



Optimized Method of Extracting Rice Chloroplast DNA for High-Quality Plastome Resequencing and *de Novo* Assembly

Takeshi Takamatsu^{1,2}, Marouane Baslam², Takuya Inomata¹, Kazusato Oikawa^{2†}, Kimiko Itoh^{1,2}, Takayuki Ohnishi³, Tetsu Kinoshita⁴ and Toshiaki Mitsui^{1,2*}

¹ Department of Life and Food Sciences, Graduate School of Science and Technology, Niigata University, Niigata, Japan, ² Laboratory of Biochemistry, Faculty of Agriculture, Niigata University, Niigata, Japan, ³ Center for Education and Research of Community Collaboration, Utsunomiya University, Utsunomiya, Japan, ⁴ Kihara Institute for Biological Research, Yokohama City University, Yokohama, Japan

OPEN ACCESS

Edited by:

Roger Deal,
Emory University, United States

Reviewed by:

Masato Nakai,
Osaka University, Japan
Aureliano Bombarely,
Virginia Tech, United States

*Correspondence:

Toshiaki Mitsui
t.mitsui@agr.niigata-u.ac.jp

† Present address:

Kazusato Oikawa,
Enzyme Research Team, Center
for Sustainable Resource Science,
RIKEN, Wako-shi, Japan

Specialty section:

This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: 06 December 2017

Accepted: 14 February 2018

Published: 28 February 2018

Citation:

Takamatsu T, Baslam M, Inomata T,
Oikawa K, Itoh K, Ohnishi T,
Kinoshita T and Mitsui T (2018)
Optimized Method of Extracting Rice
Chloroplast DNA for High-Quality
Plastome Resequencing and *de*
Novo Assembly.
Front. Plant Sci. 9:266.
doi: 10.3389/fpls.2018.00266

Chloroplasts, which perform photosynthesis, are one of the most important organelles in green plants and algae. Chloroplasts maintain an independent genome that includes important genes encoding their photosynthetic machinery and various housekeeping functions. Owing to its non-recombinant nature, low mutation rates, and uniparental inheritance, the chloroplast genome (plastome) can give insights into plant evolution and ecology and in the development of biotechnological and breeding applications. However, efficient methods to obtain high-quality chloroplast DNA (cpDNA) are currently not available, impeding powerful sequencing and further functional genomics research. To investigate effects on rice chloroplast genome quality, we compared cpDNA extraction by three extraction protocols: liquid nitrogen coupled with sucrose density gradient centrifugation, high-salt buffer, and Percoll gradient centrifugation. The liquid nitrogen–sucrose gradient method gave a high yield of high-quality cpDNA with reliable purity. The cpDNA isolated by this technique was evaluated, resequenced, and assembled *de novo* to build a robust framework for genomic and genetic studies. Comparison of this high-purity cpDNA with total DNAs revealed the read coverage of the sequenced regions; next-generation sequencing data showed that the high-quality cpDNA eliminated noise derived from contamination by nuclear and mitochondrial DNA, which frequently occurs in total DNA. The assembly process produced highly accurate, long contigs. We summarize the extent to which this improved method of isolating cpDNA from rice can provide practical progress in overcoming challenges related to chloroplast genomes and in further exploring the development of new sequencing technologies.

Keywords: *de novo* assembly, chloroplast DNA, next-generation sequencing, plastid genome, NUPTs, MTPTs, *Oryza sativa*

Abbreviations: BAM, binary alignment map; cpDNA, chloroplast DNA; HS, high salt method; InDel, small insertion/deletion; IRs, inverted repeats; LN, liquid nitrogen coupled with sucrose gradient centrifugation method; LSC, large single-copy region; Mt, mitochondria; MTPTs, mitochondrial plastid like sequences; NGS, next-generation sequencing; Pt, plastid; NUPTs, nuclear plastid DNA; PG, Percoll gradient centrifugation method; SNP, single-nucleotide polymorphism; SSC, single copy region.

INTRODUCTION

Chloroplasts, which are important cellular organelles that provide energy to plants, have an independent, circular, double-stranded DNA. The chloroplast genome (plastome), which ranges in size from 110 to 200 kb, consists of a pair of IRs, a LSC, and a SSC. The first whole-chloroplast-genome sequencing of rice (Hirai et al., 1985; Hiratsuka et al., 1989), *Arabidopsis* (Sato et al., 1999), and maize (Maier et al., 1995), based on Sanger technology, provided a basic understanding of the genome's structure and function. NGS technology now offers the promise to further the development of chloroplast genome studies, and has accelerated analyses to the point where more than 1,600 accessions of complete chloroplast genome sequences from land plants are available in public databases¹. The advance of high-throughput, high-resolution analyses has facilitated population genetics and evolutionary studies focused on the chloroplast genome (Morris et al., 2011; Wu et al., 2015; Tong et al., 2016). Organelle genomes have less diversity and exert less influence on phenotype than the nuclear genome (Wolfe et al., 1987; Kahlau et al., 2006; Drouin et al., 2008), although some studies have shown that variation in organelle genomes can influence variation in phenotypes (Moison et al., 2010; Joseph et al., 2013; Tang et al., 2013). Joseph et al. (2013) demonstrated that the cytoplasmic genome plays a central role in controlling natural variations in metabolomic networks within a reciprocal *Arabidopsis* Kas × Tsu recombinant inbred line population. Roux et al. (2016) reported cyto-nuclear co-adaptation by creating a unique series of 56 cytokines resulting from cytoplasmic substitutions among eight natural *Arabidopsis* species. More recent widespread reports show that interactions between nuclear and cytoplasmic genomes shape natural variation (Greiner and Bock, 2013; Sloan, 2015; Roux et al., 2016). This factor gives rise to new opportunities for using organelle genomes as novel breeding targets (Maliga, 2001; Wang et al., 2008). The plastome thus presents an attractive target for genome engineering and a promising alternative to nuclear transformation (Olejniczak et al., 2016). Indeed, chloroplast genomic studies have crucial translational and biotechnological applications owing to the genome's ability to express > 120 foreign genes from different organisms (Daniell et al., 2016). Accordingly, the efficient sequencing of plastid genomes, which requires highly purified plastid DNA, will allow the production of transplastomic plants (Diekmann et al., 2008).

Rice, one of the most important food crops in the world, has important syntenic relationships with the other cereal species and is a model for monocots and grasses. The chloroplast genome of *Oryza sativa* is 134,525 base pairs long, with 159 unique genes, including 38 tRNA, 8 rRNA, and 108 protein-coding genes (Hiratsuka et al., 1989). The evolutionary transfer of plastid DNA fragments to the nuclear and mitochondrial genomes is frequently found in plants (Lewin, 1984; Martin and Herrmann, 1998; Matsuo et al., 2005). Such transfers to the nuclear genome

(nuclear plastid DNA, NUPTs) and to the mitochondrial genome (mitochondrial plastid-like sequences, MTPTs) are more abundant in rice than in other higher plants (Yoshida et al., 2014). As plastome sequencing is frequently based on total DNA, this might reduce the mapping accuracy owing to the difficulty in selecting plastid-derived reads from the whole-genome sequence, which includes NUPT- and MTPT-derived reads, obtained by short read sequencing. In addition, total cellular DNA contains only 1–10% cpDNA, and the amount decreases during plant development (Baumgartner et al., 1989; Oldenburg and Bendich, 1991; Shaver et al., 1995; Oldenburg and Bendich, 2004). These problems reduce the efficiency of sample multiplex analysis in a single sequencing run on low-output sequencers such as MiSeq, MiniSeq, and PacBio RS II. Therefore, the isolation of high-purity cpDNA will improve the accuracy and cost-efficiency of plastome sequencing. Four procedures for cpDNA isolation have been reported: Percoll density gradient centrifugation to obtain intact chloroplasts free of other organelles (Lang and Burger, 2007; Kaneko et al., 2016); high salt concentration to remove contaminating DNA ionically attached to the chloroplast surface (Shi et al., 2012); the use of DNase I to digest DNA bound to the chloroplast surface (Kolodner and Tewari, 1979); and liquid nitrogen pre-treatment to prevent nuclear breakage (Hirai et al., 1985).

Shi et al. (2012) reported that the DNase I method digested not only the contaminating DNA but also cpDNA within chloroplasts. So we compared Percoll density gradient centrifugation (PG), high salt (HS), and liquid nitrogen coupled with a sucrose gradient (LN) to optimize cpDNA analysis by NGS. We sequenced the highly purified cpDNA to compare the advantage of cpDNA sequencing with whole-genome sequencing. We also assessed the performance of SNP/insertion–deletion (indel) calling and *de novo* assembly on the plastome to evaluate the effect of cpDNA purity on NGS analysis.

