

## The complete plastid genome of the wild rice species *Oryza brachyantha* (Poaceae)

Fengqi Liu<sup>a,b</sup>, Luke R. Tembrock<sup>c</sup>, Cheng Sun<sup>d</sup>, Guiqing Han<sup>a,b</sup>, Changhong Guo<sup>a</sup> and Zhiqiang Wu<sup>c</sup>

<sup>a</sup>Key Laboratory of Molecular Cytogenetics and Genetic Breeding of Heilongjiang Province, College of Life Science and Technology, Harbin Normal University, Harbin, China; <sup>b</sup>Institute of Pratacultural Science, Heilongjiang Academy of Agricultural Science, Harbin, China; <sup>c</sup>Department of Biology, Colorado State University, Fort Collins, CO, USA; <sup>d</sup>Key Laboratory of Pollinating Insect Biology of the Ministry of Agriculture, Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, China

### ABSTRACT

The whole plastid genome of wild rice (*Oryza brachyantha*) is characterized in this study. The genome is 134 604 bp in length and is arranged in a typical circular structure, including a pair of inverted repeats (IRs) of 20 832 bp in size separated by a large single-copy region (LSC) of 80 411 bp in length and a small single-copy region (SSC) of 12 529 bp in length. The overall GC content of the genome is 38.98%. One hundred and ten unique genes were annotated from the chloroplast genome, including 76 protein-coding genes, 4 ribosomal RNA genes and 30 tRNA genes. A total of 20 of these genes are duplicated in the IR regions, 13 genes contain 1 intron and 2 genes (*rps12* and *ycf3*) have 2 introns.

### ARTICLE HISTORY

Received 11 February 2016  
Accepted 13 February 2016

### KEYWORDS

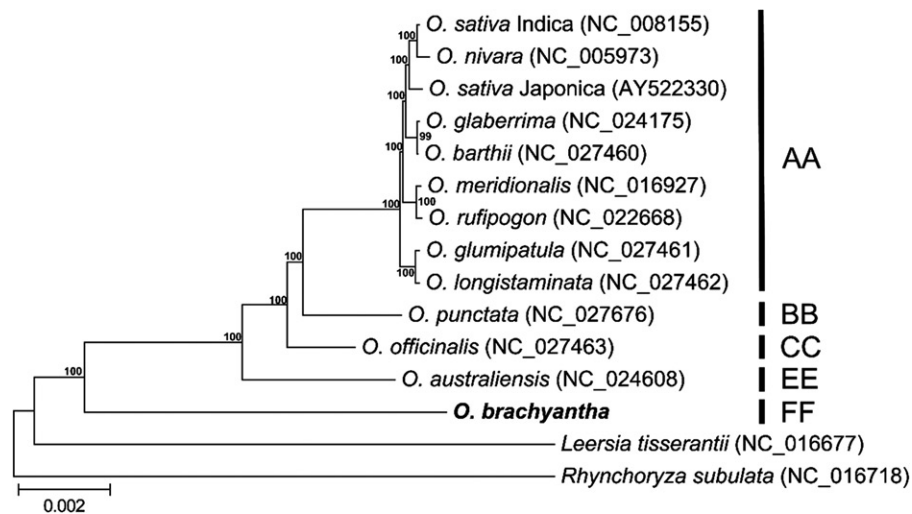
Next generation sequencing;  
*Oryza brachyantha*; plastid  
genome; rice

