

Pan-Genome in the Research of Bamboology

Yun Zhang¹, Zhenguo Xu², Lifang Yang¹, Zhiyong Zhu¹, Xiaodong Chen¹,
Hong Ran^{1,3}, Ying Zhang^{1,4}, Qirong Guo^{1*}

¹SFA Key Laboratory of Bamboo and Rattan Science and Technology, International Center for Bamboo and Rattan, Beijing

²Guangxi Academy of Forestry, Nanning Guangxi

³Chongqing Academy of Forestry, Chongqing

⁴Guangdong Academy of Forestry, Guangzhou Guangdong

Email: 13681062477@163.com, `QRGUO@126.com

Received: Nov. 28th, 2016; accepted: Dec. 11th, 2016; published: Dec. 14th, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Pan-genome was firstly proposed by American scientists Herve Tettel *et al.* in 2005 in bacteria. Pan-genome includes core genome, specific genome and dispensable genome. In this paper, based on retrospection of the concept of pan-genome, development and evolution process, we pay attention to the research process in biological field significantly, and get main research achievements, especially in systematic significations, biological diversity, gene-explore, adaptability and resistance. Because of the particularity of flowering only once in bamboo's lifetime, the traditional classification method which contains flower and fruit has great restrictions. According to the problems of phenotype, physiology and ecology in the study, we offer a proposal in bamboology; pan-genome should be strengthened to resolve some difficult problems of bamboo.

Keywords

Pan-Genome, Core Genome, Specific Genome, Dispensable Genome, Bamboology

竹学研究中的泛基因组

张 韞¹, 徐振国², 杨丽芳¹, 朱志勇¹, 陈晓东¹, 冉 洪^{1,3}, 张 莹^{1,4}, 郭起荣^{1*}

¹国际竹藤中心, 国家林业局竹藤科学与技术重点开放实验室, 北京

²广西林业科学研究院, 广西 南宁

*通讯作者。

文章引用: 张韞, 徐振国, 杨丽芳, 朱志勇, 陈晓东, 冉洪, 张莹, 郭起荣. 竹学研究中的泛基因组[J]. 农业科学, 2016, 6(6): 187-193. <http://dx.doi.org/10.12677/hjas.2016.66029>

³重庆市林业科学研究院, 重庆

⁴广东林业科学研究院, 广东 广州

Email: 13681062477@163.com, *QRGUO@126.com

收稿日期: 2016年11月28日; 录用日期: 2016年12月11日; 发布日期: 2016年12月14日

摘要

最早在细菌中, 由美国科学家Herve Tettelin等, 于2005年提出了泛基因组(Pan-genome)概念。泛基因组展示由核心基因组、特有基因组、非必须基因组3部分组成。本文在回溯泛基因组概念的提出、发展、衍变过程的基础上, 重点关注了其在生物领域的主要研究历程, 特别是在系统演化、生物多样性、基因挖掘、适应性与抗逆性等方面的主要研究成果。由于竹类是多年生一次开花的独特植物, 故依据传统的以花、果等器官进行植物分类存在很大的局限性。根据我们在研究竹类植物的表型、生理、生态等方面的多样性遇到的问题, 提出竹类植物需要强化泛基因组研究, 以破解更多竹学难题。

关键词

泛基因组, 核心基因组, 特有基因组, 非必需基因组, 竹学

1. 引言

泛基因组(Pan-genome)现指利用高通量测序和生物信息分析手段, 对不同种质材料进行较高深度测序、单独组装, 以构建泛基因组图谱, 最终获得该物种完整遗传信息的一种技术手段。文献调研发现, 其最早出现在细菌研究中, 现在在真菌、藻类、拟南芥, 海洋浮游植物, 农作物如大豆, 林木如杨树等都有了重要进展[1] [2] [3] [4]。

随着越来越多的细菌基因组测序的完成, 研究人员发现一些细菌株系的基因组至少有 20%的 DNA 序列为该株系所特有, 同时也不存在于同种的其他株系之中, 同一细菌不同株系间存在着明显的基因差异[5]。直到 2005 年, 美国基因组研究所(The Institute for Genome Research, TIGR)的科学家 Herve Tettelin 等人从 8 株 B 群链球菌(GBS)的独立基因组序列中, 发现菌株间有 1806 个基因共有, 而链球菌基因组有 2000~2400 个基因, 也就是说还有一定数量的基因在某个或某些菌株内是缺失的。为了更加准确的描述细菌基因组的信息, Herve Tettelin 等人首次提出了泛基因组(Pan-genome)的概念, 指同种细菌不同菌株全部遗传信息的集合[1]。

