术产出TOP20学者合作网络图，基于Scopus导出数据，利用CiteSpace 5.0工具绘制科研团队合作图谱，可视化展示学者学术产出情况以及相互间合作关系。图2中包括TOP20及其合作者共210个节点，因篇幅限制，只显示文献数量相对较多的节点。各节点的大小代表作者文献数量的多少，数量越多节点越大；节点之间的连线表示在文献中的共现关系，其粗细表明共现的强度，颜色则对应节点第1次共现的时间；图中颜色由冷色调到暖色调的变化对应发表年份由远及近。图2揭示植物表型组学及相关研究已经形成一定规模的合作网络。比如，以万建民、江玲和张欣等学者为核心的科研团队构成了图2中最大的合作子网，显示出中国农业科学院与南京农业大学密切的合作关系。国外的主要研究团队规模则相对较小，如A. R. Fernie、R. K. Varshney等学者发文量进入了该领域的TOP10，但跨团队协作规模不大。



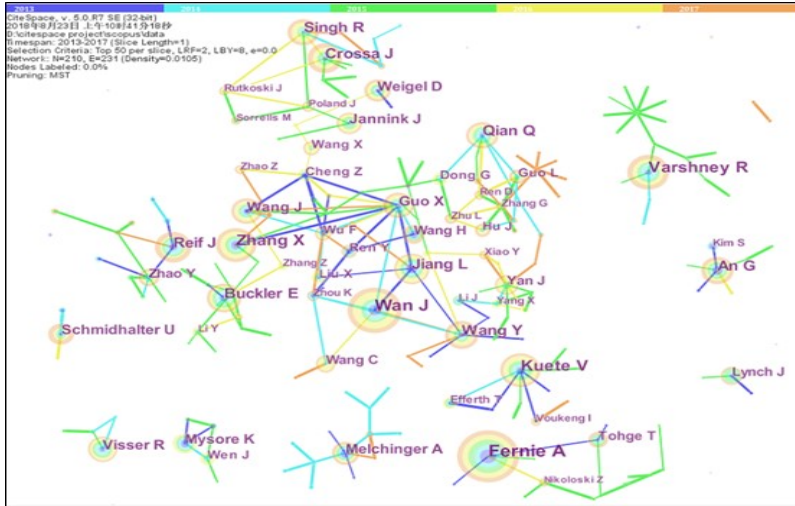


图2 2013—2017 植物表型组学及相关研究领域学术产出TOP20 学者合作图  
Fig.2 Collaborative mapping of TOP20 scholars in plant phenomics related research between 2013 and 2017

### 2.3 学术产出载体情况

虽然植物表型组学是一个多学科交叉发展的研究方向，但论文多发表在农业与生物学期刊上，发文比较分散。从表7可以看出：2013—2017年间《PLoS One》是植物表型组学及相关研究领域发文最多的期刊，共发表1242篇研究论文。而植物表型组学领域的高水平论文主要发表在《Plant Cell》《New Phytologist》和《Plant Physiology》等期刊上，TOP10%高被引论文比例分别达到68.7、59.6和54.2。

表7 2013—2017年植物表型组学及相关研究论文发表期刊分布

期刊名 Journal name	引用分 Cite score	发 文 数 Number of publications	篇均被引数 Average citations per paper	归一化文 献引用次 数FWCI	TOP10%高被引 论文比例 The proportion of TOP10%HCP
PLoS ONE	3.01	1242	10.82	1.25	20.7
Plant Physiology	6.01	666	17.02	2.14	54.2
Theoretical And Applied Genetics	4.42	504	12.09	1.66	29.8
Journal of Experimental Botany	5.78	492	14.36	1.70	36.8
Frontiers in Plant Science	4.14	468	8.24	1.61	23.7
Plant Journal	6.01	389	18.22	1.98	48.6
New Phytologist	6.65	354	18.08	2.67	59.6
Plant Cell	7.33	297	26.2	2.88	68.7
BMC Genomics	4.08	293	12.03	1.38	30.7

