



本图片由作者提供

种质资源，特别是野生种质资源是国家社会经济可持续发展不可缺少的条件之一，是一个国家的战略生物种质资源，直接影响到国家经济未来的发展潜力。

在用于优异基因鉴定和发掘的基因组学研究中，一方面需要依赖全世界已完成的基因组信息数据，另一方面需要依赖于实验室中利用各种大型仪器设备现实分析测试的实验结果。同时，对于这种海量数据的分析，还需要不同的学科交叉（植物学专家、计算机专家、基因组学专家、数学专家等）来相互配合。这样的工作模式没有一种协同工作的环境是无法完成的。

国家战略种质资源保藏与创新系统的e-Science环境

王雨华 李德铤
中国科学院昆明植物研究所, 昆明 650204

摘要: 中国西南野生生物种质资源库是国家大科学工程, 其保藏的野生种质资源是人类社会可持续发展所依赖的战略资源。本文结合国家对战略种质资源的保护和持续利用研究的需求, 利用当前先进的e-Science的科学理念和信息技术, 针对如何快速、有效地鉴定和发掘优异基因这一科学问题, 以重要野生禾本科植物的比较基因组学研究为例, 提出实施e-Science应用环境的建设方案。通过对种质资源库保藏与创新活动的全过程分析, 抓住4个关键环节(采集计划、监测评价、实验分析、创新研究), 深入了解每个环节上的信息需求, 合理部署科研工具、数据标准、模型分析和计算过程, 为科学家和支撑服务人员打造一个全面服务于种质资源采集、整理、保存、共享、传递和创新的软硬件结合、协同工作的信息化科研环境。该项工作的实施将有利于提升种质资源库的国际地位, 为我国在国际生物战略资源竞争做出贡献。

关键词: 种质资源; e-Science环境; 建设方案

The e-Science Environment for Conservation and Innovation of Germplasm Resources of Wild Species

Wang Yuhua, Li Dezhu
Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan
650204, China

Abstract: The Germplasm Bank of Wild Species is a national key project of scientific facilitation in China. The wild germplasm resources, including seeds, micro-propagated samples, genomic DNA/cDNA libraries, and mycological spores/bodies, are the basis for human sustainability. Considering the national demands to conserve and sustainably use of wild germplasm resources, the paper proposed to set up an e-Science environment for conservation and innovation of germplasm resources of wild species of plants, animals and microbes with advanced e-Science and information technology. Comparative genomic research of wild gramineae plants was taken as an example, focusing on how to identify and clone useful genes quickly and effectively. By analyzing the process of conservation and innovation of the wild plant germplasm resources, it is proposed to build a Collaboration Environment for the research of conservation technology and utilization of genetic resources, especially on the four critical links, i.e. data collection, evaluation, examination and innovation. It is hoped that the international recognition of the Germplasm Bank of Wild Species will be enhanced by making such an initiative, and it will be beneficial for China to be more competitive on the global strategic resources.

Keywords: Wild germplasm resources; e-Science environment; Proposal

1. 前言

科学大突破需要大科学工程^[1]。中国西南野生种质资源库(下文简称“种质资源库”^[2])是著名植物学家吴征镒院士向国务院建议立项的国家大科学工程,总投资1.48亿元。2005年3月开工建设,2007年建成并投入试运行。其目标是建成国际一流的、较为完整的野生生物种质资源保护设施和科学研究体系。种质资源,特别是野生种质资源是国家社会经济可持续发展不可缺少的条件之一,是一个国家的战略生物种质资源,直接影响到国家经济未来的发展潜力。二十世纪中期,美国大豆感染了胞囊线虫病,使其大豆生产濒于毁灭,后从野生大豆种质资源中筛选出抗胞囊线虫病的“北京小黑豆”,育成了高产抗病新品种,从而挽救了美国的大豆产业,并使美国大豆产量跃居世界第一位^[3]。

