

DOI: 10.13376/j.cbbs/2018125

文章编号: 1004-0374(2018)10-1038-06



黎志康, 二级研究员, 博士生导师, 中国农业科学院农作物基因资源与基因改良国家重大科学工程首席科学家。1977年毕业于安徽农业大学, 1983年获得中国农业科学院研究生院硕士学位, 1989年获得美国加州大学戴维斯校区遗传学博士学位并开展博士后工作, 1990年以来在美国德州农工大学和菲律宾国际水稻所工作22年。2003年, 任国际水稻研究所驻中国代表科学家, 兼任中国农业科学院作物科学研究所研究员、博士生导师和中国农作物基因资源与基因改良国家重大科学工程首席科学家, 2012年全职回国。国际动植物基因组年会植物分子育种分会主持人, 第二、三届国际植物分子育种研讨会主席。《Plant Breeding》、《Plant Genome》、《JIA》杂志编委, 《分子植物育种》杂志副主编。美国遗传学学会和作物科学学会会员。发表论文260余篇, 其中在《Nature》、《Science》、《PNAS》、《Nucleic Acids Res》等SCI刊物发表论文160余篇, 出版英文专著章节15篇。论文被引15521次, 连续3年入选Elsevier中国高被引学者农业与生物科学学科榜单前10名。主持完成的“3000份水稻基因组计划”项目将大大推动水稻种质资源优异基因的规模化发掘和我国功能基因组研究成果向水稻育种中的转化。

水稻基因组多样性及其在绿色超级稻育种中的应用

黎 珉¹, 王文生², 徐建龙², 黎志康^{1,2*}

(1 安徽农业大学农学院, 合肥 230036; 2 中国农业科学院作物科学研究所, 北京 100081)

摘 要: 绿色超级稻育种策略是应对未来全球粮食安全问题的有效途径, 水稻种质资源中丰富的遗传多样性为绿色超级稻的培育提供了坚实的遗传基础。近年来, 水稻基因组研究取得了重大进展, 特别是“3000份水稻基因组计划”的完成, 充分揭示了水稻种内丰富的遗传多样性, 绘制了全球首个接近完整、高质量的亚洲栽培稻泛基因组图谱, 为绿色超级稻的培育提供了更加完整的数据平台与理论基础。现介绍“3000份水稻基因组计划”的最新研究进展及其对绿色超级稻育种的应用意义。

关键词: 基因组多样性; 绿色超级稻; 泛基因组; 分子育种

中图分类号: S511

文献标志码: A

The diversity of rice genome and its application in the breeding of “Green Super Rice”

LI Min¹, WANG Wen-Sheng², XU Jian-Long², LI Zhi-Kang^{1,2*}

(1 College of Agronomy, Anhui Agricultural University, Hefei 230036, China;

2 Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China)

Abstract: The strategy of “Green Super Rice” breeding is an effective way to address future global food security

收稿日期: 2018-10-18

基金项目: 国家高技术研究发展计划(“863”计划)(2014AA10A601); 比尔及梅琳达·盖茨基金会(OPPGD51587, 1393, 1130530)

*通信作者: E-mail: zhkli1953@126.com

issues. The rich genetic diversity of rice germplasm resources provides a solid genetic basis for the development of Green Super Rice. In recent years, significant progress has been made in rice genome research, especially the completion of “3 000 rice genome projects”, which fully reveals the rich genetic diversity in rice species and, the world's first near-complete, high-quality pan-genome map of Asian cultivated rice has been built, which provided a more complete data platform and theoretical basis for the development of Green Super Rice. This article introduces the latest research progress of “3 000 rice genome projects” and its application in the breeding of Green Super Rice.