细菌通过相应的分子机制对自身基因筛选, 以适应外界环境条件, 推动自身进化。许多功能基因会在利于细菌的前提下发生水平转移, 同时, 也为防止基因组容量的大幅增长而进行基因缺失。基因的水平转移与基因缺失是基因组进化的重要机制。缺失的基因对细菌可能是有害的、中性的、有利的, 总的来说对细菌生长繁殖是非必需的[6]。

2. 泛基因组研究的重要成果

最新研究成果认为, 泛基因组可以区分为核心基因组(Core genome)和非必须基因组(dispensable genome)2 大类(图 1):

核心基因组是指在同一种细菌所有菌株中都存在的基因。它是处于高选择压力下的基因, 代表某一物种基本表型特征。核心基因保守性较强, 有研究认为更适合作为疫苗制备的候选基因[7], 在疫苗设计

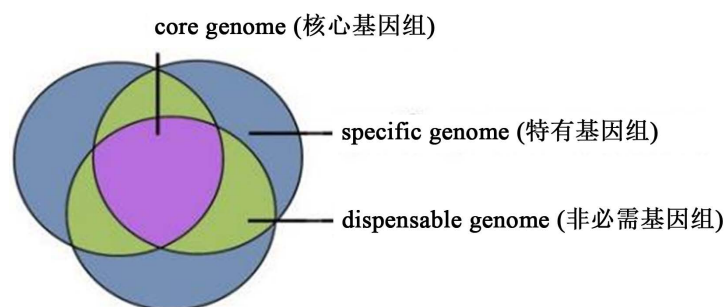


Figure 1. The composition of pan-genome
图 1. 泛基因组构成

中具有广阔应用前景[8], 泛基因组的反向疫苗学(Reverse vaccinology)成果呈现[9]。链球菌适应性发生改变的重要机制是其核心基因组遭受到正选择, 彰显其致病机制研究成果[10]。

非必需基因组包括部分菌株中存在的基因和菌株的特有基因(specific genome)。在细菌整个进化史上, 为适应环境条件而不断发生水平转移与基因缺失, 属于非必需基因。非必需基因组使不同菌株间在表型、生活方式、代谢等方面发生明显的分化, 最终表现为丰富的遗传多样性。

泛基因组学研究中最重要工作之一就是判断某一段序列在该物种基因组中的分布宽度, 即这一段序列在多少菌株中存在。基于多菌株基因组的测序方法, 主要通过“序列比对”来判断两个菌株甚至多个菌株序列之间的同源关系, 包括直系同源基因鉴定、基因组同源比对。

2.1. 生物

细菌菌株鉴定多采用的 MLST (Multilocus sequence typing)方法, 选用多个基因的策略来进行分析, 存在很大的局限性。泛基因组的分析方法可从选择所有单拷贝的核心基因用于演化分析, 或基于基因得失的进化事件, 计算不同菌株之间非核心基因比例以及差异, 作为演化分析矩阵。泛基因组在分类与鉴定中, 方法更为灵敏[11]。

在植物界, 2011年, 见德国科学家对来自8个区域的80株拟南芥 *Arabidopsis thaliana* 进行测序, 研究了拟南芥泛基因组中小的核酸多态性、大片段的插入和删除, 并对基因功能、变异特征进行局部、全局的关联分析[12]。

Emiliania huxleyi 是一种海洋浮游植物, 科研人员对来自互相隔离的不同地区的14个植株进行了基因组测序并对其进行了序列比对。采用泛基因组分析表明, 其基因组由核心基因和不同种系存在变异基因组成, 使得其适应广泛的气候环境, 从赤道到北极的全球范围内都有广泛分布[13]。