The genus *Oryza*, is an ideal model system for plant comparative genomics, as well as studies of evolutionary and functional biology (Lu et al. 2009). Within *Oryza*, two species are cultivated and the remaining ~22 wild species are wild. Ten distinct nuclear genome types, including six diploid (A, B, C, E, F and G) and four allotetraploid (BC, CD, HJ and HK), have been verified in this model genus (Ge et al. 1999). The cultivated rice has played an essential role in human nutrition and culture for millennia, and has important syntenic relationships with the other cereal species. As such, *Oryza* is among the most well-known and important study systems for Poaceae evolution and domestication (Lu et al. 2009). The wild relatives of *Oryza* have provided invaluable genomic resources for rice improvement and breeding (Dong et al. 2013). As the most distant wild relative of cultivated rice, *O. brachyantha* (F genome) is a basal lineage in *Oryza*, and may possess novel disease resistance and stress tolerance genes not found in more closely related species. Chen et al. (2013) found that *O. brachyantha* has the most compact genome within the genus *Oryza*, suggesting that a reduced genome is either plesiomorphic or genome reduction happened in this species since there is divergence from the most common recent ancestor of *Oryza*. Despite the numerous studies in *Oryza*, the complete plastid genome of *O. brachyantha* has not been assembled until now. By assembling the plastid genome of *O. brachyantha*, all plastid genome changes among *Oryza* species can now be examined. While plastomes are relatively small, they have supplied valuable informative markers in plant molecular systematics (Jansen et al. 2007; Wang et al. 2010; Wu & Ge 2012), biogeographical relationships among populations (Wang et al. 2011), plant DNA barcoding (Group CPBOL et al.

2011) and plant genetic transformation (Cui et al. 2011). In this study, we determined the whole plastid genome of *O. brachyantha* (Accession number from International Rice Research Institute as IRGC101232) by using next generation sequencing technology and Sanger sequencing.

By extracting the genome data (SRA046388) used for whole genome assembly from Chen et al. (2013), the trimmed and high-quality paired end (PE) reads were used to assemble the plastid genome from the CLC workbench (ver. 7.01 beta, CLC Inc., Aarhus, Denmark). We utilized reference mapping and *de novo* assembly methods to assemble the plastome by employing *O. australiensis* (NC\_024608) as reference following the method from Wu and Tembrock (2015). We also employed the Sanger sequencing to verify the final assembly. This draft sequence was validated and manually corrected by mapping the raw PE reads, back to the complete chloroplast genome of *O. brachyantha* (KT992850). Structural features and genes in the chloroplast genome were predicted following the method from Wu and Ge (2016).

The 134 604 bp long *O. brachyantha* plastome possesses a typical plastid double-stranded circular arrangement with a quadripartite structure common to most of the land plant plastomes (Raubeson & Jansen 2005; Wu & Ge 2012; Wu & Ge 2016). The large single copy (LSC) regions are 80 411 bp in length and the small single copy (SSC) regions are 12 529 bp in length separated by two inverted repeats (IRs) of 20 832 bp in size. The *O. brachyantha* plastome was found to contain 128 functional genes: of which 110 are in single copy and 18 are duplicated in the IR regions. Among the 110 annotated single-copy genes, 76 are protein-coding genes, 30 tRNA genes, and four rRNA genes, amounting to 44.19%, 2.13%,



**Figure 1.** NJ phylogenetic tree of *O. brachyantha* with other 14 species in the Oryzeae order based on complete plastome sequences. Numbers in the nodes are bootstrap values from 1000 replicates. The plastome of *Rhynchosyza subulata* (NC\_016718), a species from Zizaniinae subtribe, was set as an outgroup.

and 6.83% of the genome, respectively. The remaining 46.85% of the genome is made up of non-coding introns and intergenic spacers. Similar to other Poaceae plastomes (Wu & Ge 2012; Wu & Ge 2016), the *O. brachyantha* plastome is also AT-rich (61.02%). The four rRNA genes are all located in the IR regions. Twenty one tRNA genes are located in the single-copy region, whereas the other nine are located in the IR. Fifteen genes contain introns: *ycf3* contains two introns and the rest of the genes contain a single intron. *Rps12* is trans-spliced, with one of its exons in the LSC region (5' end) and the other two in the IR regions (3' end) separated by an intron. *NdhH* spans the IR and SSC junction. The *matK* gene was located within the intron of *trnK-UUU*. In the *O. brachyantha* plastome, the pairs of genes *atpB-atpE*, *psbC-psbD* and *ndhC-ndhK* had 4 bp, 53 bp and 10 bp overlapping regions, respectively.