全世界共收集各类植物种质资源已超过600万余份,然而在上应用上并没有收到预期的效果,主要原因是如何从丰富的种质资源中快速、准确地鉴定出育种上迫切需要的新的优异基因仍是摆在科学家面前的一个关键问题,特别是“十一五”期间,我国启动了“转基因生物新品种培育”重大专项,使这一科学问题更加突出。比较基因组学是优异性状基因鉴定和发掘的重要工具,是利用某些基因组图谱和测序获得的信息推测其他生物基因组的基

因数目、位置、功能、表达机制和物种演化关系的学科。比较基因组学的发展与序列数据的积累密切相关,庞大的基因组数据信息源源不断地从一系列新技术中产生,而且未来还会进一步呈指数倍增长。探索对这些信息进行解析的方式已成为这一时代生命科学最大的挑战之一。人类科学研究史表明,科学数据的大量积累将导致重大的科学规律的发现。同时,随着大量的新型先进科学观察研究装置和技术的应用,科学研究日益依赖于从海量的信息中发现新的研究点,实现新发展^[4,5]。综合这些信息能进一步帮助我们了解物种形成的机制、基因或基因组上非编码区的功能。这个过程要面对众多不同的数学模型、方程式和算法,涉及数学、统计学和计算机学的问题。因此,离开超算、数据库和计算机,比较基因组研究将寸步难行。

2. 建设目标

紧密围绕国家大科学工程“中国西南野生生物种质资源库”的建设,面向国家对战略种质资源的需求,在现有实际科研活动和数据积累的基础上,针对如何快速、有效地鉴定和发掘优异基因这一科学问题,以重要野生禾本科植物的比较基因组学研究为例,利用当前先进的e-Science的科学理念、数据挖掘技术、协同工作技术和网络信息

技术,为种质资源的保藏和创新打造一个科研信息化环境。通过对种质资源库在重要禾本科植物的比较基因组学研究中涉及到的保藏与创新活动的全过程分析,抓住4个关键环节(采集计划、监测评价、实验分析、创新研究),合理部署科研工具、数据标准、模型分析和计算过程,真正实现一个支撑种质创新的信息化协同工作平台。

3. 建设方案

3.1 建设思路

如图1,种质资源的保藏与创新有着一套成熟的工作流程,基本分成9个部分:采集计划——野外采集——种子清理——质量检测——入库保存——发芽实验——创新研究——再生扩繁——开发利用。根据科研活动的方式可以分成四个关键环节:采集计划、监测评价、实验分析和研究创新。在每一个活动部门都有严格的操作流程和标准规范,在每一个关键环节都需要依赖原有的科学积累和分析实验设备,并产生着大量的数据。因此,这四个环节内部及其之间如何有效、科学地协作并达到数据积累和分析的目的,是种质资源保藏与创新成功的关键。