MATERIALS AND METHODS

Plant Material

We germinated seeds of several rice (*O. sativa*) cultivars: temperate japonica Nipponbare (Shiga Prefecture Agricultural Research Center) and Koshihikari (Niigata Agricultural Research Institute); tropical japonica Sensho, Urasan, Padi Perak, and Khao Nok (NARO, Genetic Resources Center); aus Chinsurah Boro 2 (Tohoku University) and Kasalath; and indica 93-11 (National Institute of Genetics).

Protocols for Chloroplast Isolation

The three techniques selected for cpDNA isolation are described as follow: To determine the best method, we first performed three independent experiments to extract the cpDNA from bulk samples of Nipponbare by using each method (LN, HS, PG method). Then, the cpDNA from bulk samples of other cultivars were extracted once or twice by LN method.

¹<https://www.ncbi.nlm.nih.gov/genome/browse/?report=5>

Updated Liquid Nitrogen–Sucrose Density Gradient Centrifugation (LN) Method

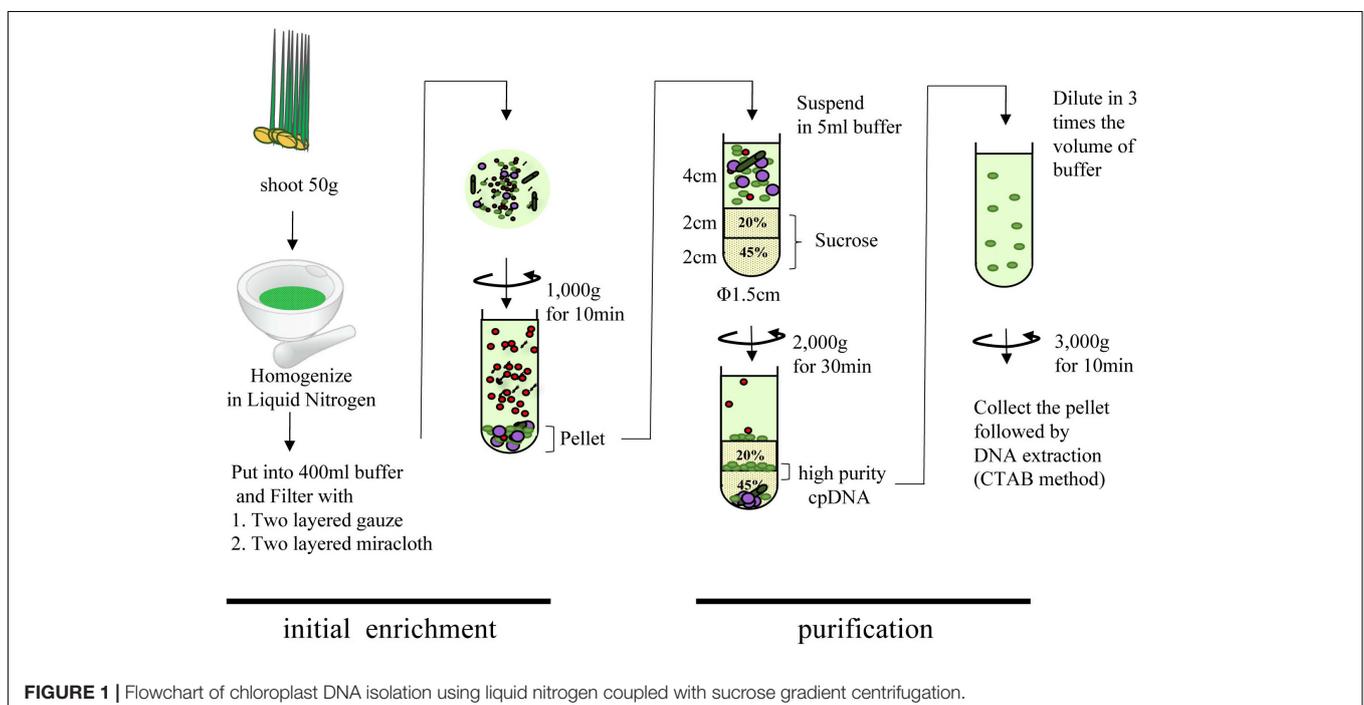
Rice seeds were sterilized and germinated on moist filter paper in culture dishes in the dark at 28°C for up to 4 days. The germinated seeds were dispersed onto a layer of water-agarose gel laid to block contamination by fungi over 1/2 MS medium agarose gel. Plants were placed in a growth chamber (28:23°C, 12:12 h, light:dark; 20,000 lux) for 8 days. cpDNA was isolated as described (Hirai et al., 1985) with some modifications (Figure 1). All procedures were performed at 4°C, and all centrifugations were performed in a CP80NX ultracentrifuge (Hitachi) with a P40ST swing rotor and 13PA tubes. In brief, 50 g of fresh shoot was cut into pieces (~3 cm), frozen in liquid nitrogen for 3 min, and gently ground into a fine powder with a mortar and pestle. The material was suspended in 400 mL of isolation buffer (50 mM Tris-HCl pH 8.0, 0.35 M sucrose, 7 mM EDTA, 5 mM 2-mercaptoethanol, 0.1% BSA) and incubated for 5 min in the dark. The suspension was filtered through two layers of gauze and then two layers of Miracloth (Merck). The filtrate was centrifuged at 1000 × g for 10 min. The pellet was suspended in 5 mL of isolation buffer, and the suspension was loaded slowly onto a stepwise 20%/45% discontinuous sucrose density gradient in 50 mM Tris-HCl (pH 8.0), 0.3 M sorbitol, and 7 mM EDTA. The gradient was centrifuged at 2000 × g for 30 min in a swinging bucket rotor. The green band at the 20%/45% sucrose interface was collected, diluted with three volumes of isolation buffer, and centrifuged at 3000 × g for 10 min in a swinging bucket rotor. Genomic DNA was isolated from the pellet by the cetyltrimethylammonium bromide method (Shi et al., 2012) or with a DNeasy Plant Mini Kit (Qiagen).

Modified High Salt (HS) Method

Sterilized seeds were grown on 0.8% agar in the dark at 28°C for 7 days and then in natural daylight at 28°C for 14 days. cpDNA was isolated using the protocol described in Shi et al. (2012). All procedures were performed at 4°C. In brief, 20 g of fresh leaves was cut into small pieces (~1 cm) and homogenized for 30 s in 400 mL of buffer A (1.25 M NaCl, 0.25 M ascorbic acid, 10 mM sodium metabisulfite, 12.5 mM borax, 50 mM Tris-HCl pH 8.0, 7 mM EDTA, 1% [w/v] PVP-40, 0.1% [w/v] BSA, and 1 mM DTT). The homogenate was filtered through two layers of Miracloth, and the filtrate was centrifuged at 200 × g for 20 min to remove starch granules, nuclei, tissue debris, and aggregates. The supernatant was centrifuged at 3500 × g for 20 min, and the pellet was suspended in 250 mL buffer B (1.25 M NaCl, 12.5 M borax, 1% [w/v] PVP-40, 50 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.1% [w/v] BSA, and 1 mM DTT) to increase the purity of the isolated cpDNAs. This step was performed twice. cpDNA was extracted with a DNeasy Plant Mini Kit (Qiagen).

Percoll Gradient (PG) Centrifugation Method

Sterilized seeds were grown on 0.8% agar at 30°C in the dark for 11 days, and then under continuous light at 28°C for 3 days for greening. cpDNA was isolated as in our previous report (Kaneko et al., 2016). All procedures were performed at 4°C. In brief, rice shoots (20 g) were homogenized in 20 mL of isolation buffer (50 mM HEPES-KOH pH 7.5, 0.33 M sorbitol, 5 mM MgCl₂, 5 mM MnCl₂, 5 mM EDTA, and 50 mM sodium ascorbate). The homogenate was filtered through four layers of gauze and then four layers of Miracloth. The filtrate was layered onto a cushion of 80% (v/v) Percoll (Sigma) in the above isolation buffer (except for the sodium ascorbate) and centrifuged at



2000 × *g* for 4 min. The crude chloroplasts on the Percoll surface were collected and diluted with more than twice the volume of isolation buffer, and then layered onto a discontinuous density gradient of 40 and 80% Percoll. The gradient was centrifuged at 4000 × *g* for 10 min. Intact chloroplasts enriched around the 40%/80% Percoll interface were collected and centrifuged again as before. Intact chloroplasts were collected, diluted with five volumes of isolation buffer, and centrifuged at 2000 × *g* for 4 min. cpDNA was extracted with a DNeasy Plant Mini Kit (Qiagen).