Key words: genomic diversity; Green Super Rice; pan-genome; molecular breeding

水稻 (*Oryza Sativa* L.) 是世界主要的粮食作物之一, 是全球超过 30 亿人口的食物来源。按照目前的人口增长速度, 预计到 2050 年全球人口将超过 90 亿。这一增长更多发生在贫穷人口高密度区域 (如南亚、非洲), 这些地区多以水稻为主食, 粮食需增产 60%~70% 才能够满足人口增长的需求^[1]。中国作为全球最大的水稻生产国和消费国^[2], 水稻的可持续性生产对于我国乃至世界的粮食安全具有举足轻重的作用。过去的 60 年里, 我国水稻生产分别经历了矮化育种和杂交水稻的研究突破, 产量已有显著提高。当前, 农业生产与环境资源的矛盾日益突出。水稻用水约占农业用水总量的 70%, 农药和化肥的使用量也远高于世界平均水平, 而且我国相当部分耕地是中低产田。在水资源缺乏、耕地减少、自然 / 生物灾害频发等严峻形势下, 加快转变传统生产方式, 逐步形成资源节约型、环境友好型的“两型农业”生产体系, 是解决上述问题的有效途径^[3-4]。其中, 培育并大面积推广“节水抗旱、抗病抗虫、肥料利用高效、优质高产”的绿色超级稻 (Green Super Rice) 新品种^[5]是实现我国两型农业的核心目标, 也是实现我国农业可持续发展的国家战略。

与现有的超级稻相比, 绿色超级稻不仅要高产优质, 更要具有多种绿色性状。这些绿色性状包括对各种逆境 (干旱、盐、碱、高温、低温、作物营养元素包括微量元素缺乏、各种病虫害等) 的抗性和水肥的高利用效率。这些绿色性状在遗传上几乎都涉及大量的基因。育种家要同步改良多个复杂性状面临着极大的挑战。针对这一挑战, 我国的科学家早在 1998 年就启动了“参与全球水稻分子育种计划研究”的项目。20 年来, 我国十多个研究单位充分利用水稻种质资源中的丰富遗传多样性, 通过大规模杂交、回交和目标性状的定向选择, 将一套水稻微核心种质中与高产、优质、抗性绿色性状相关的有利基因导入到现有优良推广品种中, 培育出大量优良遗传背景下的导入系群体^[6], 建立了全

球最大的水稻分子育种材料平台。利用目标性状导入系和分子标记剖析并发掘控制绿色性状有利基因和遗传网络的理论技术体系已经建立, 并正在逐步完善^[7-9]。另一方面, 我国水稻功能基因组研究取得了重大进展, 克隆了大量重要的水稻基因^[3]。然而, 这些研究成果还不能充分地应用到育种上, 一个主要的原因在于尚未充分了解作为育种基础的水稻种质资源基因组的多样性, 而这些遗传信息是未来新品种培育所必需的。

水稻从野生稻驯化 1 万多年来, 亚洲栽培稻及其祖先普通野生稻存在多种类群, 分布广泛, 适应多样的生态环境。水稻种群中丰富的遗传多样性是其进化和现代育种的基础, 也是应对未来粮食需求和环境变化进行品种改良的关键资源, 无论过去还是未来, 育种成果都取决于对种质资源遗传多样性的利用。全球共计保留了 78 万份水稻种质资源, 如何在育种上充分利用这些种质资源中的有利遗传变异, 一直是水稻育种研究面临的巨大挑战。培育高产高效的绿色超级稻的一个基本策略, 就是要将种质资源和基因组学技术进行整合, 建成基于资源组合的遗传信息平台, 使未来育种建立在大数据的精准遗传信息上。为了充分揭示亚洲栽培稻种内基因组遗传多样性和群体遗传结构, 为我国水稻功能基因组研究和绿色超级稻分子设计育种提供必需的基因组遗传信息, 中国农业科学院、国际水稻研究所、华大基因等十多个国内外研究机构共同完成了“3 000 份水稻基因组计划”, 发现了蕴藏在水稻种质资源中的巨量的自然变异。这 3 000 份材料来源于全球 89 个国家, 代表了全球水稻种质资源 95% 的遗传多样性。该计划是迄今为止全球最大的植物基因组测序项目, 共产生了 17 T 的测序原始数据^[10-11]。