中国农科院与北京诺禾致源生物信息科技有限公司等机构, 2014年, 率先报道了野生大豆(*Glycine soja*)的泛基因组研究成果, 对7个从头组装(*De novo assembly*)的野生大豆基因组进行比较, 发现共有59,080个基因家族, 其中48.6%为核心基因, 51.4%为非必需基因。发现与野生大豆生物逆境相关的R(抗病)基因类型远高于栽培大豆。此外, 还发现一系列重要的开花基因(如 *PHYA*, *FT* 和 *LEAFY* 等)、与脂肪含量相关的三酰甘油(TAG)合成、蔓生性及器官大小(产量相关)等相关候选基因在野生大豆和栽培大豆之间发生了重要的遗传变异, 为大豆品质育种、光周期、产量等相关基础研究与应用提供了依据[3]。有资料显示, 诺禾致源还与国际热带农业研究中心(CIAT)进行着木薯(*Manihot esculenta*)泛基因组研究。等等。

在林木中, 2016年最新首现杨树泛基因组研究成果。首次关注了杨树泛基因组的结构与功能, 研究人员对3个杂交杨树品种基因组的结构变异(structural variation, SV)、插入缺失(InDels)、拷贝数(copy number variation, CNV)进行了全基因组研究, 杂交变异后的基因种类丰富, 主要与其抗性相关[14]。

此外, 基于人类基因组中存在着种群特异、甚至个体独有的 DNA 序列和功能基因, 中国科学家在 2009 年率先提出了“人类泛基因组”的概念, 着手研究人类基因序列的总和, 成果不断涌现。

2.2. 方法学

泛基因组的研究分析工具不断完善。中科院北京基因所基于 Perl 语言开发了一套泛基因组自动化分析工具 PGAP (Pan-Genomes Analysis Pipeline), 基于蛋白编码基因数据对多株细菌进行泛基因组自动化分析, 包括功能基因聚类、泛基因组特征、功能基因遗传变异、物种演化、基因簇功能富集等, 用户可以通过一条命令完成这些工作[15]。2016 年, 荷兰瓦格宁根大学生物信息学组 Siavash Sheikhzadeh 等人, 开发了一种泛基因组工具(PanTools), 适用于高等真核生物, 其由代表泛基因组的 *De Bruijn* 图和在线构建泛基因组的计算法则两部分组成, 可用于泛基因组注释、序列添加、基因分类、基因序列或基因组区域检索、基因组重建、泛基因组的比较与查询等, 用来综合表征多种注释的基因组, 便于分析基因组成分的组成与结构[16]。

3. 泛基因组在竹学中的运用

竹子多是多年生一次性开花的独特植物, 其笋期快速生长, 常见于热带、亚热带地区, 但北可至萨哈林群岛, 在喜马拉雅可分布到雪线(4200 m)。因为竹子的形态性状复杂多变, 常缺少花、果实等繁殖器官的关键性状, 竹子不但对竹族、亚族的划分存在巨大分歧, 而且属的界定存在很大争议。虽然以李德铎研究团队为代表的队伍等在经典分类基础上, 卓越运用叶绿体、核基因和简化基因组、转录组、生命条形码等数据, 创新竹类系统发育研究领域, 取得了优异成果, 但竹子仍然因为杂交、渐渗、非协同进化、拷贝数变异、基因水平转移、化石证据不足等原因, 导致得到的基因树冲突[17]-[24]。采用泛基因组的研究思路和方法, 是解决竹类系统发育与演化科学难题的途径之一。

籐竹属 - 牡竹属 - 巨竹属是一个内部系统发育极其复杂的复合群, 已有研究揭示其为杂交起源类群[25]的复杂关系, 建议进一步采用泛基因组手段, 仔细探求之。

2013 年, 毛竹(*Phyllostachys edulis*)全基因组草图的公布, 阐明: 在数万年前, 毛竹从禾本科中分化形成, 在进化的过程中出现过染色体加倍成为四倍体, 继续进化成现在的二倍体植物[26]。毛竹原产中国, 24°N~32°N 间生长着世界上绝对数量的资源, 其垂直分布从海拔几米到 1000 m 左右, 个别地区可达 1400 m 左右的高度, 甚至达到 1600 m [27] (图 2)。毛竹先后引种至日本、韩国、菲律宾、越南及欧美各国。在日本, 1736 年从中国长沙引进的毛竹, 已经长到了北海道, 直线距离超过 2800 km (图 3) [28]。从表型看, 毛竹已经发表了 22 个种下类型[29]。如此广袤分布的毛竹, 虽然呈现强单克隆优势[28], 但也表现了良好的地理适应性, 和早竹[30] [31]、桂竹[32]、麻竹[33]、慈竹[34]等一样, 初步发现其具有多元的遗传格局。众所周知, 龙头竹(*Bambusa vulgaris*)是世界分布面积最广大的竹子, 世称“泛热带竹种”; 神农箭竹(*Fargesia murielae*)已经栽进了北极圈……从泛基因组的角度, 分析这些竹类的强适应性、抗逆性机制, 同样具有重要的科学、产业和生态意义。

