

成果与应用

水稻(籼稻)基因组

工作框架图与精细图的绘制*

中国科学院基因组生物信息学研究中心暨北京华大基因研究中心

(北京 101300)

摘要 2002年12月12日,中国科学院、国家科技部、国家计划委员会、国家自然科学基金委员会联合宣布“中国水稻(籼稻)基因组精细图完成”。此前,我国科学家于2001年10月完成了其工作框架图的绘制。水稻(籼稻)基因组工作框架图与精细图是继人类基因组计划之后所测定完成的最大的植物基因组,也是迄今为止惟一用“霰弹法”完成的最大的基因组精细图。这表明中国在基因组科学领域处于世界领先地位;也标志着中国的高性能计算在生物信息领域的综合实力。文章概要地介绍了该项成果的研究背景和意义、主要内容、研制过程和方法、应用价值及国内外的评价。

关键词 水稻基因组,工作框架图,精细图

1 研究背景与意义

水稻是世界上最重要的粮食作物,全球近半数人口以其为生。水稻基因组是迄今为止开展基因组测序的最大植物基因组,约为人类基因组的1/7,约4.3亿对碱基。通过对水稻全基因组序列分析,可以获得大量的水稻遗传信息,全面了解其遗传机理,并可获得大量用于农作物改良的极具价值的基因。

由于水稻是禾本科作物中的模式生物,对其遗传密码的破译,将促进玉米、小麦等其它重要农作物的研究和应用发展,从而带动整个粮食作物的研究。因其所蕴含的巨大社会效益和经济效益,国际上对水稻基因组研究的竞争非常激烈。日本于1991年将水稻基因组制图计划列入水稻基因组研究规划。1998年2月,由日本科学家牵头的“国际水稻基因组计划”(IRGSP)正式启动,选取日本主要栽培品种——粳稻“日本晴”为研究材料。随后,美国Monsanto公司和瑞士Syngenta公司相继开展了对“日本晴”的基因组测序研究工作。

中国及东南亚等主要水稻生产国均以籼稻及以籼稻为遗传背景的杂交稻为主要栽培品种,其种

植面积占世界稻谷生产的80%以上。中国是世界上水稻种植面积最大的国家之一,袁隆平院士培育的超级杂交水稻较其它品种增产20%—30%,为解决我国及世界的粮食问题做出了重大贡献。为开发这一宝贵的国家资源,为继续保持我国在杂交水稻育种领域的国际领先地位,由中国科学院基因组生物信息学研究中心暨北京华大基因研究中心(BGI)发起,遗传与发育生物学研究所和国家杂交水稻工程技术研究中心等共同参与的“中国杂交水稻基因组计划”于2000年5月11日正式启动,选取超级杂交稻的父本——纯种籼稻93-11为研究对象。

2 研制过程及主要内容

“工作框架图”是指覆盖基因组的大于90%序列的基因图谱,但仍存在碱基、序列不确定的区域和未在染色体上定位的序列。“精细图”则是指基因组中所有基因覆盖区域碱基序列精确、基因精确定位于染色体上的基因图谱。目前,国际通用的测序方法主要有两种——霰弹法和克隆法,这是两种不同的技术路线。BGI采用的是“全基因组鸟枪法测序”,即霰弹法,是将基因组DNA打成约2Kb—3Kb

