



Whole genome sequence analysis of unidentified genetically modified papaya for development of a specific detection method



Kosuke Nakamura^{a,*}, Kazunari Kondo^{a,*}, Hiroshi Akiyama^a, Takumi Ishigaki^a, Akio Noguchi^a, Hiroshi Katsumata^b, Kazuto Takasaki^b, Satoshi Futo^b, Kozue Sakata^a, Nozomi Fukuda^a, Junichi Mano^c, Kazumi Kitta^c, Hidenori Tanaka^d, Ryo Akashi^d, Tomoko Nishimaki-Mogami^a

^a National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan

^b FASMAC CO., LTD, 5-1-3 Midorigaoka, Atsugi, Kanagawa 243-0041, Japan

^c Analytical Science Division, National Food Research Institute, National Agriculture and Food Research Organization, 2-1-12 Kannondai, Tsukuba, Ibaraki 305-8642, Japan

^d Graduate School of Agriculture, University of Miyazaki, 1-1 Gakuen Kibanadai Nishi, Miyazaki 889-2192, Japan

ARTICLE INFO

Article history:

Received 2 November 2015

Received in revised form 22 February 2016

Accepted 27 February 2016

Available online 3 March 2016

Keywords:

Genetically modified

Transgenic sequence

Next-generation sequencer

Genome sequence

Real-time PCR

Carica papaya L.

ABSTRACT

Identification of transgenic sequences in an unknown genetically modified (GM) papaya (*Carica papaya* L.) by whole genome sequence analysis was demonstrated. Whole genome sequence data were generated for a GM-positive fresh papaya fruit commodity detected in monitoring using real-time polymerase chain reaction (PCR). The sequences obtained were mapped against an open database for papaya genome sequence. Transgenic construct- and event-specific sequences were identified as a GM papaya developed to resist infection from a *Papaya ringspot virus*. Based on the transgenic sequences, a specific real-time PCR detection method for GM papaya applicable to various food commodities was developed. Whole genome sequence analysis enabled identifying unknown transgenic construct- and event-specific sequences in GM papaya and development of a reliable method for detecting them in papaya food commodities.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Genetically modified (GM) crops for food (GM food) are being increasingly developed (James, 2014). Potential toxicity or allergenicity derived from introduction of transgenic sequences in the genome is one of the concerns over acceptance of GM food by consumers. Therefore, use of GM food is regulated by many countries. Safety assessments of GM food are mandatory and independently performed based on safety assessment criteria, such as in the EU (Regulation (EC) No 1829/2003, 2003) and Japan (Standards for the Safety Assessment of GM Foods (Seed Plants), 2004). In these

countries, any unauthorized GM foods are prohibited from the market since authorization of GM food use is stipulated under law.

In some countries, unintended use of unauthorized GM food has been reported in various food commodities, such as rice (*Oryza sativa* L.) (Grohmann et al., 2015; Nakamura et al., 2013a), flax (*Linum usitatissimum* L.) (Young et al., 2015), wheat (*Triticum aestivum* L.) (Kim et al., 2015), potato (*Solanum tuberosum* L.) (Akiyama et al., 2002a, 2002b), and papaya (*Carica papaya* L.). In the case of papaya, an unauthorized GM papaya, which resists infection from *Papaya ringspot virus* (PRSV), was found in some processed papaya commodities (papaya-leaf tea, papaya jam, papaya pickles) in Japan in 2011 (Nakamura et al., 2011). Another unauthorized GM papaya line developed to effectively resist infection from the PRSV strains in Thailand was found in a processed papaya commodity (dried papaya) in Japan in 2013 (Nakamura et al., 2014). In the EU, GM papaya has been found in some imported foods (RASFF Portal. Unauthorised genetically modified papaya from unknown origin, 2015). As compared to non-GM papaya, no *de novo* toxicity or allergenicity by GM papaya ingestion has been reported so far (Fermín et al., 2011; Guo et al., 2009; Lin et al., 2013; Lin et al., 2015). Some GM papaya lines have been safely consumed as food by human more than a decade. Monitoring foods

Abbreviations: bp, base pair; Chy, chymopapain; Ct, threshold cycle; ΔRn, change in normalized reporter signal; DNA, deoxyribonucleic acid; FAM, 6-carboxyfluorescein; GM, genetically modified; LOD, limit of detection; NOST, nopaline synthase terminator; NPTII, neomycin phosphotransferase II; P35S, Cauliflower mosaic virus 35S RNA promoter; PRSV, Papaya ringspot virus; CP, coat protein; PCR, polymerase chain reaction; QV, quality value; RSD, relative standard deviation; TAMRA, 6-carboxytetramethyl-rhodamine; UV, ultraviolet.