厚壁毛竹(*Phyllostachys edulis* “Pachyloen”)自发现以来, 因其竹壁厚, 具有重要经济性状, 已从生物学、解剖学、繁育学、引种学、生理生化、种质资源学、材料学、分子生物学、多种组学等不同视角, 进行了系列研究[35]-[43]。同样, 野外调查发现, 胸径可达 30cm 以上, 秆高 30 多米高的巨龙竹(*Dendrocalamus sinicus*)也有“厚壁”、“薄壁”类型[44] [45]。此外, 箭竹类(*Fargesia* spp.)、泰竹类(*Thyrsostachys* spp.)多有厚壁性状。竹子也有竹壁极薄的种类, 如泡竹(*Pseudostachyum polymorphum*)、薄竹(*Cephalostachyum chinense*)、沙罗单竹(*Schizostachyum funghomii*), 多种慈竹(*Schizostachyum* spp.)等等, 泛基因组可以探索竹子具“中空结构”的竹壁的形成机制。其它诸如花秆、花叶、花斑、条纹等众多的特异资源, 具有泛基因组研究的着力点。

基金项目

科技部科技支撑项目 2013BAD01B06-3。

参考文献 (References)

- [1] Tettelin, H., Masignani, V., Cieslewicz, M.J., *et al.* (2005) Genome Analysis of Multiple Pathogenic Isolates of *Streptococcus agalactiae*: Implications for the Microbial "Pan-Genome". *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 13950-13955. <https://doi.org/10.1073/pnas.0506758102>
- [2] Li, R.Q., Li, Y.R., Zheng, H.C., Luo, R.B., Zhu, H.M., Li, Q.B., Qian, W.B., Ren, Y.Y., Tian, G., Li, J.X., Zhou, G.Y., Zhu, X., Wu, H.L., Qin, J.N., Jin, X., Li, D.F., Cao, H.Z., Hu, X.D., Blanche, H., Cann, H., Zhang, X.Q., Li, S.G., Bolund, L., Kristiansen, K., Yang, H.M., *et al.* (2010) Building the Sequence Map of the Human Pan-Genome. *Nature Biotechnology*, **28**, 57-63. <https://doi.org/10.1038/nbt.1596>
- [3] Li, Y.H., Zhou, G.Y., Ma, J.X., *et al.* (2014) *De Novo* Assembly of Soybean Wild Relatives for Pan-Genome Analysis of Diversity and Agronomic Traits. *Nature Biotechnology*, **32**, 1045-1052. <https://doi.org/10.1038/nbt.2979>
- [4] Pinosio, S., Giacomello, S., Faivre-Rampant, P., *et al.* (2016) Characterization of the Poplar Pan-Genome by Genome-Wide Identification of Structural Variation. *Molecular Biology and Evolution*, **2016**, msw161. <https://doi.org/10.1093/molbev/msw161>
- [5] 周冬升, 杨瑞馥. 细菌比较基因组学和进化基因组学[J]. 微生物学杂志, 2003, 23(5): 31-43.
- [6] 庄绪冉, 朱冰璋. 细菌泛基因组学研究[J]. 上海交通大学学报(医学版), 2012, 32(1): 1440-1443.
- [7] Medini, D., Donati, C., Tettelin, H., *et al.* (2005) The Microbial Pan-Genome. *Current Opinion in Genetics & Development*, **15**, 589-594. <https://doi.org/10.1016/j.gde.2005.09.006>
- [8] Pallen, M.J. and Wren, B.W. (2007) Bacterial Pathogenomics. *Nature*, **449**, 835-842. <https://doi.org/10.1038/nature06248>
- [9] Bambini, S. and Rappuoli, R. (2009) The Use of Genomics in Microbial Vaccine Development. *Drug Discovery Today*, **14**, 252-260. <https://doi.org/10.1016/j.drudis.2008.12.007>
- [10] Lefebure, T. and Stanhope, M.J. (2007) Evolution of the Core and Pan-Genome of *Streptococcus*: Positive Selection, Recombination, and Genome Composition. *Genome Biology*, **8**, R71. <https://doi.org/10.1186/gb-2007-8-5-r71>