Scientific Reports	4.36	288	9.21	0.88	28.8
--------------------	------	-----	------	------	------

## 2.4 研究主题热点分析

### 2.4.1 研究热词图谱分析

利用 SciVal 提供的科学主题显著性研究模块对植物表型组学 2013 年—2018 年 9 月共发表的 20 953 篇论文进行关键词分析。首先使用爱思唯尔的“指纹技术”(FingerPrint)<sup>[20]</sup>提取文献题目和摘要里出现的 TOP50 关键词, 并加以归并和处理, 最后画出主题图谱(图 3)。图 3 中包括了 2013 年—2018 年 9 月所有学术产出的研究热词图谱, 以及 2013、2015、2017 年的研究热词图谱。其中, 绿色代表发文增加, 蓝色代表发文减少, 灰色代表基本持平, 字体越大代表与植物表型组学的关联性越强, 字体的位置是随机的。从图 4 可以看出植物表型组学研究的主题特征和变化趋势, 出现最多的关键词包括数量性状基因座

(Quantitative Trait Loci)、特征 (traits)、突变体 (mutants)、基因表达调控 (Gene Expression Regulation)、表现型 (phenotype)、干旱 (Droughts)、抗病性 (Disease Resistance)、图像分析 (image analysis) 等, 与作物品种相关联的词包括拟南芥 (Arabidopsis)、植物 (Plants)、水稻 (rice)、稻瘟病菌 (*Oryza sativa*)、小麦 (Triticum、wheat)、玉米 (Zea mays) 等。通过对 2013、2015 到 2017 年热词图谱进行对比, 可以看出植物表型组学研究的主题变化。2013 年主要以拟南芥为主, 研究拟南芥蛋白、数量性状基因座、基因表达等, 这些主题词呈现逐渐减少的趋势; 2015 年除了拟南芥, 又出现了水稻、小麦等作物表型的研究; 2017 年研究水稻表型的不断增多, 在研究手段上“图像分析”出现较多, 目前还处于不断增加的趋势。

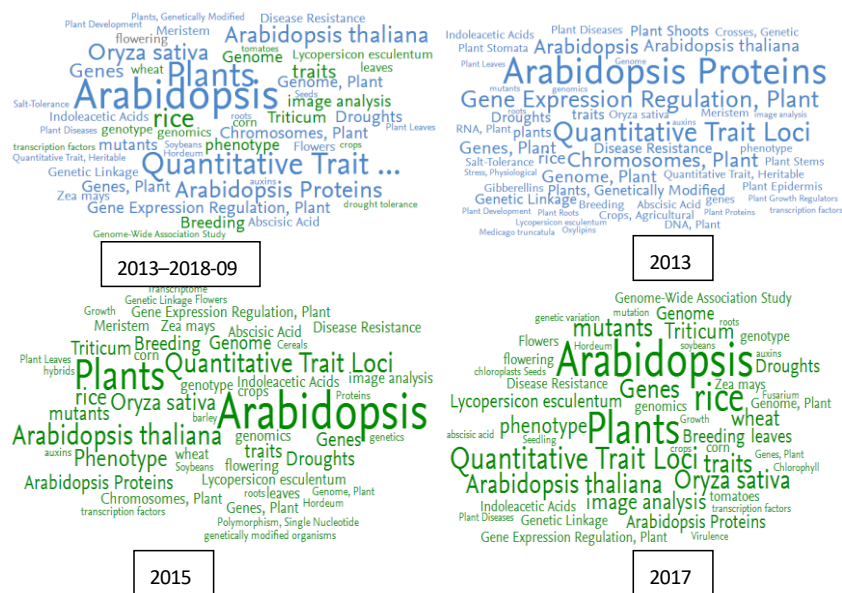


图 3 2013 年—2018 年 9 月植物表型组学及相关研究热词图谱比较

Fig. 3 Top keyword maps of plant phenomics and related research fields between 2013 and 2018-09