* Corresponding authors.

E-mail addresses: kosnakamura@nihs.go.jp (K. Nakamura), kondo@nihs.go.jp (K. Kondo).

containing any novel unauthorized GM papaya lines is required to follow the regulation and to ensure food safety.

Some GM papaya lines have already been successful in production and commercialization for foods. PRSV-resistant GM papaya line 55-1 (line 55-1) from Hawaii, for example, was the first commercialized GM papaya for food use after regulatory approval in countries such as the United States, Canada and Japan (Fermín et al., 2011; Nakamura et al., 2013b). Many GM papaya lines having different transgenic constructs have been developed worldwide (Mendoza, Laurena, & Botella, 2008). In Japan, only line 55-1 for food use has been authorized. All other GM papaya is unauthorized and prohibited from being distributed to the market in Japan before gaining Japanese regulatory approval. Therefore, laboratories testing GM food have been monitoring food commodities to avoid distribution of unauthorized GM papaya in Japan.

The most frequently used approach to detect the presence of GM food is detection of common transgenic construct sequences derived from viral and bacterial genomes, such as *Cauliflower mosaic virus* 35S RNA promoter (P35S), the 3'-untranslated region of the *Rhizobium radiobacter nopaline synthase* terminator (NOST) or *neomycin phosphotransferase II* (NPTII) using conventional polymerase chain reaction (PCR) and real-time PCR methods (Mano et al., 2009; Reiting, Broll, Waiblinger, & Grohmann, 2007). However, false-positive results are frequently observed in test monitoring various food commodities due to viral and bacterial contamination (Kitagawa, Nakamura, Kondo, Ubukata, & Akiyama, 2014). Also, detection and identification of transgenic construct- or event-specific sequences in an unknown GM food using methods such as chromosome walking and inverse PCR require highly trained experimental skills and are time consuming.

In this study, whole genome sequence data were generated for a GM-positive fresh papaya fruit commodity detected in a monitoring test using real-time PCR. Common transgenic construct sequences, P35S and border regions in *Agrobacterium* T-DNA transformation, used in GM papaya development were exploited to search for sequence reads having sequences specific to GM papaya. Transgenic construct- and event-specific DNA sequences of the GM papaya were identified from the sequencing data. Based on the identified sequences, a specific, repeatable and reproducible real-time PCR detection method applicable to testing various types of papaya food commodities was developed.

2. Materials and methods

2.1. Papaya samples

Papaya food commodities (pickled papaya, papaya-leaf tea, jam) were purchased through the Internet. A fresh papaya fruit was obtained as GM papaya line PRSV-YK using transgenic construct-specific real-time PCR method during a monitoring test for commodities covered by regulations on GM food use in Japan. Fresh non-GM papaya (cultivar Sunset) fruit was purchased from the Hawaii Papaya Industry Association through the Consumer Affairs Agency of Japan.

2.2. Purification of papaya DNA

Sarcocarp was collected from a peeled fresh papaya fruit with the seeds removed, then 10 g aliquots were processed with a food processor and used for DNA purification. Other papaya food commodities, except jam, were ground using a food processor, then 10 g aliquots used for DNA purification. DNA was purified using an ion-exchange resin-type DNA extraction and purification kit (Genomic-tip; QIAGEN, Hilden, Germany) according to methods described previously (Ohmori et al., 2013). Purified DNA was

quantified for purity and concentration by measuring UV absorption using an ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA).

2.3. Transgenic construct sequence detection using real-time PCR

DNA sequences coding for NPTII and regulatory elements in transgenic constructs, P35S and NOST, were targeted for real-time PCR detection to monitor papaya food commodities for GM papaya use. Detection of *chymopapain* (*Chy*, GenBank accession No. AY803756) in the papaya genome was used for endogenous gene detection. All nucleotide sequences of primers and probes used are listed in Table 1.