采集是种质资源库所有活动的开始,其根本目标是采集具有“3E”特性(珍稀濒危性、特有性、重要经济性)的重要野生种质,这是决定保藏什么、创新

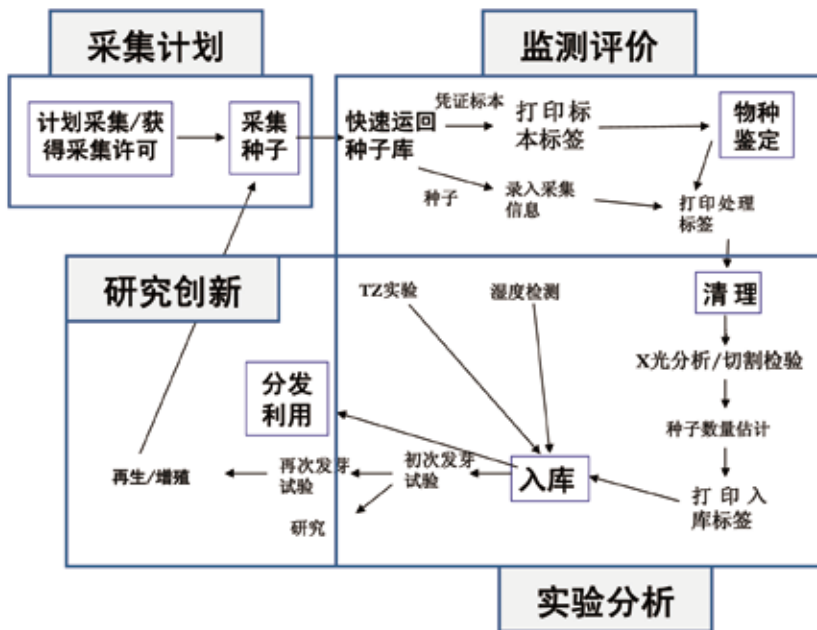


图1 种质资源库保藏与创新研究结构流程图

▶ 什么以及是否具有“3E”质量的关键一步。有计划地对西南重要野生植物种质资源进行收集是一项长期而艰巨的任务，是一个系统工程，在开始这项工作前，十分有必要进行科学的安排、开展一系列的前期工作。监测评价不仅证明采集来的物种是什么，而且要及时说明这些采集来的材料是否符合要求、从哪里来、目前状况如何等，这是决定是否保存入库的第一步。实验分析是对采集来的材料做一个全面的清理和体检分析，是决定是否能够按标准规范长期保藏的关键。研究创新就是保藏的最终目标所在，这不仅包括着本文中所提到的应用基因组学进行优异性状基因的发掘，而且涵盖了其它一切对种质资源开发利用的创新活动。在以

上4个关键环节中都需要谋划与决策，而谋划与决策一方面需要依赖于已有的积累数据，另一方面需要科学地积累其新生的数据。因此急需解决的是如何提供更方便、强大的数据积累、分析、决策工具。

建成后的种质资源库将在第一个五年内保存野生种质资源达到6450种66500份（株），十五年内将达到19000种190000份（株）。对于这种用于活体保藏的野外采集，这是一件极其艰巨的任务，因此种质资源库联系了国内外相关科研院所和大专院校的力量，一方面与具有野外采集实力的全国各地的科研院校建立紧密的合作关系，另一方面招聘一些具有致力于国家战略种质资源保护的青年志愿者。从2005

年底就开始有效地组织了辐射到包括西南、西北、华中、华北，以及华南和华东地区部分省市20多个单位近300人的团队从事这种野外采集工作。针对这种工作方式，如果没有一个可以遵循的工作规则、没有一个协调配合的模式、没有一个同步的机制和平台，不仅无法提高野外采集效率，而且将会使所有的采集变得一片混乱。

科学研究高速发展到今天，兵团作战、协力攻关已成为一种必然。在用于优异基因鉴定和发掘的基因组学研究中，一方面需要依赖全世界已完成的基因组信息数据，另一方面需要依赖于实验室中利用各种大型仪器设备现实分析测试的实验结果。同时，对于这种海量数据的分析，一方面需要以团队的形式分工、分解这些工作，另一方面又需要不同的学科交叉（植物学专家、计算机专家、基因组学专家、数学专家等）来相互配合。这样的工作模式没有一种协同工作的环境是无法完成的^[5]。随着计算机信息技术的发展，信息化的协同工作环境更加有效地促进了这种研究模式的发展。

3.2 技术路线

正如上文中所提到的，比较基因组学的研究已将研究手段紧密地与数据库和信息化结合起来。比较基因组学研究中要利用FAsTA、BLAST和CLUSTAL W等序列

比对工具,在其序列分析时要面对众多不同的数学模型、方程式和算法,涉及数学、统计学和计算机学的问题,因此,首先大型仪器设备、超级计算机及数据库是该信息化环境最直接解决的问题。基因数据分析的结果来源于微观,对其真实性的验证仍需要该物种自然环境下的性状评价,因此,对保存的种质资源进行资源评价、多样性、种子生理、生长发育规律、分子育种及遗传多样性的研究仍是信息化环境必不可少的环节。所有的科学研究和实验分析必须落实到实物材料上,特别是材料的科学选择是最终科学研究成败的关键,因此,重要野生植物资源的科学采集是信息化环境中根本解决的源头。