#### 2.4.2 主题显著度分析 基于 Scopus 数据库采用 Topic of Prominence 模型<sup>[21]</sup>

从全球范围内聚类成 96 000 个研究主题方向。笔者从中筛选与植物表型组学相关的研究主题，经过匹配只得到 1 个研究方向与植物表型组学相关，这个主题是“植物表型组学高通量图像和数据分析”，主题显著度指数达到 98.8%，发表在 TOP10%期刊上文献比例达到 52.8%，说明这个研究方向与表型组学研究有极强的关联性。基于图像和大数据的分析方法获取有实际生物学意义的信息是表型组学的研究重心<sup>[1]</sup>。笔者通过构建植物表型组学研究热词图谱，出现了很多与计算机科学领域相关的词，包括软件树冠方法学、计算机断层扫描、图像处理、计算机辅助等，体现出信息学科与农业学科交叉性的特点。

#### 2.5 植物类型分布

将与植物种类相关的检索词进行归并，筛选出相对较多的植物种类，与水稻相关主题包括 rice、*Oryza sativa* 等，与小麦相关的主题词包括 wheat、*Triticum* 等。从检索结果（图 4）看，目前植物表型组学研究的植物种类较少，主要包括拟南芥（*Arabidopsis*）、水稻（rice）、小麦（wheat）、玉米（*Zea mays*）、番茄（tomato）和大豆（Soybean）。拟南芥作为重要的模式植物，2014 年以前拟南芥的表型组学文献超过了其他植物种类的总和，此后呈现下降的趋势。水稻的表型组学文献数量占第 2 位，而且学术产出相对稳定。近几年针对小麦表型组学的研究文献也呈增加趋势。水稻和小麦表型组学研究整体学术产出总量每年不到 1 000 篇，而水稻和小麦每年总学术产出总量在 15 000 篇左右，可以看出目前植物表型组学研究还处于新兴阶段。

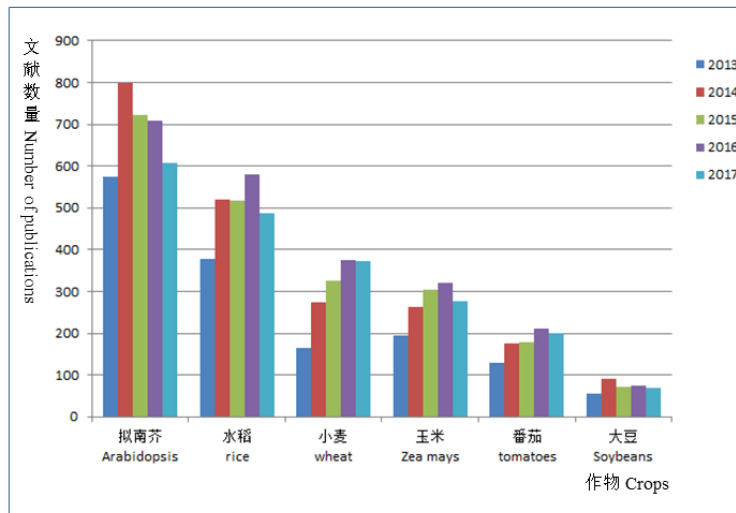


图 4 2013—2018 年 9 月植物表型组学不同植物学术产出量对比

Fig.4 Different plant model species studied in plant phenomics related publications between 2013 and 2018-09

### 3 讨论

表型信息获取与分析是植物功能基因组学和现代作物育种研究的基础。长期以来，对于表型的研究一直停留在传统人工获取阶段，劳动量大，效率低，对大批量植物样本的生长测量几乎是不可行的；同时表型数据的质量受人工主观因素的影响也比较大。植物表型的研究经历了相对较长的发展过程，本文的数据也说明了这个情况，从1935年到1984年植物表型研究领域以论文为主体的学术产出总量年均不足100篇。计算机和网络技术的快速发展为表型组学研究提供了新方法。近10年植物表型组学相关研究论文数量从2000篇增加到4000篇，逐渐成为研究热点，并且出现了各种专业化的表型测量平台<sup>[22]</sup>。植物表型的研究已由单纯的农业学科研究演变成由农业、自动控制、光学成像、图像分析及计算机技术等多领域多学科共同关注的研究方向，多领域的深度合作给作物育种学科的发展带来革命性的推动作用。