In all three protocols, DNA were eluted from the DNeasy spin column using 80 μl (LN and PG method) or 40 μl (PG method) of elution buffer. Two microliters of the isolated cpDNA was loaded and separated in 0.8% agarose gel and visualized with ethidium bromide.

Genome Copy Number Analysis by qPCR

We tested the cpDNA purity of DNA isolates by quantifying the number of copies of chloroplast, mitochondrial, and nuclear genomes. cpDNA, mtDNA, and ncDNA were quantified by quantitative real-time PCR (qPCR) using SsoFast EvaGreen Supermix (Bio-Rad) on a CFX96 real-time PCR system/C1000 Thermal Cycler (Bio-Rad). The thermocycling conditions were denaturation at 98°C for 2 min, and 39 cycles of 98°C for 2 s and 60°C for 5 s. To analyze genome copy number, we designed two sets of primers for each genome (Supplementary Table 1) to improve the accuracy of quantification, and used the mean as the threshold cycle (Ct) value of each genome. First, we calculated the PCR amplification efficiencies of each primer pair with a dilution series of a plasmid standard (10⁴ to 10⁹ copies of pGEM-T::Actin1::GAPDH::atpI::psbA::cob::coxII). The amplification efficiencies of all six genes were close to 2.0 (1.93–2.03), and *R*² values were between 0.989 and 0.999. Next, we calculated the copy ratios of cpDNA/ncDNA and cpDNA/mtDNA from the standard curves drawn from the above dilutions and compared these results of absolute quantification with the results of relative quantification using the 2^{-ΔΔCt} method (Schmittgen and Livak, 2008). Since the results of relative quantification were almost the same as those of absolute quantification, we used the relative quantification method in the subsequent qPCR analysis, giving priority to efficiency. Samples were assayed in at least three technical replicates, and the average copy ratios were calculated. Then, each genome DNA content in the extracted DNA was estimated from the copy ratio and genome size (plastid, 134,525 bp; mitochondrial, 490,520 bp; nuclear, 373,245,519 bp).

DNA Library Construction and Next-Generation Sequencing

We used 1 ng of Nipponbare purified cpDNA as input for the Nextera XT DNA library preparation kit and the Nextera XT index kit (Illumina). Constructed libraries were sequenced on an Illumina MiSeq sequencer (300 bp paired-end) with the MiSeq reagent kit v3 (Illumina). All procedures followed the manufacturer's instructions.

Sequence Mapping and Variant Detection

The sequenced paired-end reads from the purified cpDNA and three sets of whole-genome sequencing reads (total DNA [tDNA] 1–3) downloaded from public databases were trimmed in Trimmomatic v. 0.33 software (Bolger et al., 2014) with the following parameters: SLIDINGWINDOW: 8:20; TRAILING: 20; MINLEN: 90 (tDNA 1), 100 (purified cpDNA and tDNA 2), 76 (tDNA 3). The processed reads were aligned to the rice plastid reference genome (X15901.1) or a combined plastid (X15901.1)-mitochondrial (BA000029) reference genome by using the BWA-MEM v. 0.7.15 algorithm (Li and Durbin, 2009) with default parameters. PCR duplicates in BAM files were marked with Picard tools v. 1.68 software². Then local realignment of reads around indels was done in GATK (Genome Analysis Toolkit) IndelRealigner software (DePristo et al., 2011). To estimate cpDNA purity, we extracted unaligned reads from BAM files in SAMtools software (Li et al., 2009) and realigned them on the rice mitochondrial reference genome (BA000029). These unmapped hits were extracted again and realigned on the rice nuclear reference genome (IRGSP-1.0). The coverage depth of each genome was calculated from the number and length of high-quality reads in a 250-nt sliding window. To calculate allele frequency at individual plastid genome positions, we generated wig files describing the base (A/C/G/T) content in a 1-nt sliding window from BAM files in igtools v. 2.2 software (Robinson et al., 2011; Thorvaldsdóttir et al., 2013). After removal of data neighboring indels because of low reliability, we calculated first and second allele frequencies and coverage depths from wig files with a custom Perl script and then visualized them in 3D scatter plots using the scatterplot3d v. 0.3-37 tool of R (Ligges and Mächler, 2003). For variant calling, we used the SAMtools mpileup v. 1.4.1 tool (Li et al., 2009) with default parameters and GATK HaplotypeCaller v. 3.6 software (DePristo et al., 2011) with the '-ploidy 1' parameter to compare SNPs and small indels from BAM files. We filtered out heterozygous and low-quality variants (QUAL < 20) in SAMtools, and low-quality variants (QUAL < 20) in GATK.

De Novo Assembly

PCR duplicates were removed from paired-end reads using the *k*-mer-based method implemented in a Perl script³. From paired-reads of total DNA in BAM files, aligned plastid genome reads were extracted for enrichment of plastid reads, and then PCR duplicate reads were filtered out. Contigs were assembled from these reads in SOAP-denovo2 software (Luo et al., 2012) with various sets of *k*-mer parameters (Supplementary Table 4). After assembled scaffolds shorter than 500 bp were filtered out, sequences were compared against the plastid reference genome by NCBI BLAST 2 (Tatusova and Madden, 1999). Alignment results and detected SNPs/indels were visualized by Circos software (Krzywinski et al., 2009).

