Improved and robust method to efficiently deplete repetitive elements from complex plant genomes†

H. Ichida*[1] and T. Abe*[1]

Wheat is one of the most important staple food crops around the world, occupying 17% (one-sixth) of the global crop acreage and serving as a food source for approximately 40% of the world's population. Despite the availability of a high-quality reference genome sequence, whole genome re-sequencing of wheat is still a major challenge owing to its large and complex genome. Wheat has an estimated haploid genome size of 15.4–15.8 Gb, of which 84.7% is made up of repetitive sequences. To successfully apply genome sequencing to practical agricultural breeding situations, the establishment of an effective strategy to eliminate highly abundant repetitive elements within the genome is required.

Duplex-specific nuclease (DSN) provides an enzymatic method to reduce the whole-genome redundancy and normalization. However, this method has not been widely used in the genome analysis of agricultural crops owing to its sensitivity to reaction conditions and difficulty in DNA recovery after the treatment. We developed an improved design for Illumina-compatible sequencing adapters that avoids duplex formation within a typical annealing temperature range (∼68°C) and accidental degradation by the DSN enzyme. The newly designed adapter includes the 21 bp sequences immediately adjacent to the original TruSeq adapter and forms a hairpin-like structure by placing the necessary adapter sequences in a complementary orientation at the 5' and 3' ends of a single oligonucleotide at room temperature during library preparation. Using this design, all necessary sequence elements for Illumina sequencing could then be added using full-length adapter primers during the recovery PCR, which generates a library with exactly the same structure as the original TruSeq libraries.

We compared the relative abundance of 18S and 25S rDNA elements in genomic DNA and the DNA libraries before and after DSN treatment (DSN^- and DSN^+, respectively) in rice and diploid, tetraploid, and hexaploid wheat cultivars by quantitative PCR. In summary, the relative abundances of 18S and 25S rDNA were reduced to 1.15% and 3.54% of the DSN^- samples in rice. The effectiveness of the DSN treatment was essentially the same among the three wheat genomes. In particular, the relative abundance of 18S rDNA was decreased to 7.13%, 4.95%, and 4.33% considering the DSN^- in diploid, tetraploid, and hexaploid cultivars, respectively. The relative abundance of 25S rDNA was also decreased to 21.73%, 8.96%, and 7.54% considering DSN^-. These results indicate that the DSN treatment effectively depleted repetitive sequences in different wheat cultivars, which have much larger and more complex genomes compared with rice.

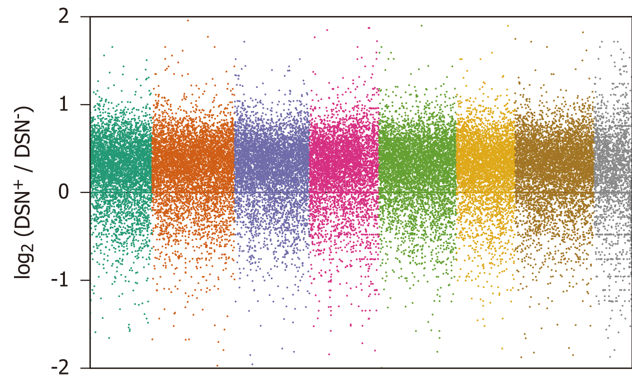

Fig. 1. Comparison of read counts before and after duplex-specific nuclease-based repeat depletion in gene regions. Reprinted from Plant Science, https://doi.org/10.1016/j.plantsci.2018.10.021, with permission from Elsevier.

We evaluated the enzymatic depletion of highly repetitive elements by DSN treatment in an actual Illumina sequencing run using the diploid wheat cultivar KU104-1, which has an estimated genome size of 6.1 Gb, as a model. The sequencing reads mapped on the regions encoding Ty1-Copia and Ty3-Gypsy were decreased by 61.6% and 33.4%, respectively, compared with the DSN^- samples. DNA-type repetitive elements were also successfully depleted by the DSN treatment. In particular, reads mapped to DNA-type transposons and simple sequence repeats (SSRs) were respectively reduced by 17.1% and 9.0% compared with the untreated control. Therefore, we conclude that this method should be useful in a broad range of species in which genomic approaches are not currently applicable owing to their large and complex genome structures and the consequent high expenses of sequencing and analysis.

† Condensed from the article in Plant Science, 280, 455-460 (2019)

*1 RIKEN Nishina Center