Title: Whole-genome sequence analysis of mutations in rice plants regenerated from zygotes, mature embryos, and immature embryos

Authors: Masako Ichikawa^{1*}, Norio Kato¹²⁴, Erika Toda²⁴, Masakazu Kashiara¹, Yuji Ishida¹, Yukoh Hiei¹, Sachiko N. Isobe³, Kenta Shirasawa³, Hideki Hirakawa³, Takashi Okamoto²⁴, and Toshihiko Komari¹

Affiliation: ¹Plant Innovation Center, Japan Tobacco Inc., Iwata, Shizuoka, JAPAN
²RIKEN Cluster for Science, Technology and Innovation Hub, Yokohama, JAPAN
³Kazusa DNA Research Institute, Kisarazu, Chiba, JAPAN
⁴Department of Biological Sciences, Tokyo Metropolitan University, Hachioji, JAPAN

* Correspondence

Correspondence Author:

Masako Ichikawa

Plant Innovation Center, Japan Tobacco Inc.,

700 Higashibara, Iwata, Shizuoka 438-0802, JAPAN

E-mail: masako.ichikawa@jt.com, Telephone: +81-538-32-7120, Fax: +81-538-33-6046

Running title: Sequencing of rice from cultured zygotes and embryos

Abbreviations: SNV: single nucleotide variant, indels: insertions and deletions

Abstract

Somaclonal variation was studied by whole-genome sequencing in rice plants (*Oryza sativa* L., 'Nipponbare') regenerated from the zygotes, mature embryos, and immature embryos of a single mother plant. The mother plant and its seed-propagated progeny were also sequenced. A total of 338 variants of the mother plant sequence were detected in the progeny, and mean values ranged from 9.0 of the seed-propagated plants to 37.4 of regenerants from mature embryos. The ratio of single nucleotide variants among the variants was 74.3%, and the natural mutation rate calculated using the variants in the seed-propagated plants was 1.2×10^{-8} . The percentage and the mutation rate were consistent with the values reported previously. Plants regenerated from mature embryos had significantly more variants than different progeny types. Therefore, using zygotes and immature embryos can reduce somaclonal variation during the genetic manipulation of rice.

Keywords

Oryza sativa L., somaclonal variation, sequencing, zygote culture, callus

Cell culture enables foreign gene transfer and gene editing in plant science. Genetic modifications are first introduced into a single cell, and then plants are regenerated from culturing that modified cell. However, previous research has shown that heritable changes, known as somaclonal variation (Nishi et al. 1968, Heinz and Mee 1969, Larkin and Scowcroft 1981), frequently occur in regenerants of dedifferentiated cells. The characteristics and induction mechanisms of somaclonal variation are poorly understood. Some variation is attributed to differences in DNA methylation levels between cells that propagate *in vivo* and

in vitro. For example, mutations caused by transposable elements are normally suppressed *in vivo* by DNA methylation but are induced in cultured cells (Hirochika et al. 1996). Prolonged culturing leads to higher mutation rates (Qin et al. 2018), and so it is typically avoided in gene transfer, editing, and other cell engineering practices. However, when somaclonal mutations are desired (e.g., to introduce more variation in breeding populations), prolonged culturing can be useful (Larkin and Scowcroft 1981).

Studying genetic variation in regenerants from cell cultures, transgenic plants, and edited plants by whole-genome sequencing has become feasible because of recent DNA sequencing technology advancements. Rice (*Oryza sativa* L.) is one of the most frequently studied plant species. Somaclonal variation in rice plants, including transgenic and edited plants, has been analyzed by whole-genome sequencing (Miyao et al. 2012, Kawakatsu et al. 2013, Zhang et al. 2014, Wei et al. 2016, Qing et al. 2018, Tang et al. 2018, and Park et al. 2019). Such studies detected numerous mutations that were induced by cell culture. Varietal differences in somaclonal variation in rice have also been observed with amplified fragment length polymorphism (AFLP) analysis (Wang et al. 2013).