Each reaction contained 12.5 μ L TaqMan Gene Expression Master Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA), 0.8 μ M each primer, 0.1 μ M probe and 50 ng DNA in a final volume of 25 μ L. Each sample was tested in duplicate. PCR was performed in MicroAmp Optical 96-well Reaction Plates (Thermo Fisher Scientific Inc.). An ABI PRISM 7900HT instrument was used for the analyses. The PCR conditions were 2 min at 50 °C, 95 °C for 10 min followed by 50 cycles of 15 s at 95 °C and 1 min at 60 °C. After completion of the amplification reaction, data were analyzed using SDS 2.1 sequence detection software (Thermo Fisher Scientific Inc.) for the ABI PRISM 7900HT Sequence Detection System. The Δ Rn threshold was set to 0.2 during exponential amplification and threshold cycle (Ct) values were recorded. When both an exponential increase in fluorescence intensity of dye from the target (6-carboxyfluorescein) and a Ct value <48.00 were confirmed, the sample was scored as positive. If no Ct value at the threshold value of 0.2 could be obtained, the reaction was scored as negative. Also, reactions with Ct values <48.00 but without exponential amplification, as judged by visual inspection of respective Δ Rn and multi-component plots, were scored as negative.

2.4. Whole genome sequencing and data analyses

Purified DNA samples from fresh papaya fruit were used to prepare DNA libraries. DNA (250 ng) was fragmented using NEBNext dsDNA Fragmentase (New England Biolabs Japan, Inc., Tokyo, Japan). DNA libraries were prepared using NEBNext Ultra DNA Library Preparation Kit (New England Biolabs Japan, Inc.). The quality of the constructed libraries was confirmed using a Qubit DNA HS Kit (Thermo Fisher Scientific Inc.) and Bioanalyzer DNA High Sensitivity Kit (Agilent Technologies, Santa Clara, CA, USA). A single whole genome sequencing was run by the MiSeq sequencing system (Illumina, San Diego, CA, USA) using a Reagent V3 (600 cycles) Kit. The obtained reads were trimmed at quality value (QV) 20 using Sickle ver. 1.290 software (<https://github.com/najoshi/sickle>). Sequencing and quality trimming of low quality reads resulted in a data set containing 18,177,038 paired-end reads (10,906,222,800 bases). Within the data set, a total of 17,375,285 paired-end reads having more than 50 bases at QV 20 were mapped using Bowtie 2 software (Langmead & Salzberg, 2012) against the papaya genome sequence (GenBank assembly accession: GCA_000150535.1) (Ming et al., 2008) as a reference. Mapped paired reads were extracted and assembled together using Velvet (Zerbino & Birney, 2008) at *k*-mer = 61. PRICE contig extension software (Ruby, Bellare, & DeRisi, 2013) was used to iteratively increase the size of existing contigs. Sequence reads were mapped to the reference genome sequences and visualized using Integrative Genomics Viewer version 2.3.40 (Robinson et al., 2011).

2.5. Analysis of transgenic construct sequences in GM papaya

DNA was PCR-amplified using oligonucleotide primers designed using the obtained genome sequences. PCR reactions were

Table 1
Oligonucleotide primers and probes used in this study.