- [11] Hall, B.G., Ehrlich, G.D. and Hu, F.Z. (2010) Pan-Genome Analysis Provides Much Higher Strain Typing Resolution than Multi-Locus Sequence Typing. *Microbiology*, **156**, 1060-1068. <https://doi.org/10.1099/mic.0.035188-0>
- [12] Cao, J., Schneeberger, K., Ossowski, S., *et al.* (2011) Whole-Genome Sequencing of Multiple *Arabidopsis thaliana* Populations. *Nature Genetics*, **43**, 956-963. <https://doi.org/10.1038/ng.911>
- [13] Read, B.A., Kegel, J., Klute, M.J., *et al.* (2013) Pan Genome of the Phytoplankton *Emiliania* Underpins Its Global Distribution. *Nature*, **499**, 209-213. <https://doi.org/10.1038/nature12221>
- [14] Pinosio, S., Giacomello, S., Faivre-Rampant, P., *et al.* (2016) Characterization of the Poplar Pan-Genome by Genome-Wide Identification of Structural Variation. *Molecular Biology and Evolution*, **33**, 2706-2719. <https://doi.org/10.1093/molbev/msw161>
- [15] 赵永兵. 泛基因组学分析方法开发及应用[D]: [博士学位论文]. 北京: 中国科学院北京基因组研究所, 2014.
- [16] Sheikhzadeh, S., Schranz, M.E., Akdel, M., de Ridder, D. and Smit, S. (2016) PanTools: Representation, Storage and Exploration of Pan-Genomic Data. *Bioinformatics*, **32**, i487-i493. <https://doi.org/10.1093/bioinformatics/btw455>
- [17] Flora of China Editorial Committee (2006) Flora of China. Vol. 22, Science Press and Missouri Botanical Garden Press, Beijing, 7.
- [18] Kelchner, S.A. and Group, B.P. (2013) Higher Level Phylogenetic Relationships within the Bamboos (Poaceae: Bambusoideae) Based on Five Plastid Markers. *Molecular Phylogenetics and Evolution*, **67**, 404-413. <https://doi.org/10.1016/j.ympev.2013.02.005>
- [19] Triplett, J.K., Clark, L.G., Fisher, A.E. and Wen, J. (2014) Independent Allopolyploidization Events Preceded Speciation in the Temperate and Tropical Woody Bamboos. *New Phytologist*, **204**, 66-73. <https://doi.org/10.1111/nph.12988>
- [20] Yang, H.Q., Peng, S. and Li, D.Z. (2007) Generic Delimitations of *Schizostachyum* and Its Allies (Gramineae: Bambusoideae) Inferred from GBSSI and trnL-F Sequence Phylogenies. *Taxon*, **56**, 45-54.
- [21] Yang, H.M., Zhang, Y.X., Yang, J.B. and Li, D. (2013) The Monophyly of *Chimonocalamus* and Conflicting Gene Trees in Arundinarieae (Poaceae: Bambusoideae) Inferred from Four Plastid and Two Nuclear Markers. *Molecular Phylogenetics and Evolution*, **68**, 340-356. <https://doi.org/10.1016/j.ympev.2013.04.002>
- [22] Zhang, L.N., Zhang, X.Z., Zhang, Y.X., *et al.* (2014) Identification of Putative Orthologous Genes for the Phylogenetic