本文将概述该研究的主要成果以及水稻种内基因组多样性的特征及其在育种上的可能应用, 并就如何以此为基础, 利用构建的目标性状导入系逐步建立高效的绿色超级稻培育分子育种技术体系提出进一步的设想。

1 水稻基因组多样性的类型及其在遗传育种上的应用

1.1 单核苷酸多态性

与国际测序品种日本晴的参考基因组比对, 在 3 010 份水稻基因组中发现了超过 2 900 万个单核苷酸多态性 (single nucleotide polymorphisms, SNPs) 和超过 25 万个小插入缺失标记 (small insertion and deletion polymorphisms, InDels)。这些 SNPs 中, 效应中等 (导致氨基酸变异) 及大效应 (导致移码或终止突变) 的分别达到 1 356 313 和 56 786 个^[12]。与 5 个其他不同的参考基因组 (IR8、N22、93-11、明恢 63、珍汕 97) 比对, 发现的 SNPs 总数超过 4 200 万。结果表明, 水稻群体基因组中存在极其丰富的 SNPs 变异, 数量多且密度高。基于其中 2 859 份水稻基因组的 SNPs 及 InDels 变异数据, 建立了综合性的水稻功能基因组育种数据库的 SNPs 与 InDels 多态性子数据库 (<http://www.oryzasnp.org/iric-portal/>, <http://www.rmbreeding.cn/snp3k/>)^[13-14]。该子数据库包含了 3K 水稻多态性信息检索、基因组浏览器可视化系统、特定区段基因组数据导出系统等多项功能。该数据库的建立将为研究水稻基因功能、指导水稻全基因组选择育种提供重要平台。为了方便在水稻遗传育种上的应用, 将这些 SNPs 分为不同数量的五套, 分别是 27M、17M、4.8M、1M 和 404K, 以满足不同的研究需求。与湖南华智水稻生物技术有限公司合作, 利用这些 SNPs 开发了一款新的基于 Affymetrix 的高通量水稻 SNPs 基因型芯

片。这款芯片共携带 56 000 多个均匀分布在 12 条水稻染色体上的 SNPs, 每两个相邻 SNPs 的平均间距为 6.4 kb (中数为 5.3 kb), 对已知功能的 200 多个重要水稻基因, 每个位点上的 SNPs 超过 20 个 (图 1)。目前, 该款芯片已在全基因组关联分析 (genome-wide association studies, GWAS)、指纹图谱构建、水稻分子设计育种等方面得到广泛应用。

依据 SNPs 数据对亚洲栽培稻群体进行了精细的种群分类, 将亚洲栽培稻分为 9 个群体, 包括籼稻亚种的 4 个群体 (中国籼稻群、南亚籼稻群、东南亚籼稻群和现代籼稻品种群)、粳稻亚种的 3 个群体 (温带粳稻群、热带粳稻群、亚热带粳稻群) 和来自南亚 (印度和孟加拉国) 的 Aus 群和香稻 (Aromatic) 群, 充分揭示了水稻种内丰富的群体结构和遗传多样性 (图 2)^[12]。水稻群体分化的结果对于水稻遗传育种研究具有重要的指导意义, 其中包括如何选择 3 000 份水稻的部分材料进行全基因组关联分析 (GWAS), 挖掘影响种群间或种群内重要性状遗传差异的有利基因, 如何利用来自不同群体的材料建立杂交水稻育种中的不同异质群 (heterotic groups), 以充分发掘不同种群间的杂种优势或利用来自不同种群的材料作为亲本拓宽育种群体的遗传基础等。

1.2 水稻基因组的微细结构变异

与传统细胞遗传学中可检测到的染色体结构变异或基因组研究中所指的微小 (< 30 bp) 插入或缺失 (InDels) 不同, 在 3 010 份水稻核心种质中的 453 份高覆盖度基因组中检测到超过 93 000 个缺失 [平

