

Building two indica rice reference genomes with PacBio long-read and Illumina paired-end sequencing data

2016

Posted by : webmaster

Posted on : 2016/9/14 9:51:41

Jianwei Zhang, Ling-Ling Chen, Shuai Sun, Dave Kudrna, Dario Copetti, Weiming Li, Ting Mu, Wen-Biao Jiao, Feng Xing, Seunghee Lee, Jayson Talag, Jia-Ming Song, Bogu Du, Weibo Xie, Meizhong Luo, Carlos Ernesto Maldonado, Jose Luis Goicoechea, Lizhong Xiong, Changyin Wu, Yongzhong Xing, Dao-xiu Zhou, Sibin Yu, Yu Zhao, Gongwei Wang, Yeisoo Yu, Yijie Luo, Beatriz Elena Padilla Hurtado, Ann Danowitz, Rod A. Wing & Qifa Zhang

Scientific Data 3, Article number: 160076 (2016)

doi:10.1038/sdata.2016.76

Received: 22 April 2016

Accepted: 29 July 2016

Published online: 13 September 2016

Abstract

Over the past 30 years, we have performed many fundamental studies on two *Oryza sativa* subsp. *indica* varieties, Zhenshan 97 (ZS97) and Minghui 63 (MH63). To improve the resolution of many of these investigations, we generated two reference-quality reference genome assemblies using the most advanced sequencing technologies. Using PacBio SMRT technology, we produced over 108 (ZS97) and 174 (MH63) Gb of raw sequence data from 166 (ZS97) and 209 (MH63) pools of BAC clones, and generated ~97 (ZS97) and ~74 (MH63) Gb of paired-end whole-genome shotgun (WGS) sequence data with Illumina sequencing technology. With these data, we successfully assembled two platinum standard reference genomes that have been publicly released. Here we provide the full sets of raw data used to generate these two reference genome assemblies. These data sets can be used to test new programs for better genome assembly and annotation, aid in the discovery of new insights into genome structure, function, and evolution, and help to provide essential support to biological research in general.

<http://www.nature.com/articles/sdata201676>