以鉴定和发掘优异性状基因为最终目的,围绕开展重要禾本科野生植物的比较基因组学全程的科研活动,从重要禾本科野生植物种质资源收集、保藏、监测,到实验室中的分析测试和海量数据的分析挖掘,充分利用数据挖掘技术、协同工作技术、计算机网络技术等,通过编制数据标准、建立硬件应用接口、设计模型工具等手段,从数据、设备、应用3个水平上,分别建立应用体系,从而搭建一个种质资源创新活动的协同工作的信息化环境。

4. 主要任务和内容

1) 采集计划和决策工具的研发

基于已经积累的大量参考型数据,在进一步了解用户需求的基础上,优化数据结构,修订数据质量标准;在严格遵循原始分类体系的基础上,针对比较基因组学研究中与研究材料的特殊需求,开发数据管理平台,研制数据挖掘工具,为用户提供一个科学、便捷的计划和决策工具。

2) 野生种质资源监测评价体系的构建

有效评价野外种质资源的状况是种质资源创新的重要依据,根据野外生长状态的特殊生境和优良性状,可以有效地判断优异基因的存在状况。通过生物地理学的研究,可以为比较基因组学研究中的遗传、演化提供依据。另外,种质资源库中一项重要的工作就是如何有效地监测物种种质资源保存过程中的状态,从而有效地评价该种质保存的质量状况和优异基因生存能力。该项工作主要涉及到3个阶段:保存前期的基本信息的记载、保存过程中各种保存条件和生理状况数据的收集、通过一定的评价模型进行评价。主要任务是仪器和传感器如何与平台系统有效对接,从而进行监测数据的记录,并通过模型的开发进行质量状况的评价。

3) 种质资源协同工作体系的建设

无论是种质资源的采集、保藏,还是利用种质资源进行比较基因组学研究,多个科学家或多个研究单位的参与都是必须的,

这种异地的为同一目标的参与必须有一个协同的工作环境。本项建设主要是基于现有的网络环境和信息数据,支持科研人员在完成合作科研项目过程中的相关工作。支持虚拟组织的创建、运行、管理、维护直到销毁。科研人员以虚拟组织为单元进行灵活的组合,提供面向虚拟组织的权限管理。支持一个虚拟组织当中的成员便捷、高效地共享和协同开发各类数字化文档,协助科研人员组织开展各类科研活动。

4) 种质资源创新研究体系的建设

种质资源创新主要是利用现代分子生物学、基因组学、蛋白质组学、谱系分析、数量遗传学和分子标记等研究方法和手段,对保存的种质资源进行资源评价、多样性、种子生理、生长发育规律、分子育种及遗传多样性的研究。在这些研究中涉及到对各类大型仪器设备的利用以及海量数据的产生和应用,因此,本项内容主要是仪器设备的数据标准编制、数据的采集方法、数据的提交标准和模型工具的开发。

5. 讨论及前景

中国西南野生生物种质资源库e-Science环境系统的建设是国内外关心和关注该区域的人们的迫切需求,是提高种质资源保存质量和效率的必然,是节约科研和传递成本的最终选择。种质资源库的保存根据不同的材料有着复杂的保存

►形式，有些是人们无法亲自接触的极端环境，有些是需要严格超净或超静的保存场所而不允许人们经常干扰和接触，有些是由于数量的巨大而靠人们手工无法完成的监测和监视，这就急需借助当前高度发达的信息化技术。


随着中国西南野生生物种质资源库建设的不断推进和信息化手段的应用，目前种质资源库已设计和建设了用于各个工作活动的数据库，并进行了很好的数据积累。但是，若真正实现种质资源库在质上的提升和飞跃，必须对种质资源在保藏和创新活动中存在的如何做科学的采集计划、如何提升数据质量、如何评价野生资源状况、如何引入模型进行计算等问题进行解决，这是目前种质资源库急需做好的关键环节，这种解决的最好办法就是以e-Science的理念打造一个信息化的工作环境，在采集计划、监测评价、资源共享、创新研究这4个