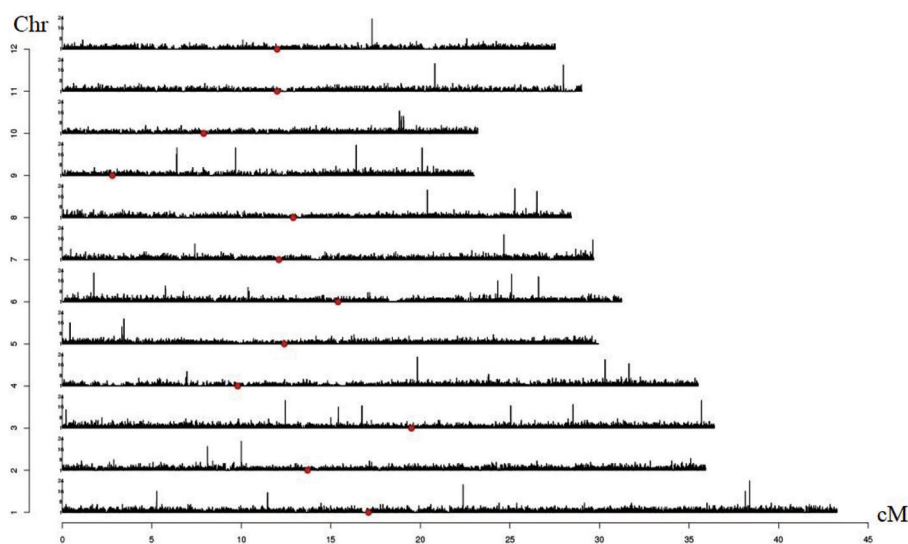
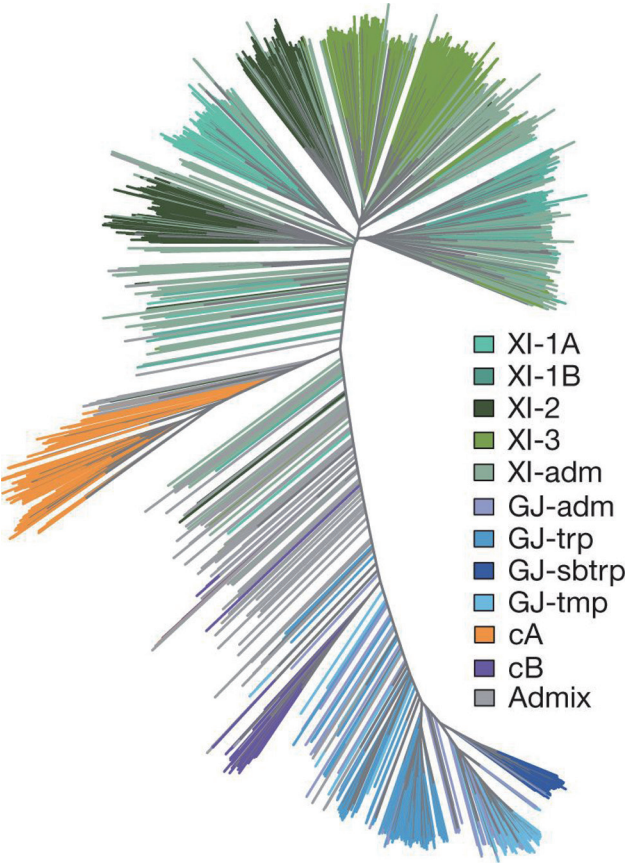


图1 Affymetrix芯片上56 000多个SNP位点在12条水稻染色体上的分布

均大小为 (4.7 ± 0.6) kb]、倒位 [平均大小为 (127.2 ± 18.9) kb]、易位、重复 [平均大小为 (111.0 ± 25.9) kb] 等微细基因组结构变异 (structural variation, SVs)^[12], 平均一个基因组超过 12 000 个。这些结构变异的大



