C. R. Biologies 337 (2014) 499–502

Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com

Molecular biology and genetics/Biologie et génétique moléculaires

Corrected sequence of the wheat plastid genome

Ahmed Bahieldin [a,*,b], Magdy A. Al-Kordy [a,c], Ahmed M. Shokry [a,d], Nour O. Gadalla [a,c], Ahmed M.M. Al-Hejin [a], Jamal S.M. Sabir [a], Sabah M. Hassan [a,b], Ahlam A. Al-Ahmadi [a], Erika N. Schwarz [e], Hala F. Eissa [f,g], Fotouh M. El-Domyati [a,b], Robert K. Jansen [a,e]

[a] Department of Biological Sciences, Faculty of Science, King Abdulaziz University (KAU), P.O. Box 80141, Jeddah 21589, Saudi Arabia
[b] Department of Genetics, Faculty of Agriculture, Ain Shams University, Cairo, Egypt
[c] Genetics and Cytology Department, Genetic Engineering and Biotechnology Division, National Research Center, Dokki, Egypt
[d] Agricultural Genetic Engineering Research Institute (AGERI), Agriculture Research Center (ARC), Giza, Egypt
[e] Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA
[f] Agricultural Genetic Engineering Research Institute (AGERI), Agriculture Research Center (ARC), Giza, Egypt
[g] Faculty of Biotechnology, Misr University for Science and Technology (MUST), 6th October city, Egypt

ARTICLE INFO

Article history:
Received 24 April 2014
Accepted after revision 9 July 2014
Available online 5 August 2014

Keywords:
Next-Generation sequencing
Rice plastid genome
GenBank

ABSTRACT

Wheat is the most important cereal in the world in terms of acreage and productivity. We sequenced and assembled the plastid genome of one Egyptian wheat cultivar using next-generation sequence data. The size of the plastid genome is 133,873 bp, which is 672 bp smaller than the published plastid genome of "Chinese Spring" cultivar, due mainly to the presence of three sequences from the rice plastid genome. The difference in size between the previously published wheat plastid genome and the sequence reported here is due to contamination of the published genome with rice plastid DNA, most of which is present in three sequences of 332, 131 and 131 bp. The corrected plastid genome of wheat has been submitted to GenBank (accession number KJ592713) and can be used in future comparisons.

1. Introduction

Wheat is considered one of the most widely cultivated and consumed food crops in the world. Cultivated wheats are either hexaploid (*Triticum aestivum*, AABBDD, 2n = 6x) or tetraploid (*Triticum durum* or *Triticum turgidum* subsp. *durum*, AABB, 2n = 4x). This genomic complexity makes it difficult to accurately sequence and assemble the entire nuclear genome. The draft genome of the A-genome progenitor species (*Triticum urartu*, AA) has been assembled and assigned as the diploid reference for further analysis of polyploid nuclear wheat genomes [1]. The available reference plastid genome of the hexaploid "Chinese Spring" cultivar was completed by Sanger sequencing of a set of cloned restriction fragments that covered the entire genome [2]. As part of a project to examine plastid single nucleotide polymorphisms (SNPs) among nine wheat cultivars from Egypt, we sequenced the complete genome and discovered that the published wheat genome contains contaminated sequence from the rice plastid genome. In this paper, we characterize the corrected plastid genome sequence for one cultivar of wheat from Egypt.

* Corresponding author.
E-mail address: bahieldin55@gmail.com (A. Bahieldin).

http://dx.doi.org/10.1016/j.crvi.2014.07.001
1631-0691/

2. Materials and methods

2.1. DNA Isolation, genome sequencing and mapping of reads to reference plastid genome

Total genomic DNA was extracted from leaf tissues (~1 g) of 14-day-old etiolated seedlings of one hexaploid wheat cultivar (Giza 168, Delta, Egypt) using the modified procedure of [3]. Purified total genomic DNA was sent to Beijing Genomics Institute (BGI), Shenzhen, China for sequencing using the Illumina HiSeq 2000 platform. Thirty million 100-bp paired-end reads were generated from a sequencing library with 500-bp inserts. Adapter sequences in reads of the raw data were deleted, and reads with 50% low quality bases (quality value $\leq$ 5) or more were removed. The remaining sequences were mapped to the published wheat plastid genome (accession number NC_002762) using CLC Genomics Workbench (version 3.0, http://www.clcbio.com/usermanuals). The raw sequence reads (SRA XXXX) and the assembled and annotated plastid genome sequence of cultivar Giza 168 (accession number KJ592713) were deposited in NCBI. Annotation of the plastid genome was performed using DOGMA [4] supplemented with tRNAscan (http://lowelab.ucsc.edu/tRNAscan-SE/) and ARAGORN (http://mbioserv2.mbioekol.lu.se/ARAGORN) for tRNAs. A circular genome map was constructed with GenomeVx [5].