然而，本文针对植物表型组学领域主题词的聚类分析只获取到“高通量图像和数据分析”一个热点，显示出目前大量已发表的植物表型组学研究手段还相对比较单一。现阶段图像识别、深度机器学习、自动驾驶等人工智能技术已经深入影响并改变着人类的<sup>[23]</sup>生活方式。因此，植物表型组学的研究也不能仅仅依靠单一手段，而应该积极寻求与其他领域的合作，由多领域共同合作推动该学科的发展。从相关的研究综述可以看出，根据最新的表型组学中英文综述[1, 24]，植物表型组学正在向如何把海量图像和传感器数据转化为真正有意义的生物学知识的方向探索前进<sup>[1, 24]</sup>。在完善大规模、高通量表型数据采集环节的基础上，使用全新的大数据分析方法对表型数据集进行数据标注、标准化、动态存储和优化整合，通过引入特别是机器学习中的深度学习算法等人工智能方法的引入，在大数据中提取可靠的植物性状信息已成为植物表型组学今后的发展趋势。当然，由于表型分析需要开发团队具备数学、算法、硬件和生命科学等多方面的专业背景，因此，来自不同学科的各个科研团队之间学科交叉和的长期合作也是表型组学必不可少的发展基础之一。另外，基于大数据的植物表型组学研究不同于以往的实验室研究，在开放环境下如果不注意数据安全和管理将会出现数据质量和保存等方面的问题，因此，在研究过程中还需要加入数据领域专家共同参与合作。

现在不同国家和地区的研究团队在植物表型组学研究中的侧重点各不相同。比如，德国注重大规模可扩展的表型数据采集技术和基于自动化的数据分析方法；荷兰注重自动化室内植物表型筛分系统以及可控环境中的植物形态特征分析；法国在结合表型和作物遗传育种研究的基础上，偏重于室内外植物环境动态互作分析和生长预测建模；英国对遥感技术和基于机器学习的计算机视觉、图像分析的表型分析十分重视；日本则对小型、高精度的定性分析有较大的科研投入；而澳大利亚的表型研究主要针对各类田间作物的系列生长表型，并开发了一系列技术手段<sup>[25-28]</sup>。中国在植物表型组领域的研究应结合我国的实际情况，同时吸收其他国家的先进经验，重点突破关键性技术，从实验逐渐走向应用。

另外，基于大数据的植物表型组学研究不同于以往的实验室研究，在开放环境下如果不注意数据安全和管理将会出现数据质量和保存等方面的问题，因此，在研究过程中还需要加入数据领域专家共同参与合作。

近5年植物表型组学的研究发展较快，受关注程度较高，但出版渠道比较单一，大部分学术论文发表在与农学、生物学相关的期刊上，这种情况不利于调动其他领域专家参与的积极性。因此，增加植物表型组学研究相关的出版载体将加快该领域的发展，有利于植物表型组学研究的传播和交流。比如，2018年南京农业大学与Science合作创办英文期刊《Plant Phenomics》，这将有利于推动植物表型组学的快速发展。

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Superscript

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Superscript

植物表型组学研究是一个新兴的研究领域，以学术论文为主体的学术产出总量还不是很多。研究主题体现出农业、栽培、育种、生物学与计算机科学等多学科交叉发展的特性，高通量图像和数据分析是现阶段植物表型组学的重要研究方向，主题显著度指数达到 98.8%，受关注度较高。目前植物表型组学研究主要植物种类包括拟南芥、水稻、小麦和玉米等。

虽然植物表型组学研究已经呈现多学科合作发展的趋势，但植物科学还是主体领域。在世界范围内，美国、中国、德国等国家实力较强。中国科学院是学术产出数量最多也是引用总量最多的研究机构，同时中国农业科学院、中国农业大学、南京农业大学等 8 个机构占据植物表型组学学术产出 TOP20 的地位，体现出中国对植物表型组学研究和相关应用已处于世界前列。同时，植物表型组学及相关技术已逐渐为各类植物研究团队所接受，加大与其他领域的合作有利于植物表型组学的快速发展。