XI-1A: Xian/indica subpopulation 1A; XI-1B: Xian/indica subpopulation 1B; XI-2: Xian/indica subpopulation 2; XI-3: Xian/indica subpopulation 3; XI-adm: Xian/indica admixed types; GJ-adm: Geng/japonica admixed types; GJ-trp: Geng/japonica tropical subpopulation; GJ-sbtrp: Geng/japonica subtropical subpopulation; GJ-tmp: Geng/japonica temperate subpopulation; cA: centum-Aus population; cB: centum-Basmati population

图2 基于3 010份水稻基因组SNPs构建的系统进化树^[12] (与日本晴基因组比对)

小绝大多数在 100 bp 到 1 Mb 之间 (表 1)。

这些结构变异在水稻群体中大多频率很低, 但有相当一部分群体特异的结构变异在某个水稻群体中的频率很高, 而在其他群体中几乎或根本不存在。平均而言, 任意两个籼-粳基因组之间的结构变异数达到 14 653 个, 粳-粳间为 4 809 个, 籼-籼间超过 5 000 个, 粳-Aus 间为 14 128 个, 粳-香稻间为 9 547 个。对于每个水稻材料而言, 这些结构变异所影响到的功能基因数平均达到 $(1\,238 \pm 415)$ 个 (缺失)、 (109 ± 42) 个 (重复) 和 $(2\,465 \pm 983)$ 个 (倒位) (表 2, 待发表资料)。水稻基因组间存在如此大量的结构变异, 一方面解释了水稻分离育种群体和杂交后代中常见的衰退 (hybrid breakdown) 现象, 另一方面说明了按照单一参考基因组定位的基因在染色体上位置的不确定性, 因此, 育种亲本基因组间结构变异的信息是未来精确设计育种必须考虑的因素。

1.3 亚洲栽培稻泛基因组和种内基因存在/缺失变异

泛基因组 (pan genome) 概念是指某一物种全部基因的总称, 包括一套在该物种所有个体基因组中都携带的核心基因或基因家族 (core genes/gene families) 以及一套只在某些个体基因组含有的分散式基因或基因家族 (distributed genes/gene families)。Zhao 等^[15] 利用 66 个来自不同水稻类群的栽培稻品种和野生稻株系构建了栽培稻-野生稻的泛基因组, 但限于样本数量以及测序深度, 所构建的水稻泛基因组完整性较低。“3 000 份水稻基因组计划” 构建

表2 各类结构变异影响的功能基因数目

群体	缺失	倒位	重复	总计
籼	$1\,449 \pm 153$	$2\,951 \pm 537$	128 ± 34	$4\,527 \pm 656$
粳	527 ± 220	893 ± 352	57 ± 22	$1\,478 \pm 560$
Aus	996 ± 156	$1\,703 \pm 477$	100 ± 27	$2\,799 \pm 590$
香稻	$1\,418 \pm 136$	$2\,752 \pm 468$	92 ± 24	$4\,262 \pm 551$
混合	$1\,141 \pm 284$	$2\,239 \pm 766$	83 ± 19	$3\,463 \pm 1\,029$
平均	$1\,238 \pm 415$	$2\,465 \pm 983$	109 ± 42	$3\,812 \pm 1\,404$

表1 453份高覆盖度基因组结构变异数目^[12]