- Reconstruction of Temperate Woody Bamboos (Poaceae: Bambusoideae). *Molecular Ecology Resources*, **14**, 988-999. <https://doi.org/10.1111/1755-0998.12248>
- [23] 吴妙丽. 竹亚科的系统发育与叶绿体基因组进化[D]: [博士学位论文]. 福州: 福建农林大学, 2016.
- [24] Ma, P.F., Zhang, Y.X., Guo, Z.H. and Li, D.Z. (2015) Evidence for Horizontal Transfer of Mitochondrial DNA to the Plastid Genome in a Bamboo Genus. *Scientific Reports*, **5**, Article Number: 11608. <https://doi.org/10.1038/srep11608>
- [25] 郑朝汉. 箣竹属 - 牡竹属 - 巨竹属复合群的物种形成——杂交的分子证据[D]: [博士学位论文]. 广州: 中国科学院华南植物园, 2014.
- [26] Peng, Z., Lu, Y., Li, L., et al. (2013) The Draft Genome of the Fast-Growing Non-Timber Forest Species Moso Bamboo (*Phyllostachys heterocycla*). *Nature Genetics*, **45**, 456-461. <https://doi.org/10.1038/ng.2569>
- [27] Chen, S.L. and Gilbert, M.G. (2006) *Flora of China*. Science Press, Beijing and Missouri Botanical Garden Press, St Louis.
- [28] Isagi, Y., Oda, T., Fukushima, K., et al. (2016) Predominance of a Single Clone of the Most Widely Distributed Bamboo Species *Phyllostachys edulis* in East Asia. *Journal of Plant Research*, **129**, 21-27. <https://doi.org/10.1007/s10265-015-0766-z>
- [29] 夏湘婉, 黄云峰, 周明兵. 毛竹生物资源多样性[J]. 竹子研究汇刊, 2014, 33(4): 6-15.
- [30] 肖创伟, 余远国, 江雄波, 等. 鄂东南早出丰产早竹种(品种)的筛选[J]. 林业科技开发, 2014, 28(3): 101-105.
- [31] 杨茹, 孙志娟, 项艳. 刚竹属 14 个品种遗传多样性的 ISSR 分析[J]. 竹子研究汇刊, 2010, 29(4): 11-14.
- [32] 任立宁. 竹子种质分子鉴别的联合分析[D]: [硕士学位论文]. 广州: 中国林业科学研究院, 2010.
- [33] 杨秀艳. 粉单竹遗传多样性和麻竹种内杂交子代遗传变异研究[D]: [博士学位论文]. 广州: 中国林业科学研究院, 2007.
- [34] 甘小洪, 陈凤, 林树燕, 等. 慈竹不同变异类型的纤维形态研究[J]. 南京林业大学学报: 自然科学版, 2013, 37(4): 99-104.
- [35] 龙春玲, 刘腾飞, 于芬, 等. 厚壁毛竹与毛竹叶片的光学解剖结构比较分析[J]. 安徽农业大学学报, 2015, 42(1): 39-44.
- [36] 李建, 杨清培, 方楷, 等. 厚壁毛竹与毛竹的抗寒生理比较研究[J]. 江西农业大学学报, 2011, 33(3): 537-541.
- [37] 郭起荣. 厚壁毛竹的种质性状及其繁育研究[D]: [博士学位论文]. 长沙: 中南林学院, 2003.
- [38] 方楷, 杨清培, 郭起荣, 等. 原产地与引种地厚壁毛竹竹材成分质量分数比较[J]. 浙江农林大学学报, 2012, 29(4): 595-599.
- [39] 郭起荣, 胡芳名, 杜天真, 等. 厚壁毛竹叶片的生理生化特性[J]. 中南林业科技大学学报, 2005, 25(3): 7-11.
- [40] 方楷, 陈尚铎, 杨光耀, 等. 厚壁毛竹结构性成分含量特征[J]. 江西农业大学学报, 2014(5): 929-933.
- [41] 张莹, 田埂, 路慧萍, 等. 厚壁毛竹六个节气笋芽发育的转录组分析[J]. 江西农业大学学报, 2015(3): 466-474.
- [42] 冉洪. 厚壁毛竹种质性状的重测序研究[D]: [硕士学位论文]. 广州: 中国林业科学研究院, 2016.
- [43] 张莹. 厚壁毛竹笋——竹发育进程的转录组分析[D]: [硕士学位论文]. 广州: 中国林业科学研究院, 2016.
- [44] 成聘睿, 史正军, 刘瑞华, 等. 厚壁型巨龙竹秆材的材性特征分析[J]. 生物质化学工程, 2015, 49(4): 45-48.
- [45] 刘瑞华, 成聘睿, 史正军, 等. 薄壁型巨龙竹物理性质研究[J]. 世界竹藤通讯, 2015, 13(1): 14-17.

期刊投稿者将享受如下服务：

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：hjas@hanspub.org