* 该研究得到中国科学院、国家科技部、国家计划委员会、国家自然科学基金委员会、北京市政府、浙江省政府及杭州市政府的大力支持

收稿日期:2002年12月29日

的片段进行测序,然后再将这些小的片段拼接起来,重新组装成一个完整的基因组。它的最大优点是经济、快速、高效,但对高性能计算的方法和设备要求非常高。

经过一年多的前期准备,中国水稻(籼稻)基因组测序工作于 2001 年 7 月 1 日开始大规模加速。2001 年 10 月 12 日,由中国科学院、国家科技部、国家计划委员会联合宣布完成了水稻(籼稻)基因组“工作框架图”的绘制,并公布数据库供无偿使用。该框架图要点包括:测序工作量达水稻基因组的 4 倍;基因组覆盖率达 92%以上,即经初步组装的“一致性”序列占整个基因组的 90%以上;经初步组装的“一致性”序列中碱基的准确率达 99%以上,即每一碱基出错的可能性(误差率)低于 1%。数据库的无偿使用,得到国内外同行的一致好评。至今,已有 22 个国家 20 多万人次访问和下载了中国的水稻数据。“水稻(籼稻)基因组工作框架序列图”以 14 页的篇幅发表在 2002 年 4 月 5 日的《科学》(美)上,并以青山衬托下的一片金灿灿的云南哈尼梯田作为该期的封面。同时,还配以大量评论文章,高度评价该工作是“具有最重要意义的里程碑性工作”,对“新世纪人类的健康与生存具有全球性的影响”,“永远改变了我们对植物学的研究”,并祝贺论文作者“对世界科学乃至全人类里程碑的贡献”。该成果还被两院院士评选为 2001 年度中国十大科技新闻之一。2002 年 5 月 28 日,江泽民主席在两院院士大会上的讲话中,将其列为近代中国生命科学对世界三个主要贡献之一:“在当代世界科技发展的史册上,我国科技工作者也书写了光辉的篇章:在世界上首次人工合成牛胰岛素;完成了人类基因组百分之一的测序研究工作;率先公布了水稻基因组框架图……”。2002 年 9 月 5 日,该项目组荣获由香港求是科技基金会颁发的 2002 年度“求是杰出科技成就集体奖”。2002 年 11 月 11 日,BGI 主任杨焕明、于军又因在水稻基因组研究中的巨大贡献,被拥有 150 多年历史并享誉世界的杂志《科学美国人》评为“2002 年度全球科研领袖”。

框架图的研究成果也极大地加速了全球水稻基因的研究工作,中国也投入了近两亿元的资金,支持基于水稻基因组的功能和应用开发研究。水稻

(籼稻)基因组精细图是我国后续水稻功能和应用研究以及水稻全基因组第三代芯片研制的基础,同时也是比较基因组学研究的基础。“精细图”和“工作框架图”的差别虽然只有几个百分点,但它对基因预测、基因功能鉴定的准确性以及基因表达、遗传育种等研究的贡献是一个质的飞跃。IRGSP 已将 2002 年底前完成 10 倍覆盖度的图谱作为第二期目标。因此,在水稻(籼稻)基因组工作框架图中已取得国际领先优势的中国籼稻基因组研究,必须在激烈的国际竞争中,加速完成“精细图”的绘制。

BGI 根据现有经济实力和“工作框架图”的结果,着重进行基因所在区域内的“精细图”绘制,对于因重复序列所造成的基因间的不连续,视经费情况再行定夺。这种技术思路,避开了最为棘手的重复序列问题,既可节省大量的时间和财力,又不降低“水稻精细图”的整体价值。终于在 2002 年 12 月初率先完成了水稻(籼稻)全基因组“精细图”的绘制。该成果的主要内容和应用如下:

(1)完成了水稻(籼稻)全基因组“精细图”。该图覆盖了 97%的基因序列,并可将其中 97%的基因精确地定位在染色体上。覆盖基因组 94%的染色体定位序列准确性达到了 99.99%。已达到国际公认的精细基因图的标准。它是迄今为止世界上惟一的基于“全基因组鸟枪法”构建的大型植物基因组高精度基因图。

(2)完成了水稻亚种内和亚种间分子遗传标记图谱。通过对籼稻(indica)和粳稻(japonica)亚种基因组已定序列的比较分析,发现了 100 多万个单核苷酸多态性,将这些分子遗传标记在染色体上定位,并整合在精细基因图谱上。这些标记物可以用来鉴别基因的来龙去脉,追踪它们在遗传群体中和杂交过程中的分布,进而指导遗传育种实践。