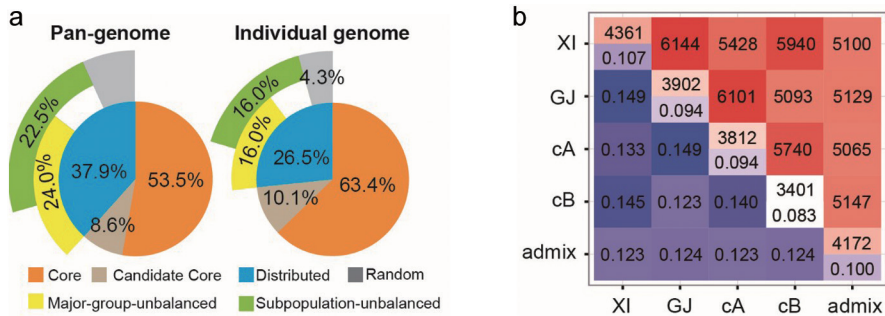
群体	缺失		倒位		重复		易位	总计
	数量	大小(kb)	数量	大小(kb)	数量	大小(kb)	数量	数量
籼	$5\,411 \pm 631$	5.2 ± 0.4	188 ± 35	127.5 ± 12.9	165 ± 33	111.8 ± 15.6	$8\,990 \pm 2\,404$	$14\,754 \pm 3\,066$
粳	$1\,670 \pm 770$	5.5 ± 1.0	59 ± 27	123.3 ± 33.0	62 ± 25	83.0 ± 29.0	$2\,397 \pm 1\,248$	$4\,189 \pm 2\,045$
Aus	$5\,068 \pm 1\,043$	4.7 ± 0.7	161 ± 37	132.0 ± 13.8	141 ± 32	111.3 ± 17.1	$7\,628 \pm 2\,031$	$12\,997 \pm 3\,055$
香稻	$3\,200 \pm 855$	5.8 ± 0.4	105 ± 37	131.3 ± 23.1	97 ± 29	89.0 ± 17.1	$4\,491 \pm 1\,863$	$7\,892 \pm 2\,768$
混合	$4\,309 \pm 1\,146$	5.0 ± 5.7	136 ± 37	128.7 ± 21.7	129 ± 36	106.5 ± 23.3	$6\,161 \pm 2\,283$	$10\,735 \pm 3\,406$
平均	$4\,535 \pm 1\,667$	5.3 ± 0.6	152 ± 62	127.1 ± 19.4	139 ± 52	105.1 ± 22.7	$7\,348 \pm 3\,429$	$12\,178 \pm 5\,133$

的亚洲栽培稻的泛基因组共包含了 23 876 个基因家族，由 14 826 (62.1%) 个核心基因家族和 9 050 (37.9%) 个分散式基因家族组成 (图 3a)，其中包含 12 465 个粳稻中没有的水稻全长新基因和 9 000 多个部分序列的新基因，这是目前首个接近完整、高质量的亚洲栽培稻泛基因组^[12]。图 3b 显示了来自两个不同水稻群体品种间的基因家族存在 / 缺失变异 (gene presence/absence variation, PAV) 数目上的平均差异。其中粳 - 籼品种间差异有 6 144 个基因家族，粳 - 粳品种间为 4 361 个，而籼 - 籼品种间也达到

3 902 个^[14]。可以预见，基因 PAV 变异在未来的功能基因组和分子设计育种中将有广泛的应用。例如，应用基因 PAV 联合 SV 进行关联分析，可以直接检测到与复杂性状变异关联的候选基因 (图 4)^[16]。