²<http://broadinstitute.github.io/picard/>

³<https://github.com/linneas/condetri>

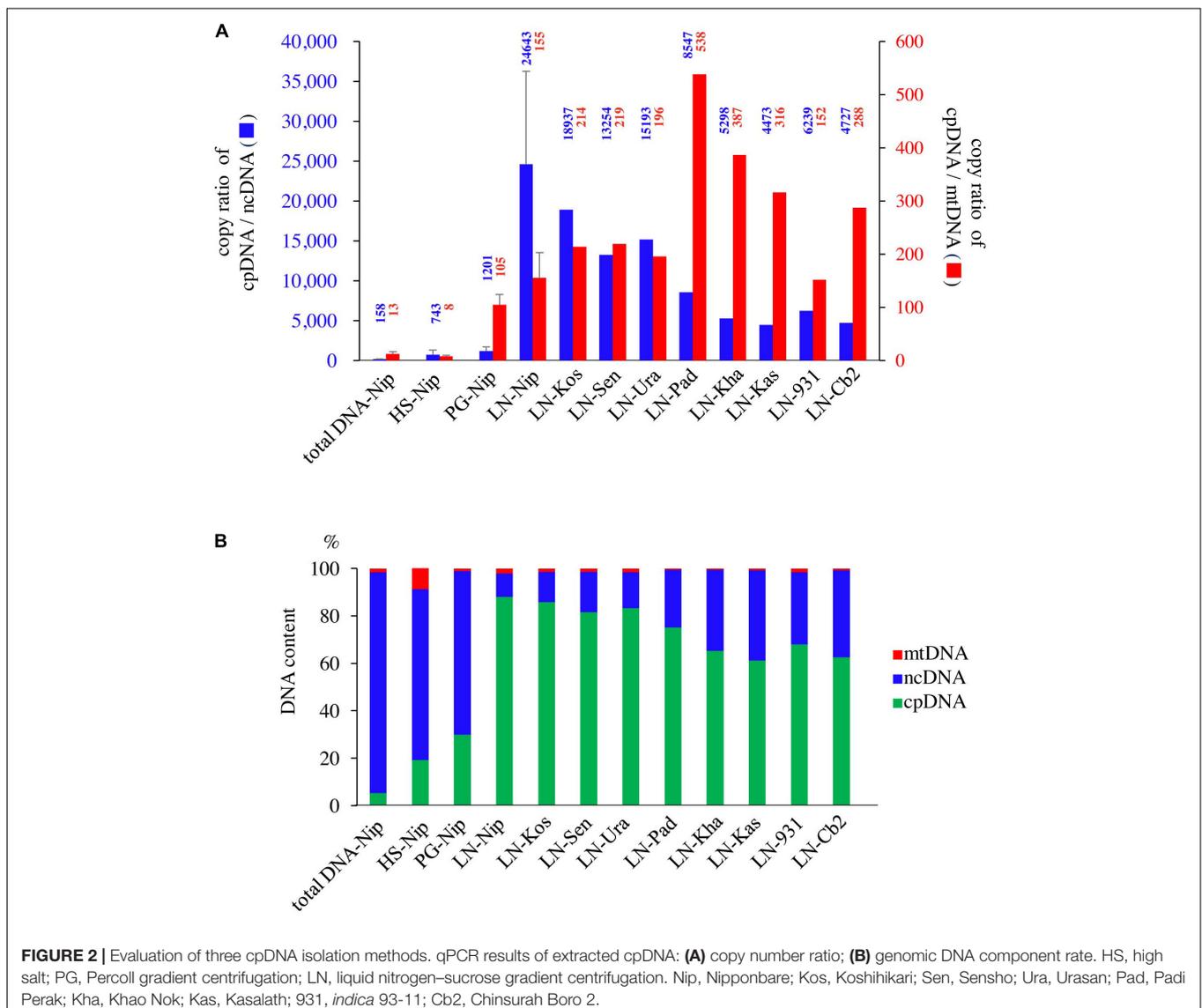
RESULTS AND DISCUSSION

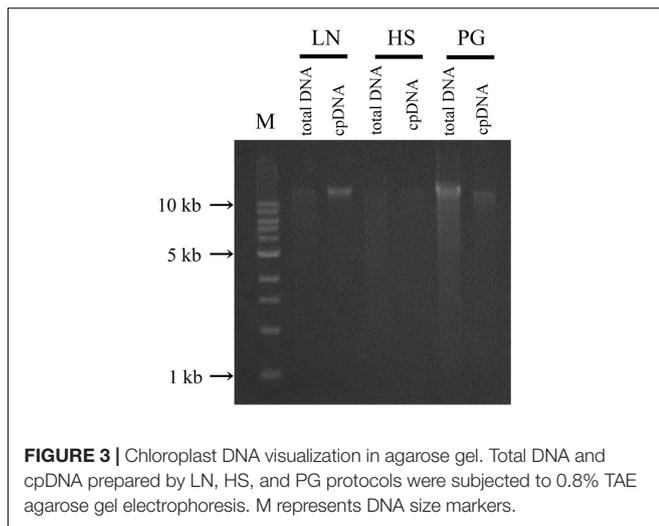
Comparison of Chloroplast DNA Isolation Methods for NGS Sequencing

Many areas of chloroplast research require assessment of the quality and quantity of cpDNA. Although assessment several methods have been developed, there is little discussion in the literature of whether they provide similar quality of information (Rowan et al., 2009; Rowan and Bendich, 2011). Large-scale studies of genome variation and evolution, which rely on large quantities of plant material, also require cpDNA isolation. Here, we compared three methods of isolating cpDNA (Supplementary Table 2): updated liquid nitrogen–sucrose density gradient centrifugation (LN), high-salt buffer (HS), and Percoll gradient centrifugation (PG). The original LN protocol was improved at various steps (Hirai et al., 1985) associated with gradient centrifugation to isolate cpDNA (Figure 1). The

main improvements, made to allow the use of general laboratory instruments and to simplify the method, were the use of a mortar and pestle instead of a liquid nitrogen-resistant mixer, and omission of 65% sucrose following a preliminary experiment that showed heavy contamination of nuclear and mitochondrial DNA at the 45%/65% interface. The protocol was further extended to the isolation of cpDNA from several rice accessions from four cultivar groups.

qPCR was used to determine the purity of the cpDNA samples for NGS. The HS method enriched not only cpDNA but also mtDNA (Figure 2), which may reduce cpDNA purity. The PG method enriched cpDNA with minimal contamination of mtDNA, but still with significant amounts of ncDNA. On the other hand, the LN technique consistently gave high purity, with cpDNA copy number ratios of 24 643× ncDNA and 155× mtDNA (Figure 2A, LN-Nip). A high ratio is critical to reducing misaligned reads on the plastome, such as NUPTs or MTPTs, and





to increasing NGS accuracy. Moreover, cpDNA accounted for up to 88% of the total isolated DNA (**Figure 2B**), which is likely to provide meaningful cost-effectiveness of an NGS run. The LN method yielded high-purity cpDNA in all nine cultivars tested. These results show the potential for the LN method to be applied to a wide range of rice cultivars, as we have since confirmed (unpublished data).

In the updated LN method, isolated DNA gave a well defined electrophoresis band, which is indicative of undegraded DNA (**Figure 3** and Supplementary Table 2). This DNA could be of sufficient quality not only for short-read sequencing, but also for mate-pair and long-read sequencing. The high-quality pure cpDNA isolated from LN method would be suitable to facilitate the plastome sequencing using the novel third-generation DNA sequencing technologies such as PacBio RS II and Oxford Nanopore MinION. This process may overcome the hard-to-assemble IRs regions composed of more than 20,000 base pairs long, and can enable new insights such as large indels and structural variations. A large amount of chloroplast pellet was collected, from which abundant cpDNA was extracted: 800 ng of cpDNA from 50 g of shoot. We also confirmed that as little as 4 g of shoot gave enough cpDNA for our NGS platform (Illumina MiSeq coupled with Nextera XT DNA library preparation).

However, the DNA obtained from the HS and PS protocols displayed very weak bands and smeared, indicative of low DNA yield (**Figure 3** and Supplementary Table 2).

Assessments of Resequencing: Mapping and Variant Calling

Illumina sequencing produced 4,105,152 paired-end reads with an average read length of 175 bp and a total of 718,401,600 bases (**Table 1**). We downloaded three sets of whole-genome sequence data (tDNA 1–3) from the public database to assess the influence of cpDNA purity on NGS analysis (**Table 1**). The sequence reads were aligned to the Nipponbare plastid reference genome (X15901.1) to discover putative SNPs and small indels. Over 78% of reads were aligned in the cpDNA purified by the LN method, versus only 1.2–10.2% in the tDNA (**Table 1**), corresponding to ratios of 7.7× to 66× tDNA. The plastid genome (pt) was highly enriched in the cpDNA compared with those in three tDNAs (**Table 2**). Massively parallel sequencing gave an increase of at least fivefold in cpDNA purity compared with leaf tDNA (range, <3–30%) (Nock et al., 2011). The LN protocol is thus a feasible replacement for the PG and HS methods. The HS method isolated only pea cpDNA (Bookjans et al., 1984), and its improvement (Shi et al., 2012) did not improve purity. Hirao et al. (2008) considered the use of sucrose density gradient centrifugation as the best method for separating ncDNA contamination from cpDNA, which qPCR results (**Figure 2**) and the Illumina sequencing

TABLE 2 | cpDNA purity: coverage depths and copy ratio of plastid (pt), mitochondrial (mt), and nuclear (nc) genomes of the purified chloroplast DNA (cpDNA) and three total genomic DNAs.

	Depth			Copy ratio	
	pt	mt	nc	pt/nc	pt/mt
Purified cpDNA	3071	67	0.3	10172	45.7
Total DNA_1	668	44	2.2	301	15.3
Total DNA_2	1094	187	18.3	60	5.8
Total DNA_3	366	66	11.0	33	5.5

cpDNA purity was calculated from coverage depth of each genome. After alignment on the plastid genome, unmapped reads were realigned on the mitochondrial genome, then the unmapped reads on the mitochondrial genome were realigned on the nuclear genome.

TABLE 1 | Summary of NGS samples and aligned results.

DNA	Library			Reads (After_QC)			Aligned reads		
	SRA	Instrument	Layout	Read1	Read2	# Reads	# pt	# pt_uniq	% pt
Purified cpDNA	This study	Miseq	Paired	204	146	4,105,152	3,233,842	3,230,265	78.8
Total DNA_1	SRR1239746	Hiseq2000	Paired	90	89	10,595,500	1,076,775	1,003,940	10.2
Total DNA_2	SRR1614244	Hiseq2000	Paired	101	101	70,075,112	927,471	898,850	1.3
Total DNA_3	SRR077421	GAll	Paired	76	76	55,180,822	673,954	648,676	1.2
	SRR077422								
	SRR077425								

Purified cpDNA: Nipponbare cpDNA obtained by LN method (**Figures 2A,B**, column 4). Total DNAs: published Nipponbare whole-genome sequencing reads downloaded from public database. Reads 1 and 2 indicate mean lengths of reads.

(Table 2) strongly supported, indicating high enough yield and purity to perform subsequent resequencing and genome assembly. The small difference between the results of mapping and qPCR analysis is likely due to the calculation methods, since qPCR calculates copy ratio from two target regions on each genome, whereas NGS aligns sequences across the whole genome. Visualization of the read alignment shows that the plastid reference genome is sufficiently covered in each sample (Figure 4). While tDNA samples show uniform coverage depth across the reference genome, the purified cpDNA shows irregular coverage depth. The inconsistency would arise from the properties of the Nextera XT library kit, since a high number of PCR enrichment cycles can easily cause PCR-dependent coverage bias (Lan et al., 2015).