One limitation of the studies cited above is that the sequenced plants were regenerated solely from mature embryos. Other cell and tissue sources should be investigated to gain a better understanding of somaclonal variation. For example, immature embryos may be more efficient for genetic transformation in cereal plants (Hiei et al. 2014). More efficient gene editing has been reported in rice zygotes (Toda et al. 2019). Another limitation of the studies cited above is that cells were cultured for more than two months in most cases, although a shorter culture period is typically recommended. Toki (1997) described a rice transformation protocol that requires only one month to grow calli from mature seeds. Therefore, in this study, rice plants were regenerated from zygotes, mature embryos, and

immature embryos using the minimal times required to achieve callus induction and growth. The mutation rates of regenerants were compared by whole-genome resequencing.

The process for obtaining regenerated plant samples is shown in Figure 1. A mother plant was cultivated in a greenhouse from a single seed of the rice cultivar 'Nipponbare' obtained from a commercial distributor. The following samples were collected from the mother plant and used immediately: young spikelets (before flowering), immature seeds (12 days after pollination), and mature seeds.

Egg and sperm cells were isolated from young spikelets and electrically fused to produce zygotes, as described by Uchiumi et al. (2006, 2007). The zygotes were cultured and regenerated into plantlets (Uchiumi et al. 2007). Husking and sterilizing mature seeds, preparing immature embryos from immature seeds, and making culture media to induce calli and regenerate plants from mature and immature embryos were performed as previously described (Hiei et al. 2008). Callus induction and cell culture from zygotes, immature embryos, and mature embryos using media that contains 2, 4-dichlorophenoxy-acetic acid (2, 4-D) were conducted for 4, 2, and 2 weeks, respectively. Periods for the plant regeneration cultures without 2, 4-D from the cells from zygotes, immature embryos, and mature embryos were 3, 2, and 2 weeks, respectively. Regenerants were grown in a greenhouse with seed-propagated plants.

The plants shown in Figure 1 were analyzed by whole-genome resequencing. No abnormalities were visually observed in greenhouse-grown plants. Approximately 1 g of young leaf segments were sampled from the mother plant and four progeny groups: (i) five regenerants from five independent zygotes, (ii) five regenerants from five independent immature embryos, (iii) five regenerants from five independent mature embryos, and (iv) five seed-propagated plants. The leaves were frozen and stored at -80°C before DNA preparation.

Total DNA was extracted from the young leaves of 21 plants using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). DNA quality was analyzed using NanoDrop™ and Qubit™ (Thermo Scientific, Waltham, Massachusetts, US). Illumina paired-end (PE) libraries were constructed for each plant with a TruSeq DNA PCR-Free Sample Prep Kit (Illumina, San Diego, CA). Illumina HiSeq X with a read length of 151 nucleotides (nt) was used to obtain genome sequences. Sequenced bases with quality scores less than 10 were filtered by PRINSEQ 0.20.4 (Schmieder and Edwards, 2011). Adaptor sequences in the reads were trimmed using FASTX-Clipper from the FASTX-Toolkit 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit).

A total of 3,066,032,798 reads were obtained from the 21 Illumina PE libraries. The average number of obtained reads per line was 146,001,562 (range: 59,349,518 to 188,539,542). The total lengths of the obtained reads per plant are shown in Figure 2. The Illumina PE reads were mapped onto the 'Nipponbare' reference genome (Os-Nipponbare-Reference-IRGSP-1.0, Kawahara et al. 2013) to detect base variances (SNVs and indels) by Bowtie2 (Langmead and Salzberg, 2012). Variant calls were performed by bcftools 1.9 in SAMtools (Li et al. 2009). The parameters used for variant calls were `bcftools-1.9/bin/bcftools mpileup -d 10000000 -Ou -b bam-list.txt -f IRGSP-1.0_genome.fasta -a DP,AD,INFO/AD | bcftools-1.9/bin/bcftools call -cv -Ov -f GQ -o output.vcf`.

Variants of the mother plant were filtered to remove false positives using the following criteria: minimum depth = 10; maximum depth = 65; minimum quality = 999; variants at positions where bases were not called in all plants were excluded; variants that were detected in two or more plants at the same position were excluded. The number of variants of the mother plant in each progeny group was statistically compared, assuming Poisson distributions. Namely, when two groups were compared, the variant numbers in the

two groups were hypothesized to be samples from the same Poisson distribution, and then the probability for such a combination of Poisson samples was calculated to test the hypothesis. Mutation rates per rice diploid genome were calculated by dividing the number of variants by $2 \times 390 \times 10^6$, as shown in Tang et al. (2018).