Detection method	Target sequence	Oligonucleotide name	Nucleotide sequence (5' → 3')	Amplicon (bp)	Reference
<i>Chy</i> (real-time PCR)	Papaya endogenous <i>chymopapain</i>	Q-Chy-1F2	CCATGCGATCCTCCCA	72	Nakamura et al., 2013a,b; Guo et al., 2009
		Q-Chy-2R Q-Chy-P ^a	CATCGTAGCCATTGTAACACTAGCTAA FAM-TTCCCTTCAI(BHQ1)C CATTCCCACTCTTGAGA		
<i>NPTII</i> (real-time PCR)	<i>Neomycin phosphotransferase II</i>	pp-nptii F	CCGGCCACAGTCGATGA	64	This study
		pp-nptii R pp-nptii P	TGCCTGCTTGCCGAATATC FAM-CAGAAAAGCGGCCAT TTCCACCA-TAMRA		
NOST (real-time PCR)	<i>Nopaline synthase</i> terminator	180-F	CATGTAATGCATGACGTTATTTATG	84	Waiblinger et al., 2008
		180-R TM-180	TTGTTTTCTATCGCGTATTAAATGT FAM-ATGGGTTTTTATGAT TAGAGTCCCGCAA-TAMRA		
P35S (real-time PCR)	<i>Cauliflower mosaic virus 35S RNA</i> promoter	35S-F	GCCTCTGCCGACAGTGGT	82	Waiblinger et al., 2008
		35S-R 35S-P	AAGACGTGGTTGGAACGTCTTC FAM-CAAAGATGGACCCCC ACCCACG-TAMRA		
GM papaya screening (qualitative PCR)	Contig_1(P35S, PRSV CP)	p981	ATGACAAGAAGAAATCTTCG	1482	This study
	Contig_344501(PRSV CP)	p982	CTATATTCTAATGCTTACATAG		
Event-specific for the line PRSV-YK (qualitative PCR)	T-DNA (NOS promoter)	Papa27	GCGTCATCGCGGGGGTCATAA	See Supplementary Table 3 for details	Fan et al., 2009
	Papaya genome DNA (Right-border flanking)	Papa31	TTGTTCTAATAAGTTGCTAC		
	Papaya genome DNA (Right-border flanking)	Papa32	AATATCAAATGGACGTGTTAGTG		
	T-DNA (Left border)	Papa56	GTTATTAAGTTGCTAAGCGTCAA		
	Papaya genome DNA (Left-border flanking)	Papa57	AGACATATATCATCAAGACCATAGTAG		
	T-DNA (Left border)	Papa58	GTCTTGCGATGATTATCAT		
	Papaya genome DNA (Left-border flanking)	Papa59	TGGTTATCAATATAGCAATTATGTAG		
PRSV-YK (real-time PCR)	Boundary sequence between <i>Papaya ringspot virus</i> coat protein gene and <i>Cauliflower mosaic virus 35S RNA</i> promoter in the GM papaya line PRSV-YK	YK-2F	ACACGGGGGACTCTAGAG	76	This study
		YK-2R	ACCGGTATCCACAGCTTC		
		YK-2P	FAM-TCCCTTCATGGCGTC-TAMRA		

^a Black-hole quencher 1 (BHQ1) of underlined T base was used.

prepared as follows. Briefly, the PCR reaction mixture (50 µL) contained 5 µL 10× *PfuUltra* HF reaction buffer (Agilent Technologies), 0.2 mM dNTP (Agilent Technologies), 0.5 µM forward and reverse primers (p981 and p982, respectively), 2.5 U *PfuUltra* HF DNA polymerase (Agilent Technologies) and 50 ng template DNA. PCR conditions were as follows: 95 °C for 2 min, followed by 30 cycles of denaturation at 95 °C for 30 s, annealing at 58 °C for 30 s, extension at 72 °C for 60 s and terminal elongation at 72 °C for 10 min. PCR was carried out using a GeneAmp PCR System 9700 (Thermo Fisher Scientific Inc.). PCR-amplified DNA fragments were extracted from agarose gels and purified using a QIAquick PCR Purification Kit (QIAGEN). Nucleotide sequences of both strands were analyzed using an ABI PRISM 3700 DNA analyzer and a BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific Inc.). To detect transgenic construct-specific sequence, DNA sequences between P35S and the PRSV coat protein (PRSV CP in YK strain, GenBank accession No. X97251) were targeted. To detect event-specific sequences, qualitative PCR primer sets were designed to amplify DNA sequences between genomic and T-DNA sequences in GM papaya line 16-0-1/17-0-5, as reported previously (Fan et al., 2009). Nucleotide sequences of all primers used for PCR are listed in Table 1.

3. Results

3.1. Qualitative monitoring test for GM papaya use

To qualitatively confirm for presence of GM papaya in papaya food commodities obtained (pickled papaya, papaya-leaf tea, jam, fresh fruit), commonly used DNA sequences (*NPTII*, NOST and P35S) in GM papaya transgenic constructs (Mendoza et al., 2008) were targeted for real-time PCR detection. Ct values (*Chy* detection method, 22.58; *NPTII* detection method, 25.97; NOST detection method, 24.17; P35S detection method, 25.52) were obtained from a fresh papaya fruit (Fig. 1, Supplementary Table 1). In addition, other papaya food commodities were tested, and targeted sequences (*NPTII*, NOST and P35S) were amplified (Ct value <48.00) in duplicate DNA preparations from five other independently packaged commodities (two pickled papayas, two papaya-leaf teas and one jam) (Supplementary Table 2). DNA purified from non-GM papaya (Sunset) and water, used as a no-DNA control sample, were negative in the real-time PCR detection tests. The non-GM papaya (Sunset) was only positive at a Ct value of 23.18 using the *Chy* detection method. To identify authorized and unauthorized GM papaya, further DNA sequence information on the