植物表型组学研究的学术产出以期刊文献为主，所有学术产出在几个主要相关学科中均处于较高水平。国际合作较多，且国际合作提升了学术产出的水平。大部分学术论文发表在与植物与生物学相关的期刊上，其中《Plant Cell》《New Phytologist》和《Plant Physiology》是发文较多且论文质量较高的期刊。

本文得到 Elsevier 公司靳辉先生的帮助，谨致谢意。

#### 参考文献 References:

- [1] 周济, Tardieu, F., Pridmore, T. 等. 植物表型组学: 发展、现状与挑战[J]. 南京农业大学学报, 2018, 41 (4): 580-588.
- Zhou, J., Tardieu, F., Pridmore, T., Doonan, J., et al. Plant phenomics: history, present status and challenges[J]. Journal of Nanjing Agricultural University, 2018, 41(4):580-588(in Chinese with English abstract).
- [2] Gowen CM, Fong S. Phenome analysis of microorganisms [M]. Edwards D, Stajich and D. Bioinformatics Tools and Applications, J Hansen, New York: editors. Springer, 2009.
- [3] Kelsoe JR, Niculescu AB 3rd. Finding genes for bipolar disorder in the functional genomics era: from convergent functional genomics to phenomics and back [J]. CNS Spectr, 2002, 7 (3): 215-216, 223-226.
- [4] Finkel E. With “Phenomics”, plant scientists hope to shift breeding into overdrive [J]. Science, 2009, 325: 380-381.
- [5] 李宏, 韦晓兰. 表型组学: 解析基因型-表型关系的科学[J]. 生物技术通报, 2013, (7): 41-47.
- Li H, Wei X L. Phenomics: A Science of Unravelling the Genotype-Phenotype Relationship[J]. Biotechnology Bulletin, 2013, (7): 41-47(in Chinese with English abstract).
- [6] Pieruschka R, Poorter H. Phenotyping plants: genes, phenes and machines Introduction [J]. Funct Plant Biol, 2012, 39: 813 - 820.
- [7] Tester M, Langridge P. Breeding technologies to increase crop production in a changing world [J]. Science, 2010, 327: 818 - 822.
- [8] Yang W, Duan L, Chen G, Xiong L, et al. Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies [J]. Curr Opin PlantBio, 2013, 16: 180 - 187.
- [9] Dhondt S, Wuyts N, Inze D. Cell to whole-plant phenotyping: the best is yet to come [J]. Trends Plant Sci, 2013, 18: 433 - 444.
- [10] White J W, Andrade-Sanchez P, Gore M A, et al. Field-based phenomics for plant genetics research [J]. Field Crops Res, 2012, 133: 101 - 112.