A total of 338 variants of the mother plant were detected in the progeny groups, and the number of the variants in each progeny group is shown in Figure 3. Mean values ranged from 9.0 of the seed-propagated plants to 37.4 of regenerants from mature embryos. The ratio of single nucleotide variants among the variants of the mother plant was 74.3%. This percentage is close to those reported by Zhang et al. (2014) and Tang et al. (2018). The natural mutation rate calculated using the variants in the seed-propagated plants was 1.2×10^{-8} , which is consistent with the natural mutation rates of 5.4×10^{-8} for rice (Tang et al. 2018), 0.7×10^{-8} for Arabidopsis (Ossowski et al. 2010), and $2.2\text{--}3.9 \times 10^{-8}$ for maize (Yang et al. 2017).

More variants were detected in regenerants from mature embryos than in regenerants from immature embryos or zygotes. The numbers of variants detected in regenerants from immature embryos and zygotes were similar to those detected in seed-propagated plants. These results suggest that the starting tissue affects mutation rates. The regenerants from zygotes were cultured longer than regenerants from other tissues because callus formation from single cells takes longer. Despite its extended duration, the process of culturing cells from zygotes was not more mutagenic than culturing cells from other tissues. The difference in the number of variants between regenerants cultured from mature and immature embryos was unexpected because calli induced from these tissues are similar in appearance. A callus is a mass of dedifferentiated cells and is generally believed to have lost the characteristics of the original tissue. More detailed comparisons, such as analyzing patterns of transcription and DNA methylation in calli from different tissues, may provide useful insight into the nature of somaclonal variation.

Mature seeds are commonly used and are a convenient starting tissue for cell culture because they can be obtained in large quantities and stored at room temperature for several months before culturing. In contrast, cell culture protocols from zygotes and immature embryos require a constant supply of fresh materials from healthy plants grown in well-conditioned greenhouses. However, the results of this study suggest that the use of immature embryos and zygotes should be added to the general guidance of shorter cell culture periods to reduce somaclonal variation during the genetic manipulation of rice.

Acknowledgments

The authors thank Naoki Takemori for his support and helpful discussions and Chitose Matsui and Natsuki Hayashi for their technical assistance.

Conflict of Interest Statement

Erika Toda, Sachiko N. Isobe, Kenta Shirasawa, Hideki Hirakawa and Takashi Okamoto declare that they have no conflict of interest. Masako Ichikawa, Norio Kato, Masakazu Kashihara, Yuji Ishida, Yukoh Hiei, and Toshihiko Komari were affiliated with Japan Tobacco Inc., which was a funder of this study.

References

- Heinz DJ, Mee GWP (1969) Plant differentiation from callus tissue of *Saccharum* species. Crop Sci 9: 346-348
- Hiei Y, Ishida Y, Komari T (2014) Progress of cereal transformation technology mediated by

- Agrobacterium tumefaciens*. Front Plant Sci 5: 628
- Hiei Y, Komari T (2008) *Agrobacterium*-mediated transformation of rice using immature embryos or calli induced from mature seed. Nat Protoc 3: 824-834
- Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci U S A 93: 7783-7788
- International Rice Genome Sequencing P (2005) The map-based sequence of the rice genome. Nature 436: 793-800
- Jupe F, Rivkin AC, Michael TP, Zander M, Motley ST, Sandoval JP, Slotkin RK, Chen H, Castanon R, Nery JR, et al. (2019) The complex architecture and epigenomic impact of plant T-DNA insertions. PLoS Genet 15: e1007819
- Kawahara Y, De La Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, et al. (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice 6: 4
- Kwakatsu T, Kawahara Y, Itoh T, Takaiwa F (2013) A whole-genome analysis of a transgenic rice seed-based edible vaccine against cedar pollen allergy. DNA Res 20: 623-631
- Koboldt DC, Zhang Q, Larson DE, Shen D, Mclellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22: 568-576
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9: 357-359
- Larkin PJ, Scowcroft WR (1981) Somaclonal variation - a novel source of variability from cell cultures for plant improvement. Theor Appl Genet 60: 197-214