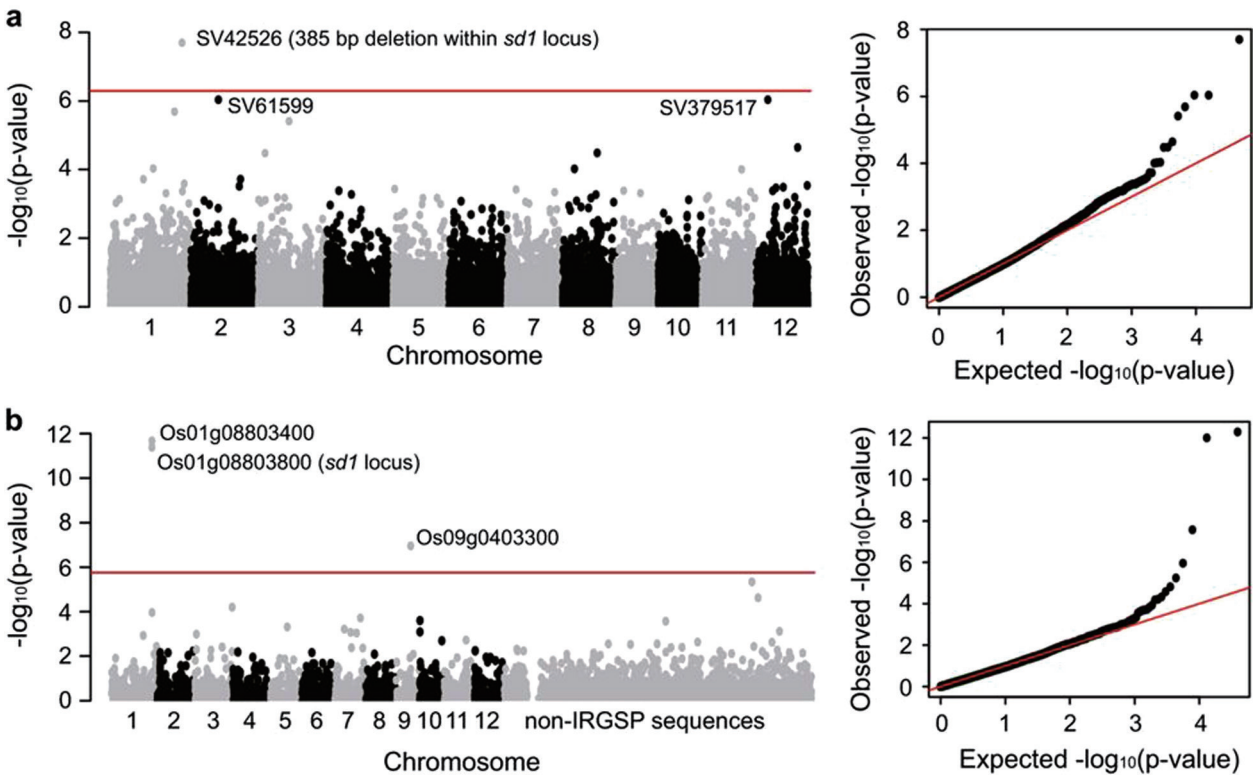
1.4 水稻基因的功能单倍型多样性

与大多数 SNPs 不同的是，由基因内重要区域的大效应 (非同义) SNPs 和 InDels 构成的基因单倍型多样性 (gene haplotype polymorphism or gene functional allelic diversity) 通常具有明显的功能或表型效应，因而代表了每个基因位点上最重要的等位基因变



a: 水稻泛基因组和单个个体基因组的构成；b: 不同水稻群体间基因家族数目和比例的平均差异。

图3 水稻泛基因组^[12]



a: 基于结构变异(缺失、重复、倒位)；b: 基于分散式基因的存在/缺失变异。

图4 基于结构变异和存在/缺失变异对株高进行全基因组关联分析^[16]

异。例如, 3 000 份水稻基因组变异研究中分析了 9 个重要的驯化相关基因的单倍型, 发现粳稻中携带的等位基因在大部分籼稻品系中都可以找到, 但是许多籼稻中携带的等位基因在粳稻中是丢失的, 籼粳亚群间在许多基因位点上的单倍型呈现明显的差异^[12]。目前, 正在根据 3 000 份水稻基因组中发现的 SNPs、InDels 和泛基因组数据构建水稻所有基因单倍型多样性的数据库, 一旦建成, 将大大促进水稻功能基因组和分子设计育种的研究。

2 展望

“3 000 份水稻基因组项目”对水稻基因组结构变异与基因存在/缺失变异的揭示完善了对水稻基因组多样性的了解, 将极大推动水稻功能基因组的研究。可以预见, 在未来 5~10 年内, 随着水稻泛基因组图谱的构建与完善, 越来越多的水稻基因功能将得到解析。目前已建立的重要数据库和资源包括 SNP 和表型数据库 (Rice SNP-Seek Database)^[13] 和 3K 泛基因组数据库 (RPAN: Rice Pan-genome Browser)^[17], 以及正在构建的结构变异及单倍型数据库。尽管如此, 这些数据库与未来水稻分子设计育种所需的大数据库尚有距离。理想的数据库应该至少包括以下 4 部分: (1) 完整的水稻泛基因组数据, 即所有水稻基因、每个基因位点上不同的等位基因及其功能 (分子功能、影响的表型以及如何影响这些表型); (2) 水稻核心种质的基因组构成 (每个种质携带的所有基因及其染色体位置); (3) 优良种质 (生产上正在使用的优良品种或杂交稻亲本) 的基因组精准构成 (高质量全基因组序列); (4) 主要农艺性状 (产量、适应性、抗病虫性、抗逆性、品质等) 的遗传基础和基因调控网络。