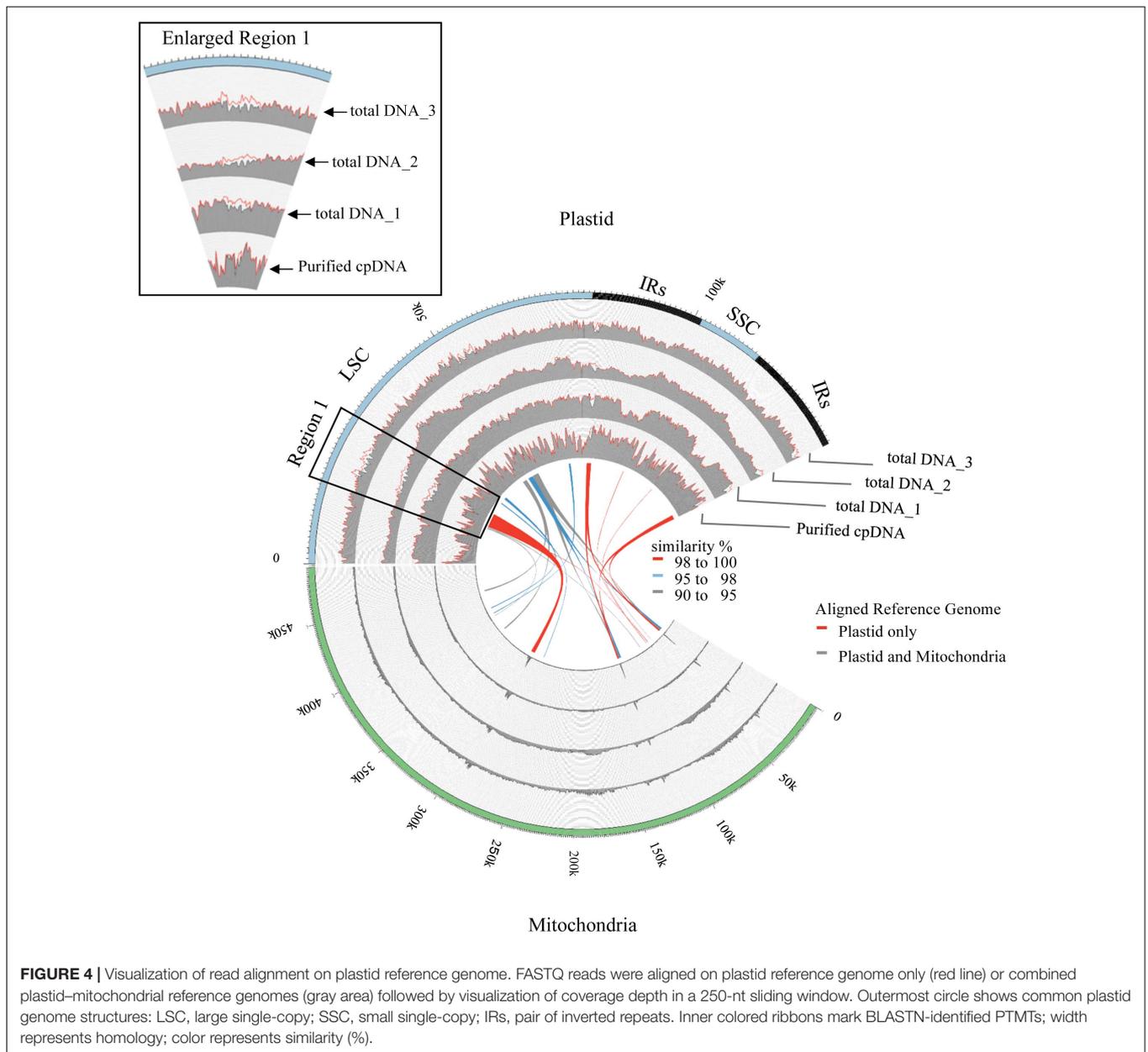
Next, we analyzed misaligned NUPT and MTPT reads, which disrupt precise analysis, on the plastid genome. We considered that MTPT reads containing a small number of mismatches against plastid genome could be separated by aligning reads to a combined plastid-mitochondria reference genome. Indeed, tDNA showed differences in coverage depth between the Pt and Pt-Mt reference genomes within regions of high similarity between the genomes, while purified cpDNA showed approximately similar coverage depth across the whole plastid genome (Figure 4; e.g., Region 1). To further detect NUPT- and MTPT-derived noise, we plotted the first and second allele frequencies and coverage depth at individual base positions on the plastid genome in 3D graphs (Figure 5). In the purified cpDNA, the first allele frequencies are close to 1.0 across the genome except in positions of low coverage depth, indicating low contamination by mitochondrial or nuclear DNA. By contrast, tDNA shows several sites where the first allele frequency was reduced and one or more other alleles were detected, even in positions with deep coverage depth. This result corroborates the frequency of this tendency and the lower purity of cpDNA (Table 2). Using the combined Pt-Mt reference genome improved the percentage of first allele frequencies. This result suggests that first allele reductions result from contamination by MTPTs, supporting observations of coverage depth differences such as in Region 1 in Figure 4. Furthermore, these results demonstrate that aligning on the combined Pt-Mt reference genome enables reduction of mtDNA contamination by a computational approach. However, it seems that the remaining low first allele frequencies are subject to noise derived from ncDNA. It is difficult to remove ncDNA-derived contamination by computation because some rice NUPTs have the same sequence as in the complete plastid reference genome and occur in multiple copies in the nuclear genome (Matsuo et al., 2005). The above results indicate that purified cpDNA can lead to high coverage depth of the chloroplast genome with a low number of reads, providing robust mapping and high-throughput sequencing of the rice plastid genome. Using tDNA sequenced on the Illumina platform is not consistently reliable, showing a higher rate of error alignment, which will likely affect later analysis such as plastid genome assembly or variant detection. It is important to highlight here the power of this technique in isolating

cpDNA with improved data quality and lowered sequencing costs.

The rice Nipponbare plastid reference genome X15901.1 was released in 1989 (Hiratsuka et al., 1989). Later, Tang et al. (2004) independently published a Nipponbare plastid reference genome, AY522330, noting 79 SNPs and 110 indels of putative sequence errors. To assess the inference of misaligned reads from the NUPTs and MTPTs by detection of these sequencing errors, we used AY522330 as correct data, and processed the BAM files describing the read alignments on X15901.1 to detect variants using two distinct variant callers, SAMtools mpileup and GATK HaplotypeCaller (Table 3 and Supplementary Table 3). GATK HaplotypeCaller had much higher sensitivity than SAMtools mpileup, as previously reported (Liu et al., 2013; Pirooznia et al., 2014; Yi et al., 2014). SAMtools mpileup failed to detect most indels (Supplementary Table 3), having very low sensitivity in indel calling (Tian et al., 2016). As sample purity decreases, SAMtools detects more heterozygous variants, which may be false positives, as they don't exist in AY522330. Although it is possible to select highly reliable variants by filtering out heterozygous and low-quality variants, some false-positive variants were found in all samples (e.g., plastid genome sites at 44772, 44775, 70290, and 70291; Supplementary Table 3). On the other hand, all variants identified by GATK HaplotypeCaller are consistent with AY522330 and show a high-quality score tolerant to the filtering process, regardless of sample purity, and no difference was found in the number of variants within the four samples. These results suggest that GATK HaplotypeCaller is superior at detecting mutations with high accuracy, not only in purified cpDNA, but also in total DNA samples, in the resequencing analysis of the plastid genome. However, GATK identified only 129 variants out of 189 reported in AY522330 (Tang et al., 2004). Most of those missing variants lie within the two IRs (Figure 6 and Supplementary Table 3), since GATK HaplotypeCaller ignores low-mapping-score reads (e.g., multi-mapping reads). Similarly, population genetic analysis of chloroplasts in 383 rice varieties also showed lower SNP/indel density within the IRs than in other plastid genome regions (Tong et al., 2016). Our efforts to detect variants in the IRs by changing several parameters of GATK HaplotypeCaller to allow low-mapping-score reads did not resolve this issue (data not shown). Although GATK HaplotypeCaller is the current gold standard variant caller, our results show that the development of other computation approaches is required for plastome resequencing. The discrepancies between our purified cpDNA and total DNA studies shed light on the importance of plant plastidial studies to thoroughly describe how to map them.

Assessments of Plastome *de Novo* Assembly: Scaffolds and Variant Calling

We generated *de novo* assembled plastid genomes for assessment of the purified cpDNA and the three tDNAs (Figure 6 and Supplementary Tables 3–5). Comparison of the scaffold alignments showed that the plastid genome was almost fully covered by the assemblies of all samples, but the tDNAs generated numerous shorter scaffolds. Repeat sequences are well known to hinder long contig formation (Kim et al., 2015). Our



results also show that the contig elongations stop around the boundaries of the evolutionally conserved pair of large IRs. Since it is difficult to construct a complete plastid genome from a single library, supplementation with libraries of different insert sizes and mate pairs, long read sequencing, and PCR analysis is also necessary (Ferrarini et al., 2013; Naito et al., 2013).