3. Results and discussion

Raw reads were mapped to the reference wheat plastid genome (accession number NC_002762). The number of reads mapped was 1,195,172, which represents 1.1% of the total reads. The read depth averaged 1,229X coverage across the genome. The assembled plastid genome of the Giza 168 wheat cultivar is 133,873 bp, which is 672-bp smaller than the published genome for the "Chinese Spring" cultivar [2]. Mapping of the Giza 168 genome to the "Chinese Spring" genome identified three DNA sequences of 332, 131 and 131 bp that are absent from the Giza 168 cultivar (Fig. 1). Other shorter sequences were also found only in the published reference plastid genome. The 332, 131 and 131 bp DNA sequences are located at positions 6,164-6,495, 83,918-84,048 and 130,960-131,090 of the reference genome, respectively. Alignment of these two genomes generated gaps within the plastid genome sequence of the Giza 168 cultivar (Fig. 1). BlastN analyses of the three extra sequences from the wheat reference plastid genome to the NCBI database indicated 100% sequence identity of these fragments to plastid genome of rice (*Oryza sativa*, Japonica group) as well as the published "Chinese Spring" reference wheat genome (Table 1). The next best Blast hits were to another rice species, *Oryza rufipogon* with 99% identity. Blast hits to plastid genomes of other cereals were not detected. To further confirm that the published wheat plastid genome sequence is contaminated with rice sequence, the sequences from the wheat GZ168 cultivar flanking the three gaps were blasted to the NCBI database (Table 2). These sequences are located in the plastid genome of cultivar GZ168 between 5,957-6,288 (332 bp), 83,385-83,515 (131 bp) and 130,192-130,322 (131 bp) bases (Fig. 1). The results indicated 100% sequence identity to plastid genome sequences of hexaploid wheat and other members of the Triticeae, while no Blast hits were detected to any of the rice plastid genome sequences available in the GenBank. Annotation of the corrected wheat plastid genome confirmed the gene content and order from the published genome (Fig. 2).

The cause of the contamination of the wheat plastid genome sequence with rice DNA is unknown, however, other cases of errors in published plastid genome sequences have been detected in the past. For example, in the plastid genome of tobacco [6], sequencing errors of 119 bp were reported seven years after it was published [7]. In the case of tobacco, 90 bp were missed because a small *Alu*I restriction fragment was missed in the cloning

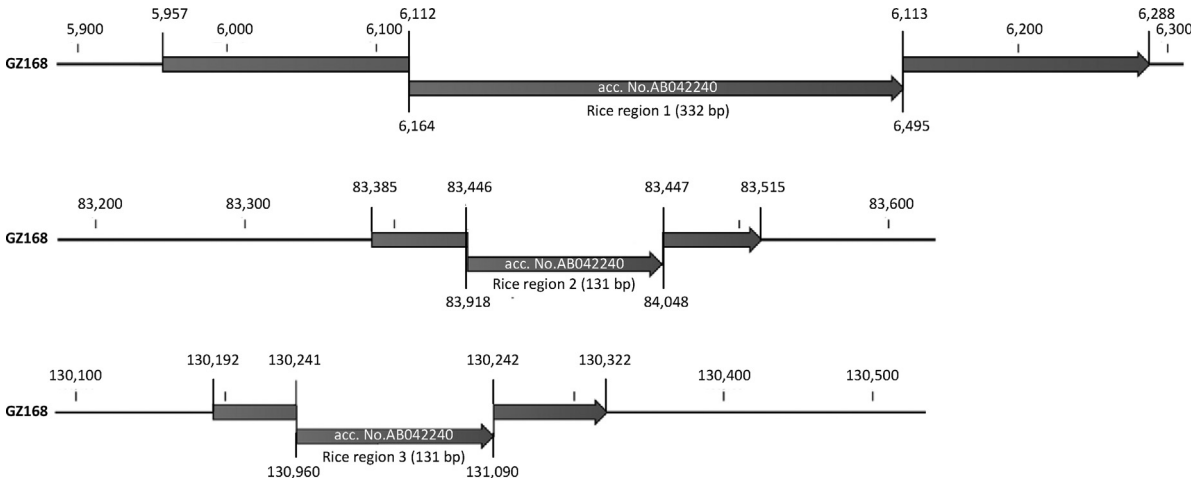

Fig. 1. Extra sequences present in the published wheat plastid genome (accession number NC_002762 = AB042240, 134,545 bp) missing in the Egyptian wheat cultivar (GZ168, accession number KJ592713, 133,873 bp). Numbers above and below maps indicate coordinates in KJ592713 and NC_002762, respectively.