随着大规模水稻种质资源重测序工作的完成和水稻功能基因组研究的飞速进展, 上述的 (1)、(2) 和 (4) 正在逐步建立和完善, 而 (3) 则将由育种单位分别在实施分子设计育种的过程中构建。目前面临的新挑战是如何充分利用数据库中水稻各类群中丰富的遗传变异, 根据具体育种目标, 设计不同生态地区的绿色超级稻新品种所需要的理想“目标基因组类型 (target genomic constitutions)”, 并运用各种现代育种技术手段, 通过精准的亲本选择、杂交

组合配制、后代群体目标性状 (目标基因) 的聚合和定向改良以及育种后代的快速稳定, 高效地培育出符合要求的绿色超级稻新品种。

[参 考 文 献]

- [1] 余四斌, 熊银, 肖景华, 等. 杂交稻与绿色超级稻. 科学通报, 2016, 35: 3797-803
- [2] Elert E. Rice by the numbers: a good grain. Nature, 2014, 514: S50-1
- [3] 余四斌, 汤欣欣, 罗利军. 功能基因组与绿色超级稻培育的研究进展. 生命科学, 2016, 10: 1287-94
- [4] Wing RA, Purugganan MD, Zhang Q. The rice genome revolution: from an ancient grain to Green Super Rice. Nat Rev Genet, 2018, 19: 505-17
- [5] 张启发. 绿色超级稻培育的设想. 分子植物育种, 2005, 3: 601-2
- [6] 黎志康. 我国水稻分子育种计划的策略. 分子植物育种, 2005, 3: 603-8
- [7] Li ZK, Fu BY, Gao YM, et al. Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (*Oryza sativa* L.). Plant Mol Biol, 2005, 59: 33-52
- [8] Zhang F, Zhai HQ, Paterson AH, et al. Dissecting genetic networks underlying complex phenotypes: the theoretical framework. PLoS One, 2011, 6: e14541
- [9] Cui Y, Zhang F, Xu J, et al. Mapping quantitative trait loci in selected breeding populations: a segregation distortion approach. Heredity, 2015, 115: 538-46
- [10] Li JY, Wang J, Zeigler RS. The 3,000 rice genomes project: new opportunities and challenges for future rice research. Gigascience, 2014, 3: 8
- [11] 3,000 rice genomes project. The 3,000 rice genomes project. Gigascience, 2014, 3: 7
- [12] Wang W, Mauleon R, Hu Z, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature, 2018, 557: 43-9
- [13] Alexandrov N, Tai S, Wang W, et al. SNP-Seek database of SNPs derived from 3000 rice genomes. Nucleic Acids Res, 2015, 43: D1023-7
- [14] 郑天清, 余泓, 张洪亮, 等. 水稻功能基因组育种数据库 (RFGB): 3K水稻SNP与InDel子数据库. 科学通报, 2015, 4: 367-71
- [15] Zhao Q, Feng Q, Lu H, et al. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat Genet, 2018, 50: 278-84
- [16] Hu Z, Wang W, Wu Z, et al. Novel sequences, structural variations and gene presence variations of Asian cultivated rice. Sci Data, 2018, 5: 180079
- [17] Sun C, Hu Z, Zheng T, et al. RPAN: rice pan-genome browser for ~3000 rice genomes. Nucleic Acids Res, 2017, 45: 597-605