(3)预测出约 6 万个水稻基因。利用这些信息,制备出了全基因组基因芯片。为功能基因组研究提供了强有力的工具,为大规模分离抗病、高产、优质的相关基因奠定了基础。

(4)通过比较基因组学研究,发现水稻和拟南芥基因组在基因组结构、基因表达和基因功能方面存在广泛差异。以水稻为代表的单子叶植物与拟南芥相比具有很多新的基因,为解释这些基因的新功

能和进一步研究单、双子叶植物在分子进化和生理生化上差异的分子机制研究奠定了基础。

(5)建立了全基因组鸟枪法测序基因组组装的计算机软件体系。运用全基因组鸟枪法测序,组装成高精度全基因组的基因图,是一项技术上的新突破,在植物基因组研究中属首创。植物基因组有大量的重复序列,它们主要分布在基因和基因之间,使基因很容易重复和翻转,同时为基因序列的组装造成巨大困难。这些重复序列的正确识别和组装,需要开发特殊的计算软件。该体系的建立,为开展其它重要物种的基因组研究开辟了一条全新的经济、快捷和可靠的研究方法。

3 应用价值

和人类基因组计划一样,水稻基因组研究是与信息科学密切结合的大科学工程。从基因测序到组装,从基因预测到全基因组芯片的研制,从分子标记图谱到基因和蛋白质表达图谱,充分体现了大科学工程从基础到应用的科学研究系统性、原创性及可持续发展性。大规模的水稻基因组测序,促进了测序技术的集成;工业规模的高通量 DNA 测序,带来了规模效应,技术革新和规范管理大大降低了成本。

基因组大量数据的产出,对我国海量信息处理体系和高性能计算的软硬件环境提出了前所未有的需求和挑战。它全面促进了相关算法发展和软件开发。计算技术研究所的大型计算机“曙光 3000”,国家北方计算中心的大型计算机“神威”,都为水稻基因组的工作做出了积极的贡献。高性能计算在生

物学领域的成功应用,对人们在海量信息的基础上探讨生物的奥秘,从方法论和观念上都带来革命性的进步。

自主基因预测软件的开发,使得准确地预测基因和覆盖全基因组的基因芯片研制成为可能。基因芯片将基因组学的基础研究与功能和应用研究密切联系起来,将大大加快水稻基因表达的动态研究。

该工作还全面带动了我国大规模 DNA 测序、全基因组的基因芯片制作、海量生物信息处理等规模化、集成性的平台体系建设。它标志着我国在基因组领域领先的地位,也是我国在攀登科学高峰的征途中进行原创性科学研究和高新技术集成等方面全面创新的综合体现。它实现了跨学科人才队伍的凝聚和培养。为进一步开展所有农业与环境相关的基因组研究打下了良好的基础。

由于科学和技术上的创新及我国科学工作者艰苦奋斗和勤俭节约,中国水稻基因组工作在“投入/产出比”上远优于国际同类项目。该项目的高速、高质和低投入,已成为国际大科学项目的重要典范之一。由此吸引了国际同行对中国基因组研究的密切关注,美国、丹麦等国的政府以及世界银行等投资机构已在基因组相关的领域与我国进行合作谈判。

该成果的取得将为全面了解水稻的生长、发育、抗病、抗逆和高产规律,推动遗传育种,解决粮食问题带来革命性的突破。

(本刊编辑部整理)

The Mapping of The Working Draft and The Fine Map of The Rice Genome
(*Oryza sativa* L. *ssp indica*)
(BGI,CAS, Beijing 101300)

Chinese Academy of Sciences, Ministry of Science and Technology, the State Development Planning Commission, and the National Natural Science Foundation of China announced together "the completion of the fine map of the Chinese rice (*Oryza sativa* L. *ssp indica*) genome" on December 12th, 2002. And Chinese scientists have mapped its working draft in October 2001. It is the biggest botanic genome sequenced after the Human Genome Project, and is the biggest one sequenced with the shotgun strategy so far. It shows that China has taken the leading place in the genome science field, and also demonstrates the general strength of Chinese high performance computation in the bioinformatics field. The article introduces summarily the research background, the significance, the main content, the research process and the methods of the achievement and the appraisal on it at home and abroad.

Keywords rice genome, working draft, fine map