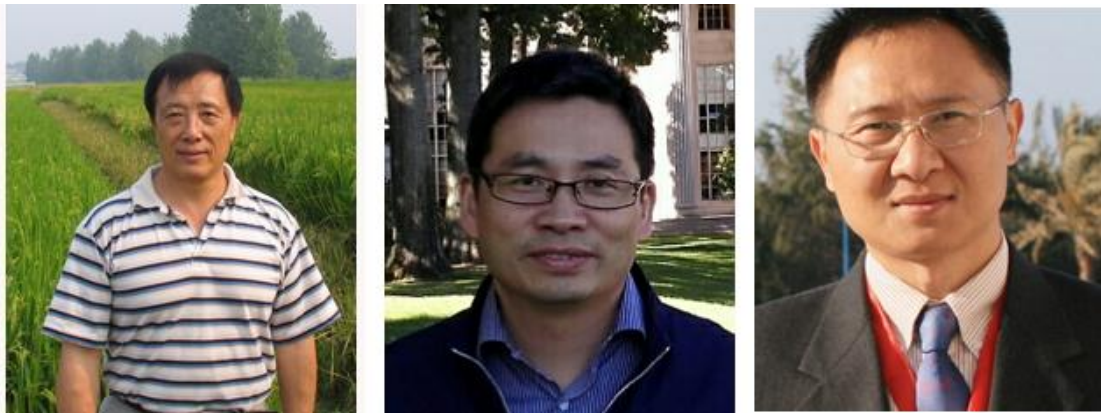


科学家完成亚洲栽培稻基因组变异研究

Scientists Complete Genomic Variation of Cultivated Rice in Asia



黎志康研究员、韦朝春教授、张耕耘教授（从左至右）

4月25日，由中国农业科学院、国际水稻研究所、上海交通大学、华大基因、深圳农业基因组研究所、安徽农业大学、美国亚利桑那大学等国内外16家单位历时7年共同协作攻关完成的“3000份水稻基因组计划”结出硕果——文章以“Genomic variation in 3,010 diverse accessions of Asian cultivated rice”为题发表在《Nature》期刊。

水稻是全世界最重要的粮食作物之一，预计到2050年，全球的稻米产量必需增加1倍以满足未来人口增加的需求。而且，在全球气候变化的趋势下，如何培育优质、高产、广适的水稻新品种一直是育种家面临的巨大挑战。过去20多年，全球科学家已经克隆了多个水稻基因，然而，如何将这些功能基因组的研究成果转化为育种家所能利用的分子设计育种信息，则必须对水稻种质资源的基因组信息进行充分了解。

“3000份水稻基因组计划”正是在这样的背景下应运而生，研究团队对来自全球89个国家的代表了78万份核心种质约95%遗传多样性的3010份水稻进行了重测序和大数据分析，针对水稻起源、分类和驯化规律进行了深入探讨，剖析了水稻核心种质资源的基因组遗传多样性；揭示了亚洲栽培稻的起源和群体基因组变异结构；建立了基于水稻基因组信息的数据库和应用平台，推进传统“经验育种”向现代“精准育种”跃升。具体成果包括：

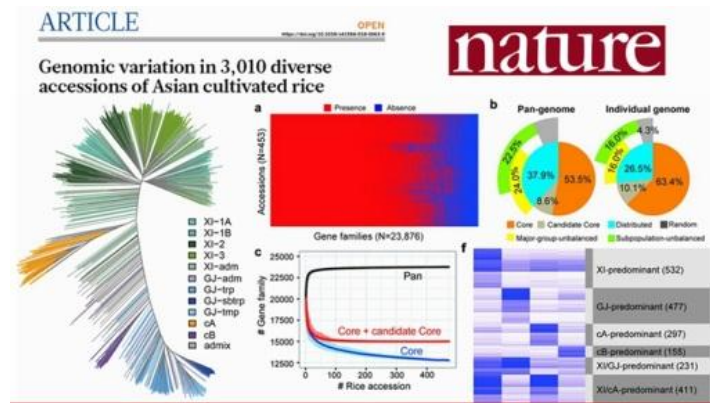
1) 检测到32M的高质量SNPs和Indels,对亚洲栽培稻群体进行了更为细致和准确的划分，由传统的5个亚群增加到9个。

2) 亚洲栽培稻中存在大量结构变异(SVs)，平均每个基因组约有1.2万个SVs，籼粳亚群间呈现明显的差异，并且可能是引起籼粳杂种不育和杂种衰退的遗传基础。

3) 使用“map-to-pan”策略构建了亚洲栽培稻的泛基因组，获得了268M的日本晴参考序列之外的非冗余序列，预测了1.2万个新的全长基因。

4) 通过对重要进化相关基因的单体型分析发现：很多的籼稻品种携带的等位基因没有出现在粳稻品种当中。该结果更支持籼稻是独立进化来的假说（多起源假说）。

5) 最后简单介绍了3000份水稻基因组的SNP和PAVs数据在关联分析挖掘水稻有利基因上的应用。



该论文还首次提出了粳、籼亚种的独立多起源假说，并恢复使用粳（*Oryza sativa* subsp. xian）、籼（*Oryza sativa* subsp. geng）亚种的正确命名，使中国源远流长的稻作文化得到正确认识和传承。

此项成果将推动水稻规模化基因发掘和水稻复杂性状分子改良，提升全球水稻基因组研究和分子育种水平，加快优质、高产、广适水稻新品种培育。该研究是国内外水稻研究专家大协作取得的成果，体现了中国在水稻基因组研究方面居于世界领先地位，并扩大了我国水稻功能基因组研究国际领先优势。在国家扩大开放的大背景下，重大数据共享和重大项目合作，将为我国乃至全球农业研究水平带来更大的飞跃。



Genomic variation in 3,010 diverse accessions of Asian cultivated rice
科学家完成亚洲栽培稻基因组变异研究

中国农业科学院深圳生物育种创新研究院黎志康研究员、阮珏研究员

深圳华大基因研究院张耕耘教授

上海交通大学韦朝春教授

2018年4月25日

<https://doi.org/10.1038/s41586-018-0063-9>

Here we analyse genetic variation, population structure and diversity among 3,010 diverse Asian cultivated rice (*Oryza sativa* L.) genomes from the 3,000 Rice Genomes Project. Our results are consistent with the five major groups previously recognized, but also suggest several unreported subpopulations that correlate with geographic location. We identified 29 million single nucleotide polymorphisms, 2.4 million small indels and over 90,000 structural variations that contribute to within- and between-population variation. Using pan-genome analyses, we identified more than 10,000 novel full-length protein-coding genes and a high number of presence-absence variations. The complex patterns of introgression observed in domestication genes are consistent with multiple independent rice domestication events. The public availability of data from the 3,000 Rice Genomes Project provides a resource for rice genomics research and breeding.