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079
- Miyao A, Nakagome M, Ohnuma T, Yamagata H, Kanamori H, Katayose Y, Takahashi A, Matsumoto T, Hirochika H (2012) Molecular spectrum of somaclonal variation in regenerated rice revealed by whole-genome sequencing. *Plant Cell Physiol* 53: 256-264
- Nishi T, Yamada Y, Takahashi E (1968) Organ redifferentiation and plant restoration in rice callus. *Nature* 219: 508-509
- Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327: 92-94
- Park D, Park SH, Kim YS, Choi BS, Kim JK, Kim NS, Choi IY (2019) NGS sequencing reveals that many of the genetic variations in transgenic rice plants match the variations found in natural rice population. *Genes Genomics* 41: 213-222
- Qin Y, Shin K-S, Woo H-J, Lim M-H (2018) Genomic variations of rice regenerants from tissue culture revealed by whole genome re-sequencing. *Plant Breed Biotech* 6: 426-433
- Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863-864
- Tang X, Liu G, Zhou J, Ren Q, You Q, Tian L, Xin X, Zhong Z, Liu B, Zheng X, et al. (2018) A large-scale whole-genome sequencing analysis reveals highly specific genome editing by both Cas9 and Cpf1 (Cas12a) nucleases in rice. *Genome Biol* 19: 84
- Toda E, Koiso N, Takebayashi A, Ichikawa M, Kiba T, Osakabe K, Osakabe Y, Sakakibara H, Kato N, Okamoto T (2019) An efficient DNA- and selectable-marker-free genome-editing system using zygotes in rice. *Nat Plants* 5: 363-368
- Toki S (1997) Rapid and efficient *Agrobacterium*-mediated transformation in rice. *Plant Mol*

Biol Rep 15: 16-21

Uchiumi T, Komatsu S, Koshihara T, Okamoto T (2006) Isolation of gametes and central cells from *Oryza sativa* L. Sex Plant Reprod 19: 37-45

Uchiumi T, Uemura I, Okamoto T (2007) Establishment of an in vitro fertilization system in rice (*Oryza sativa* L.). Planta 226: 581-589

Wang X, Wu R, Lin X, Bai Y, Song C, Yu X, Xu C, Zhao N, Dong Y, Liu B (2013) Tissue culture-induced genetic and epigenetic alterations in rice pure-lines, F1 hybrids and polyploids. BMC Plant Biol 13: 77

Wei FJ, Kuang LY, Oung HM, Cheng SY, Wu HP, Huang LT, Tseng YT, Chiou WY, Hsieh-Feng V, Chung CH, et al. (2016) Somaclonal variation does not preclude the use of rice transformants for genetic screening. Plant J 85: 648-59

Yang N, Xu XW, Wang RR, Peng WL, Cai L, Song JM, Li W, Luo X, Niu L, Wang Y, et al. (2017) Contributions of *Zea mays* subspecies *mexicana* haplotypes to modern maize. Nat Commun 8: 1874

Zhang D, Wang Z, Wang N, Gao Y, Liu Y, Wu Y, Bai Y, Zhang Z, Lin X, Dong Y, et al. (2014) Tissue culture-induced heritable genomic variation in rice, and their phenotypic implications. PLoS One 9: e96879

Figure legends

Figure 1

Sampling design for whole-genome sequencing. DNA was isolated from 5 plants derived from 5 different spikelets/seeds in each progeny group.

Figure 2

Total length of sequence reads per plant.

Figure 3

Number of variants

** Significantly different from the other three progeny groups ($P < 0.01$).

Figure 1. Sampling design for whole-genome sequencing. DNA was isolated from 5 plants derived from 5 different spikelets/seeds in each progeny group.

Figure 2. Total length of sequence reads per plant.

Figure 3. Number of variants of the mother plant. ** Significantly different from the other three progeny groups ($P < 0.01$).