关键环节上突出信息化手段的应用，从而有效地实现国家战略种质资源的保藏与创新。

人类科学研究史表明，一方面，科学大突破需要大科学工程；另一方面，科学数据的大量积累将导致重大的科学规律的发现。本项目选择中国西南野生生物种质资源这个国家大科学工程作为研究对象，将大科学工程与信息技术结合起来，以科学研究中的强强结合的手段，来保藏和研究国家的战略种质资源，具有重要的战略意义。

中国西南野生生物种质资源库已经成为世界上除挪威诺亚方舟种子库和英国皇家植物园之外的第三个保存世界重要树种种质资源的机构，在大家还都停留在数据库的水平上时，若我们能以一个全面的信息化系统来考虑和建设，将比较基因组学通过信息化的科研环境灵活、全面地用于鉴定和发掘优异基因之中，将进

一步显示我国在种质资源保藏和创新方面的先进地位和作用。

近十多年，水稻的研究蓬勃发展，目前其测序工作已经完成，通过拓展特别是禾本科更具代表性的野生近缘植物的比较基因组学研究，可以将禾本科诸基因组整合成一个“大禾本科基因组”。继水稻基因组^[6、7]之后，另外一种具有C4光合作用的禾本科作物高粱的全基因组也已完成了测序^[8]。目前，美国正在进行玉米基因组的测序，预计今年将发布测序结果。今年可望完成基因组测序的第4种禾本科植物是谷子。在国家“十一五”启动的“转基因生物新品种培育”重大专项中，以禾本科作物为主的水稻、小麦、玉米占了5个品种中的3个。因此，重要野生禾本科植物的比较基因组学研究是一个理想的模式体系，可以保障该方案的成功实施，将是我院科学研究环境信息化的一个重要案例。 

参考文献:



- [1] 聚焦中国“大科学工程”. http://www.investzj.com.cn/sanji.asp?id_forum=008324, 2004.
- [2] 中国西南野生生物种质资源库. <http://www.genobank.org/>, 2008.
- [3] 李德铎, 娄治平. 履行国际公约, 建设中国西南野生

生物种质资源库. 中国科学院院刊, 2006, 21(1): 79-81.

[4] 桂文庄. 什么是e-Science?. 科研信息化技术与应用, 2008, 1(1): 1-7.

[5] 南凯, 董科军, 马永征, 等. 支持e-Science的协同工

作环境.科研信息化技术与应用, 2008, 1(1): 35-40.

[6] Yu, J., et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, 2002, 296: 79-92.

[7] Sasaki T, Matsumoto T, Yamamoto K, et al. The genome sequence and structure of rice chromosome 1. *Nature*, 2003, 420:312-316.

[8] Chen S-L, Li D-Z, Zhu G-H, et al. Poaceae. In: Wu ZY, Raven PH, Hong DY. (eds.) *Flora of China*. Beijing: Science Press, and St. Louis: Missouri Botanical Garden Press, 2006, 22: 1-730.

[9] Paterson, A. H., Bowers, J. E., Bruggmann, R. et al. The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, 2009, 457: 551-556.

[10] Xiao, J.H., Grandillo, S., Ahn, S.N., McCouch, S.R., Tanksley, S.D., Li, J.M., and Yuan, L.P. Genes from wild rice improve yield. *Nature*, 1996, 384:223-224.

收稿时间:2009年3月29日

作者信息



王雨华

中国科学院昆明植物研究所, 博士, 研究员, 研究方向为植物资源、科学数据库及其应用环境等。



李德铢

中国科学院昆明植物研究所, 博士, 研究员, 博士生导师, 中国西南野生生物种质资源库项目总经理, 研究方向为植物分类、分子系统发育、生物地理学和生物多样性保护等。