The genome assembled *de novo* from the purified cpDNA has long contiguous sequences which were joined into three scaffolds in the correct order (Figure 6A). This is potentially another advantage of the LN approach. In contrast, *de novo* assembly using tDNA produced 5, 10, and 14 scaffolds (Figures 6B–D). A simulation study demonstrated that de Bruijn Graph Assemblers such as SOAPdenovo2, which we used, can

improve the assembly of longer contigs from longer reads (Knudsen et al., 2010). Certainly, among tDNAs 2 and 3, the plastid DNA ratio was approximately the same, but the longer read length of tDNA 3 resulted in the assembly of longer contigs (Figures 6C,D and Table 1). Moreover, high heterozygosity increases the complexity of the de Bruijn graph structure, leading to small contigs and base call errors (Kajitani et al., 2014). Other reports also suggest that sequence error and GC bias create ‘dropouts’—multiple gaps in assemblies—and hence small contigs and scaffolds, even in small genomes such as those of plastids (Li et al., 2010; Minoche et al., 2011). In the *de novo* plastome assembly, widespread NUPTs and MTPTs are likely to behave like heterozygous sites and sequence errors, and thus to interrupt contig formation. In fact, our results show a clear

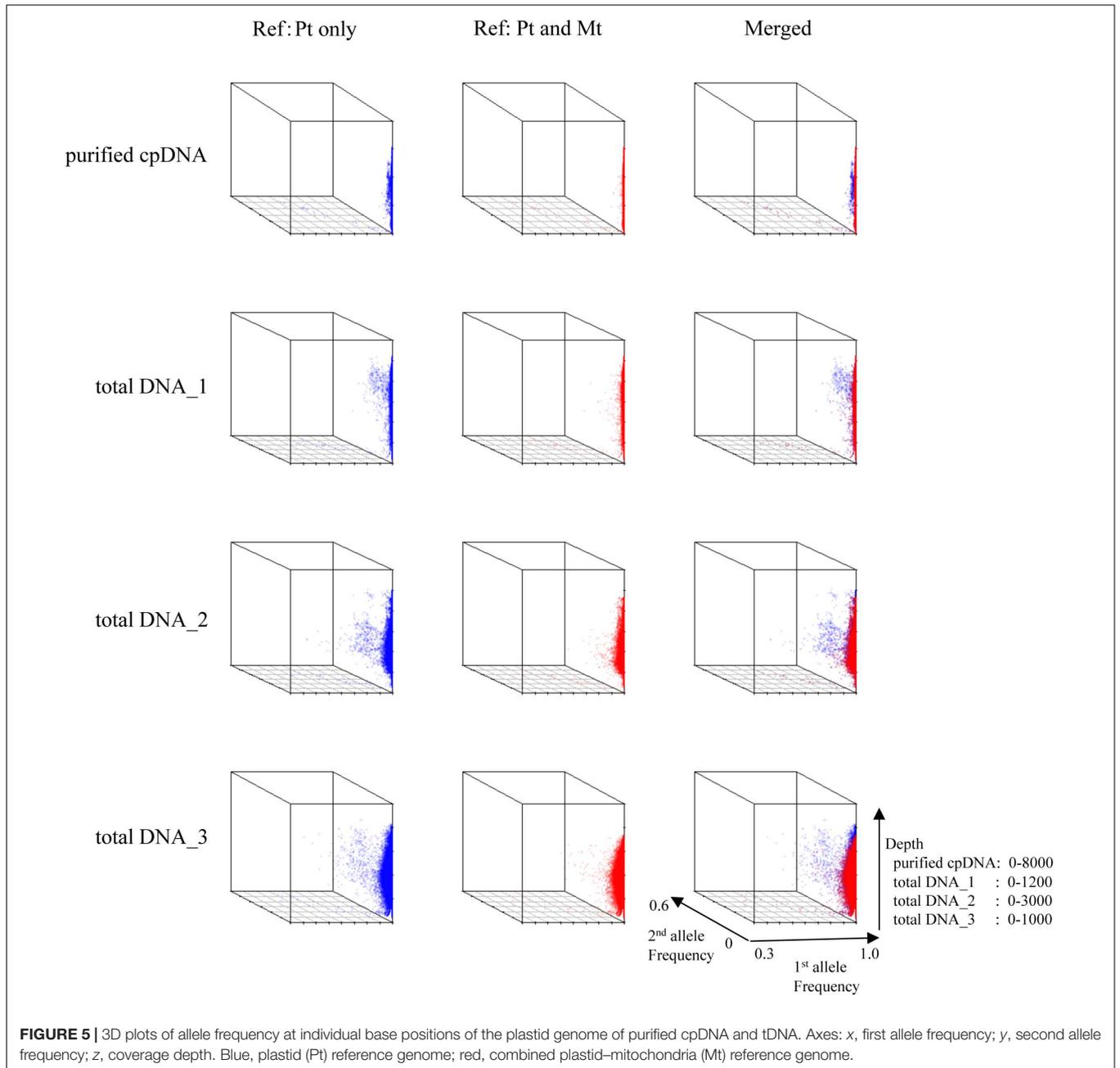
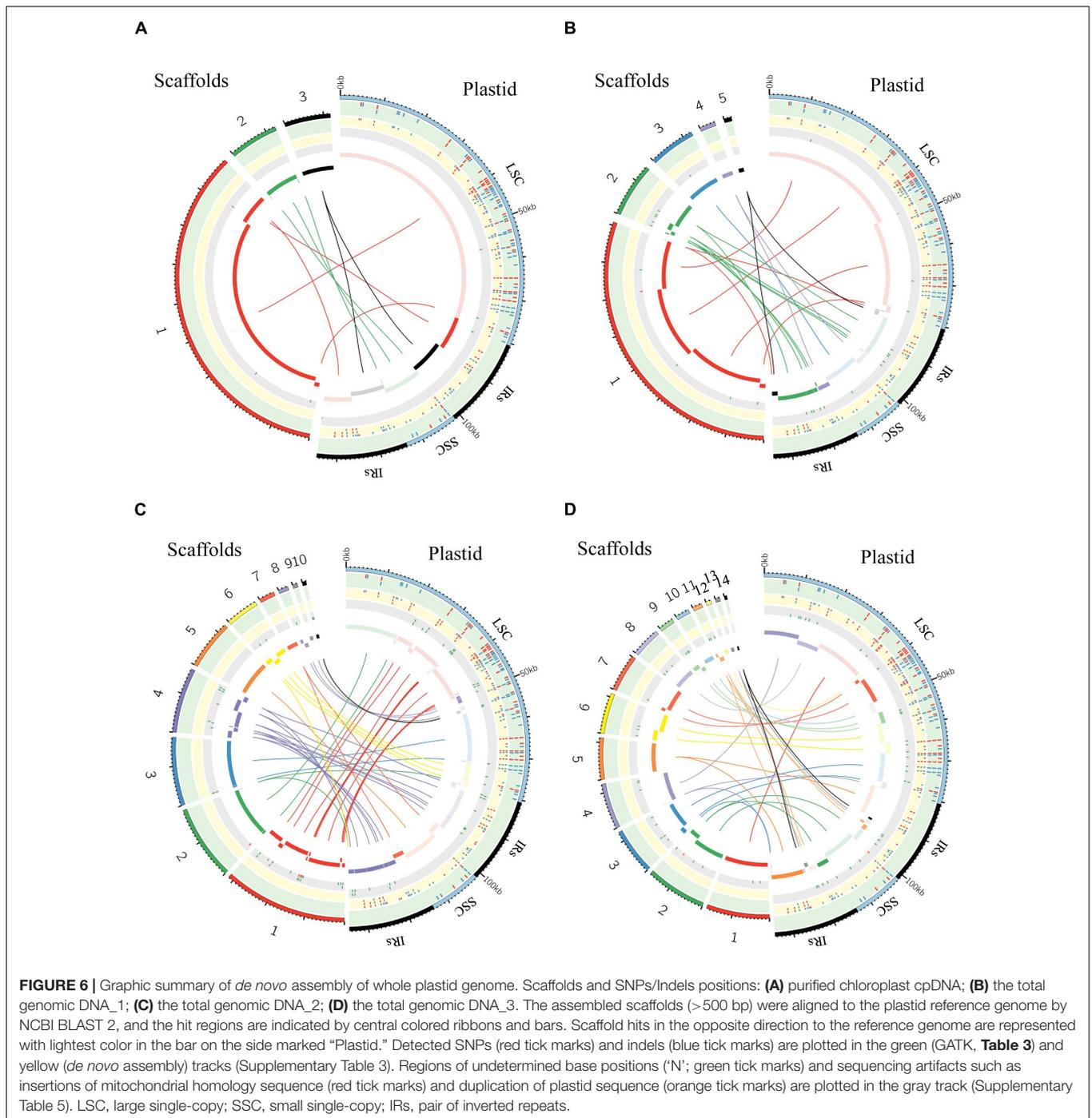


TABLE 3 | Plastid variant call results from SAMtools mpileup versus GATK HaplotypeCaller using the purified cpDNA and the three total genomic DNA samples.