- [11]潘映红. 论植物表型组和植物表型组学的概念和范畴[J]. 作物学报, 2015, 41(2): 175-186.
- PAN Y H. Analysis of Concepts and Categories of Plant Phenome and Phenomics[J]. Acta Agronomica Sinica, 2015, 41(2): 175-186(in Chinese with English abstract).
- [12]邱均平. 信息计量学(二):第二讲文献信息增长规律与应用 [J]. 情报理论与实践, 2000, 23(2): 153-157.
- Qiu J P. Informetrics (二): Second Talk Regularity and Application of Document Information Growth[J]. Intelligence theory and Practice, 2000, 23(2): 153-157.
- [13]马春晖, 张南, 周晓丽等. 基于 Scopus 和 Scival 的国内外高校食品科学研究现状分析 [J]. 食品科学技术学报, 2016, 34(4): 54-60.
- Ma C H, Zhang N, Zhou X L, et al. Analysis of research status of food science basing on Scopus and Scival[J]. Journal of Food Science and Technology, 2016, 34(4):54- 60(in Chinese with English abstract).
- [14]马春晖, 戴岳, 张南. 基于文献计量学的高分子材料领域研究现状及热点分析[J]. 中国塑料, 2017, 31(6):8-15.
- Ma C H, Dai Y, Zhang N. Research Status and Hot Spots of Polymer Materials Based on Literature Metrology[J]. China Plastics, 2017, 31(6):8-15(in Chinese with English abstract).
- [15]孙颀. 基于文献计量的国内外柿研究现状分析[J]. 果树学报, 2017, 34(6): 706-714.
- SUN J. Current situation of persimmon science and technology in the world based on bibliometrics analysis[J]. Journal of Fruit Science, 2017, 34(6): 706-714(in Chinese with English abstract).
- [16]Scopus 官方网站[EB/OL]. [2018-09-26]. <http://www.scopus.com>.
- Scopus Official Website[EB/OL]. [2018-09-26]. <http://www.scopus.com>.
- [17]SciVal 官方网站[EB/OL]. [2018-09-26]. <http://www.scival.com>.
- SciVal Official Website[EB/OL]. [2018-09-26]. <http://www.scival.com>.
- [18]侯剑华, 胡志刚. CiteSpace 软件应用研究的回顾与展望 [J]. 现代情报, 2013, 33(4): 99-103.
- Hou J H, Hu Z G. Review on the Application of CiteSpace at Home and Abroad[J]. Journal of Modern Information, 2013, 33(4): 99-103(in Chinese with English abstract).
- [19]orbit 专利检索平台[EB/OL]. [2018-09-26]. <https://www.orbit.com>
- Orbit Patent Retrieval Platform[EB/OL]. [2018-09-26]. <https://www.orbit.com>
- [20]Henry Small, Kevin W. Boyack, Richard Klavans. Identifying emerging topics in science and technology[J]. Research Policy, 2014, 43:1450-1467.
- [21]Klavans R, Boyack K W. Research portfolio analysis and topic prominence[J]. Journal of Informetrics, 2017, 11:1158-1174.
- [22]段凌凤, 杨万能. 水稻表型组学研究概况和展望[J]. 生命科学, 2016, 28(10): 1129-1137.
- DUAN L F, YANG W N. Research advances and future scenarios of rice phenomics[J]. Chinese Bulletin of Life Sciences, 2016, 28(10): 1129-1137.
- [23]钟义信. 人工智能的突破与科学方法的创新[J]. 模式识别与人工智能, 2012, 25(3): 456-461.
- ZHONG Y X. Breakthroughs in Artificial Intelligence and Innovation in Methodology[J]. PR & AI, 2012, 25(3): 456-461.
- [24]Reynolds D, Baret F, Welcker C, Bostrom A, Ball J, Cellini F, Lorence A, Chawade A, Khafif M, Noshita K, Mueller-Linow M, Zhou J, Tardieu F. What is cost-efficient phenotyping? Optimizing costs for different scenarios. Plant Sci [Internet] 2018; In Press. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0168945217311482>
- [25]Furbank RT, Tester M. Phenomics—technologies to relieve the phenotyping bottleneck. Trends Plant Sci [Internet] 2011 [cited 2014 Mar 21];16:635-44. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22074787>
- [26]Fiorani F, Schurr U. Future scenarios for plant phenotyping. Annu Rev Plant

---

Biol [Internet] 2013;64:267–91. Available from:  
<http://www.ncbi.nlm.nih.gov/pubmed/23451789>

[27] Cendrero-Mateo MP, Müller O, Albrecht H, Burkart A, Gatzke S, Janssen B, Keller B, Körber N, Kraska T, Matsubara S, Li J, Müller-Linow M, Pieruschka R, Pinto F, Rischbeck P, Schickling A, Steier A, Watt M, Schurr U, Rascher U. Field Phenotyping: Challenges and Opportunities. *Terr Ecosyst Res Infrastructures* [Internet] 2017;53–80. Available from: <http://www.crcnetbase.com/doi/10.1201/9781315368252-4>

[28] Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M. Plant Phenomics, From Sensors to Knowledge. *Curr Biol* 2017;27:R770–R783.

Formatted: English (United Kingdom)