Phylogenetic analysis, using the whole plastome alignment from 13 published *Oryza* species including *O. brachyantha* and two outgroup Oryzeae species, was performed using neighbour-joining (NJ) analysis in MEGA 6.0 (Tamura et al. 2013). Our phylogenetic analysis of plastomes divided the 13 *Oryza* species into the four predetermined genome groups (Figure 1). As expected, *O. brachyantha* was resolved as an early diverging lineage in *Oryza*. The phylogenetic relationships of the *Oryza* plastomes were identical to previous phylogenetic studies using nuclear genetic markers (Ge et al. 1999; Tang et al. 2015).

## Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## References

Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, Li B, Bai Z, Goicoechea JL, Liang C, et al. 2013. Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat Commun.* 4:1595.

Cui C, Song F, Tan Y, Zhou X, Zhao W, Ma F, Liu Y, Hussain J, Wang Y, Yang G, et al. 2011. Stable chloroplast transformation of immature scutella and inflorescences in wheat (*Triticum aestivum* L.). *Acta Biochim Biophys Sin (Shanghai).* 43:284–291.

Dong Z, Wang H, Dong Y, Wang Y, Liu W, Miao G, Liu X, Wang D, et al. 2013. Extensive microsatellite variation in rice induced by introgression from wild rice (*Zizania latifolia* Griseb.). *PLoS One.* 8:e62317.

Ge S, Sang T, Lu BR, Hong DY. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc Natl Acad Sci USA.* 96:14400–14405.

Group CPBOL, Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, et al. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc Natl Acad Sci USA.* 108:19641–19646.

Jansen RK, Cai Z, Raubeson LA, Daniell H, dePamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, et al. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA.* 104:19369–19374.

Lu F, Ammiraju JSS, Sanyal A, Zhang SL, Song RT, Chen JF, Li GS, Sui Y, Song X, Cheng Z, et al. 2009. Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. *Proc Natl Acad Sci USA.* 106:2071–2076.

Raubeson LA, Jansen RK. 2005. In: Henry R, editor. *Diversity and evolution of plants—genotypic and phenotypic variation in higher plants.* Wallingford, UK: CABI. p. 45–68.

Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.

Tang L, Zou X, Zhang L, Ge S. 2015. Multilocus species tree analyses resolve the ancient radiation of the subtribe Zizaniinae (Poaceae). *Mol Phylogenet Evol.* 84:232–239.

Wang L, Qi XP, Xiang QP, Heinrichs J, Schneider H, Zhang XC. 2010. Phylogeny of the paleotropical fern genus *Lepisorus* (Polypodiaceae, Polypodiopsida) inferred from four chloroplast DNA regions. *Mol Phylogenet Evol.* 54:211–225.

Wang L, Wu ZQ, Bystrakova N, Ansell SW, Xiang QP, Heinrichs J, Schneider H. 2011. Phylogeography of the Sino-Himalayan fern *Lepisorus clathratus* on “the roof of the world”. *PLoS One.* 6:e25896.

Wu ZQ, Ge S. 2012. The phylogeny of the BEP clade in grasses revisited: evidence from the whole-genome sequences of chloroplasts. *Mol Phylogenet Evol.* 62:573–578.

Wu ZQ, Ge S. 2016. The whole chloroplast genome of wild rice (*Oryza australiensis*). *Mitochondrial DNA.* 27:1062–1063.

Wu ZQ, Tembrock LR. 2015. Two complete chloroplast genomes of white campion (*Silene latifolia*) from male and female individuals. *Mitochondrial DNA.* DOI: 10.3109/19401736.2015.1126829.