Aligned (ref)	SAMtools mpileup				GATK HaplotypeCaller			
	Pt only		Pt and Mt		Pt only		Pt and Mt	
	Total	Pass	Total	Pass	Total	Pass	Total	Pass
Purified cpDNA	86	64	86	64	129	129	129	129
Total DNA_1	86	66	86	67	129	129	129	129
Total DNA_2	131	64	87	65	129	129	129	129
Total DNA_3	128	69	99	69	129	129	129	129

Processed reads were aligned on the plastid reference genome (Pt only) or on the combined plastid-mitochondrial reference genome (Pt and Mt). Heterozygous and low-quality variants (QUAL < 20) were filtered out in SAMtools, and low-quality variants (QUAL < 20) were filtered out in GATK.



tendency for scaffolds to be longer with purer plastid DNA (**Figure 6** and **Table 1**). Additionally, scaffolds in tDNAs contain multiple low-quality regions composed of both small and large gaps of consecutive undetermined ('N') bases and sequencing artifacts such as the insertion of regions with high homology to mitochondrial sequences and the duplication of plastid sequences (**Figure 6** and Supplementary Table 5). By contrast, we obtained scaffolds of purified cpDNA across the whole plastid genome with very few 'N' sequences.

As a result of SNP/indel detection, *de novo* sequencing identified 187 or 188 variants out of 189 reported in AY522330, with a higher sensitivity than resequencing analysis by GATK HaplotypeCaller (**Figure 6** and Supplementary Table 3). Despite this high sensitivity, tDNA 2 and 3 returned 41 and 58 additional variants, indicative of high false-positive rates, but the purified cpDNA did not return other variants. Taken together, these results reveal specificity and robustness in the identification of SNPs/indels by *de novo* sequencing with high-purity cpDNA.

Overall, read length and cpDNA purity are key to the successful *de novo* assembly of plastid genomes. Increasing cpDNA purity compensates for the low yield of MiSeq and enables the best use of its 300-bp paired-end sequencing, which is the longest read length of current Illumina next-generation sequencers. High-purity cpDNA is crucial for *de novo* assembly and SNP/indel calling without sequencing artifacts. It's worth noting that *de novo* sequencing allowed the identification of SNPs/indels within the two IRs where GATK HaplotypeCaller ignored variant calling. This result indicates that *de novo* sequencing using high-purity cpDNA could be an effective method for detecting variants within IRs.

CONCLUSION

The updated LN technique permits the extraction of enriched cpDNA, allowing the investigation of plastid genomes in a more cost-effective, time-saving manner, with huge increases in sequence throughput. We demonstrated that it is possible to obtain high-quality cpDNA with which to perform functional analysis to widen the scope for high-throughput sequencing and gain new insights for other genetic studies. Collectively, our analyses strongly support that the LN protocol increases the depth of coverage with a low output of short-read sequencing, allowing the large-scale bioinformatic/computational analysis of data. Using this protocol, we generated highly accurate plastid genome sequences without sequencing artifacts. The application of NGS followed by read mapping analysis to highly purified cpDNA would allow efficient detection of SNPs and indels within a plant population and accessions. This improvement in chloroplast sequencing technologies may help the rapid advancement of the chloroplast genomics field, the understanding of plastid genome replication and repair, the high-resolution analysis of heteroplasmy, and the development of technologies for chloroplast transformation.

REFERENCES

- Baumgartner, B. J., Rapp, J. C., and Mullet, J. E. (1989). Plastid transcription activity and DNA copy number increase early in barley chloroplast development. *Plant Physiol.* 89, 1011–1018. doi: 10.1104/PP.89.3.1011
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bookjans, G., Stummann, B. M., and Henningsen, K. W. (1984). Preparation of chloroplast DNA from pea plastids isolated in a medium of high ionic strength. *Anal. Biochem.* 141, 244–247. doi: 10.1016/0003-2697(84)90452-4
- Daniell, H., Lin, C.-S., Yu, M., and Chang, W.-J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17:134. doi: 10.1186/s13059-016-1004-2
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Diekmann, K., Hodkinson, T. R., Fricke, E., and Barth, S. (2008). An optimized chloroplast DNA extraction protocol for grasses (Poaceae) proves suitable for whole plastid genome sequencing and SNP detection. *PLoS One* 3:e2813. doi: 10.1371/journal.pone.0002813

ACCESSION CODES

The sequence data have been deposited in the DDBJ Sequence Read Archive: DRR118684.

AUTHOR CONTRIBUTIONS

TM and TK designed and supervised the research. TT performed the experiments and carried out the analysis. TI, KO, and TO contributed analytic tools. TT, MB, and TM interpreted the data. MB contributed to the conception and design of the work. TT, MB, and TM wrote the manuscript. MB, KI, and TM revised and finalized the paper. All authors read and approved the final version of the manuscript.

FUNDING

This research was supported by KAKENHI Grants-in-Aid for Scientific Research (A) (15H02486) from the Japan Society for the Promotion of Sciences and by a Grant for Promotion of KAAB Projects (Niigata University) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

ACKNOWLEDGMENTS

Computations were performed on the NIG supercomputer at the ROIS National Institute of Genetics.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00266/full#supplementary-material>

- Drouin, G., Daoud, H., and Xia, J. (2008). Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* 49, 827–831. doi: 10.1016/j.ympev.2008.09.009
- Ferrari, M., Moretto, M., Ward, J. A., Šurbanovski, N., Stevanović, V., Giongo, L., et al. (2013). An evaluation of the PacBio RS platform for sequencing and *de novo* assembly of a chloroplast genome. *BMC Genomics* 14:670. doi: 10.1186/1471-2164-14-670
- Greiner, S., and Bock, R. (2013). Tuning a ménage à trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. *BioEssays* 35, 354–365. doi: 10.1002/bies.201200137
- Hirai, A., Ishibashi, T., Morikami, A., Iwatsuki, N., Shinozaki, K., and Sugiura, M. (1985). Rice chloroplast DNA: a physical map and the location of the genes for the large subunit of ribulose 1,5-bisphosphate carboxylase and the 32 KD photosystem II reaction center protein. *Theor. Appl. Genet.* 70, 117–122. doi: 10.1007/BF00275309
- Hirao, T., Watanabe, A., Kurita, M., Kondo, T., and Takata, K. (2008). Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC Plant Biol.* 8:70. doi: 10.1186/1471-2229-8-70
- Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., et al. (1989). The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts

- for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* 217, 185–194.
- Joseph, B., Corwin, J. A., Li, B., Atwell, S., and Kliebenstein, D. J. (2013). Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome. *eLife* 2:e00776. doi: 10.7554/eLife.00776
- Kahlau, S., Aspinall, S., Gray, J. C., and Bock, R. (2006). Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J. Mol. Evol.* 63, 194–207. doi: 10.1007/s00239-005-0254-5
- Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 24, 1384–1395. doi: 10.1101/gr.170720.113
- Kaneko, K., Takamatsu, T., Inomata, T., Oikawa, K., Itoh, K., Hirose, K., et al. (2016). N-glycomic and microscopic subcellular localization analyses of NPP1, 2 and 6 strongly indicate that trans-Golgi compartments participate in the Golgi to plastid traffic of nucleotide pyrophosphatase/phosphodiesterases in rice. *Plant Cell Physiol.* 57, 1610–1628. doi: 10.1093/pcp/pcw089
- Kim, K., Lee, S.-C., Lee, J., Yu, Y., Yang, K., Choi, B.-S., et al. (2015). Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Sci. Rep.* 5:15655. doi: 10.1038/srep15655
- Knudsen, B., Forsberg, R., and Miyamoto, M. M. (2010). A computer simulator for assessing different challenges and strategies of de novo sequence assembly. *Genes* 1, 263–282. doi: 10.3390/genes1020263
- Kolodner, R., and Tewari, K. K. (1979). Inverted repeats in chloroplast DNA from higher plants. *Proc. Natl. Acad. Sci. U.S.A.* 76, 41–45.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Lan, J. H., Yin, Y., Reed, E. F., Moua, K., Thomas, K., and Zhang, Q. (2015). Impact of three Illumina library construction methods on GC bias and HLA genotype calling. *Hum. Immunol.* 76, 166–175. doi: 10.1016/j.humimm.2014.12.016
- Lang, B. F., and Burger, G. (2007). Purification of mitochondrial and plastid DNA. *Nat. Protoc.* 2, 652–660. doi: 10.1038/nprot.2007.58
- Lewin, R. (1984). No genome barriers to promiscuous DNA: the movement of DNA sequences between mitochondrial, chloroplast and nuclear genomes is even more prolific than had been expected. *Science* 224, 970–971. doi: 10.1126/science.224.4652.970
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., et al. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20, 265–272. doi: 10.1101/gr.097261.109
- Ligges, U., and Mächler, M. (2003). Scatterplot3d – an R package for visualizing multivariate. *J. Stat. Softw.* 8, 1–20.
- Liu, X., Han, S., Wang, Z., Gelernter, J., and Yang, B.-Z. (2013). Variant callers for next-generation sequencing data: a comparison study. *PLoS One* 8:e75619. doi: 10.1371/journal.pone.0075619
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1:18. doi: 10.1186/2047-217X-1-18
- Maier, R. M., Neckermann, K., Igloi, G. L., and Kössel, H. (1995). Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* 251, 614–628. doi: 10.1006/jmbi.1995.0460
- Maliga, P. (2001). Plastid engineering bears fruit. *Nat. Biotechnol.* 19, 826–827. doi: 10.1038/nbt0901-826
- Martin, W., and Herrmann, R. G. (1998). Gene transfer from organelles to the nucleus: how much, what happens, and Why? *Plant Physiol.* 118, 9–17. doi: 10.1104/PP.118.1.9
- Matsuo, M., Ito, Y., Yamauchi, R., and Obokata, J. (2005). The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. *Plant Cell* 17, 665–675. doi: 10.1105/tpc.104.027706
- Minoche, A. E., Dohm, J. C., and Himmelbauer, H. (2011). Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. *Genome Biol.* 12:R112. doi: 10.1186/gb-2011-12-11-r112
- Moison, M., Roux, F., Quadrado, M., Duval, R., Ekovich, M., Lê, D.-H., et al. (2010). Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in *Arabidopsis thaliana*. *Plant J.* 63, 728–738. doi: 10.1111/j.1365-313X.2010.04275.x
- Morris, G. P., Grabowski, P. P., and Borevitz, J. O. (2011). Genomic diversity in switchgrass (*Panicum virgatum*): from the continental scale to a dune landscape. *Mol. Ecol.* 20, 4938–4952. doi: 10.1111/j.1365-294X.2011.05335.x
- Naito, K., Kaga, A., Tomooka, N., and Kawase, M. (2013). De novo assembly of the complete organelle genome sequences of azuki bean (*Vigna angularis*) using next-generation sequencers. *Breed. Sci.* 63, 176–182. doi: 10.1270/jsbbs.63.176
- Nock, C. J., Waters, D. L. E., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., et al. (2011). Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol. J.* 9, 328–333. doi: 10.1111/j.1467-7652.2010.00558.x
- Oldenburg, D. J., and Bendich, A. J. (1991). Degradation of chloroplast DNA in second leaves of rice (*Oryza sativa*) before leaf yellowing. *Protoplasma* 160, 89–98. doi: 10.1007/BF01539960
- Oldenburg, D. J., and Bendich, A. J. (2004). Changes in the structure of DNA molecules and the amount of DNA per plastid during chloroplast development in maize. *J. Mol. Biol.* 344, 1311–1330. doi: 10.1016/j.jmb.2004.10.001
- Olejniczak, S. A., Łojewska, E., Kowalczyk, T., and Sakowicz, T. (2016). Chloroplasts: state of research and practical applications of plastome sequencing. *Planta* 244, 517–527. doi: 10.1007/s00425-016-2551-1
- Pirooznia, M., Kramer, M., Parla, J., Goes, F. S., Potash, J. B., McCombie, W., et al. (2014). Validation and assessment of variant calling pipelines for next-generation sequencing. *Hum. Genomics* 8:14. doi: 10.1186/1479-7364-8-14
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754
- Roux, F., Mary-Huard, T., Barillot, E., Wenes, E., Botran, L., Durand, S., et al. (2016). Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. *Proc. Natl. Acad. Sci. U.S.A.* 113, 3687–3692. doi: 10.1073/pnas.1520687113
- Rowan, B. A., and Bendich, A. J. (2011). “Isolation, quantification, and analysis of chloroplast DNA,” in *Chloroplast Research in Arabidopsis. Methods in Molecular Biology (Methods and Protocols)*, Vol. 774, ed. R. Jarvis (New York, NY: Humana Press), 151–170. doi: 10.1007/978-1-61779-234-2_10
- Rowan, B. A., Oldenburg, D. J., and Bendich, A. J. (2009). A multiple-method approach reveals a declining amount of chloroplast DNA during development in *Arabidopsis*. *BMC Plant Biol.* 9:3. doi: 10.1186/1471-2229-9-3
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999). Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res.* 6, 283–290. doi: 10.1093/dnares/6.5.283
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3, 1101–1108.
- Shaver, J. M., Oldenburg, D. J., and Bendich, A. J. (1995). Differential transcription of pea chloroplast genes during light-induced leaf development (continuous far-red light activates chloroplast transcription). *Plant Physiol.* 109, 105–112. doi: 10.1104/pp.89.3.1011
- Shi, C., Hu, N., Huang, H., Gao, J., Zhao, Y.-J., and Gao, L.-Z. (2012). An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. *PLoS One* 7:e31468. doi: 10.1371/journal.pone.0031468
- Sloan, D. B. (2015). Using plants to elucidate the mechanisms of cytonuclear co-evolution. *New Phytol.* 205, 1040–1046. doi: 10.1111/nph.12835
- Tang, J., Xia, H., Cao, M., Zhang, X., Zeng, W., Hu, S., et al. (2004). A comparison of rice chloroplast genomes. *Plant Physiol.* 135, 412–420. doi: 10.1104/pp.103.031245
- Tang, Z., Yang, Z., Hu, Z., Zhang, D., Lu, X., Jia, B., et al. (2013). Cytonuclear epistatic quantitative trait locus mapping for plant height and ear height in maize. *Mol. Breed.* 31, 1–14. doi: 10.1007/s11032-012-9762-3
- Tatusova, T. A., and Madden, T. L. (1999). BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* 174, 247–250.

- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi: 10.1093/bib/bbs017
- Tian, S., Yan, H., Neuhauser, C., and Slager, S. L. (2016). An analytical workflow for accurate variant discovery in highly divergent regions. *BMC Genomics* 17:703. doi: 10.1186/s12864-016-3045-z
- Tong, W., Kim, T.-S., and Park, Y.-J. (2016). Rice chloroplast genome variation architecture and phylogenetic dissection in diverse *Oryza* species assessed by whole-genome resequencing. *Rice* 9:57. doi: 10.1186/s12284-016-0129-y
- Wang, S., Liu, J., Feng, Y., Niu, X., Giovannoni, J., and Liu, Y. (2008). Altered plastid levels and potential for improved fruit nutrient content by downregulation of the tomato DDB1-interacting protein CUL4. *Plant J.* 55, 89–103. doi: 10.1111/j.1365-313X.2008.03489.x
- Wolfe, K. H., Li, W. H., and Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U.S.A.* 84, 9054–9058.
- Wu, C.-C., Ho, C.-K., and Chang, S.-H. (2015). The complete chloroplast genome of *Cinnamomum kanehirae* Hayata (Lauraceae). *Mitochondrial DNA* 27, 1–2. doi: 10.3109/19401736.2015.1043541
- Yi, M., Zhao, Y., Jia, L., He, M., Kebebew, E., and Stephens, R. M. (2014). Performance comparison of SNP detection tools with illumina exome sequencing data—an assessment using both family pedigree information and sample-matched SNP array data. *Nucleic Acids Res.* 42, e101–e101. doi: 10.1093/nar/gku392
- Yoshida, T., Furihata, H. Y., and Kawabe, A. (2014). Patterns of genomic integration of nuclear chloroplast DNA fragments in plant species. *DNA Res.* 21, 127–140. doi: 10.1093/dnares/dst045

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Takamatsu, Baslam, Inomata, Oikawa, Itoh, Ohnishi, Kinoshita and Mitsui. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.