Genomic variation in 3,010 diverse accessions of Asian cultivated rice

Wensheng Wang, Ramil Mauleon, Zhiqiang Hu, Dmytro Chebotarov, Shuaishuai Tai, Zhichao Wu, Min Li, Tianqing Zheng, Roven Rommel Fuentes, Fan Zhang, Locusede Mansueto, Dario Copetti, Milicent Sanciangco, Kevin Christian Palis, Jianlong Xu, Chen Sun, Bining Fu, Hongliang Zhang, Yongming Gao, Xiuxin Zhao, Fei Shen, Xiao Cui, Hong Yu, Zichao Li, Miaolin Chen, Jeffrey Detras, Yongli Zhou, Xinyuan Zhang, Yue Zhao, Dave Kudrna, Chunchao Wang, Rui Li, Ben Jia, Jinyuan Lu, Xianchang He, Zhaotong Dong, Jiabao Xu, Yanhong Li, Miao Wang, Jianxin Shi, Jing Li, Dabing Zhang, Seunghyee Lee, Wushu Hu, Alexander Poliakov, Inna Dubchak, Victor Jun Ulat, Frances Nikki Borja, John Robert Mendoza, Jauhan Ali, Jing Li, Qiang Gao, Yongchao Niu, Zhen Yue, Ma. Elizabeth B. Naredo, Jayson Talag, Xueqiang Wang, Jinjie Li, Xiaodong Fang, Ye Yin, Jean-Christophe Glaszmann, Jianwei Zhang, Jayang Li, Ruaraith Sackville Hamilton, Rod A. Wing, Jue Ruan, Gengyun Zhang, Chaochun Wei, Nickolai Alexandrov, Kenneth L. McNally, Zhikang Li & Hei Leung

doi:10.1038/s41586-018-0063-9
Received:
27 November 2016
Accepted:
28 February 2018
Published online:
25 April 2018
http://www.nature.com/articles/s41586-018-0063-9

Abstract
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.