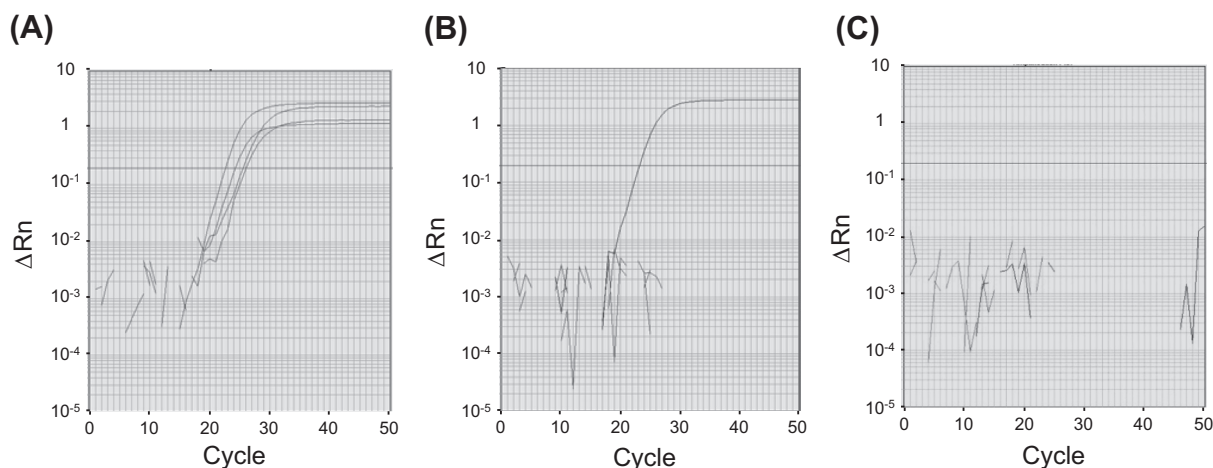


Fig. 1. Representative real-time PCR amplification plots for detecting transgenic construct sequences. Sets of primers and probe for detecting endogenous gene (*Chy*) and transgenic construct sequences (*NPTII*, *NOST*, *P35S*) were used for real-time PCR. Oligonucleotides used are listed in Table 1. DNA samples (25 ng) purified from (A) a fresh papaya fruit commodity and (B) non-GM papaya (cultivar Sunset) were used as templates. (C) Water was used as a no-template control.

transgenic construct-specific sequences inserted in the genome and the border sequences of the transgenic construct and papaya genome, which are unique to each GM papaya line (event-specific sequences), is required.

3.2. Whole genome sequence analysis of GM papaya

To analyze transgenic construct-specific and event-specific sequences in the GM papaya detected, paired-end whole genome sequencing libraries (400–500 bp) were generated using DNA purified from the unidentified GM-positive fresh papaya fruit obtained in the real-time PCR monitoring test, and sequenced. Among the quality-trimmed paired-end reads, 49.84% sequences were successfully mapped to the reference using Bowtie 2 (Fig. 2A). Some unmapped sequences were identified as *Enterobacteriaceae* genome sequences using a BLASTn search. The 411,911 contigs (267,082,851 bases) resulting from *de novo* assembly of the mapped reads were used to search for transgenic construct- and event-specific sequences in GM papaya.

We first searched for P35S sequences commonly used as a promoter in GM papaya (Mendoza et al., 2008). Contig_1 (719 bp) was found with a sequence 99.8% homologous to a 445-bp P35S in a binary vector used for *Agrobacterium* transformation of plants (GenBank accession No. EU327975), and a 68-bp 3'-sequence of Contig_1 was homologous to Contig_344501 (780 bp) in the *de novo* assembled sequences (total of 411,911 contigs). *PRSV CP* in Contig_1 and Contig_344501 had a 68-bp sequence homologous to the genome sequence coding for *PRSV CP* in a *PRSV* isolate from Taiwan (GenBank accession No. X97251) (Fig. 2B). Genome-wide sequence analysis thus indicated the presence of a transgenic construct for P35S-driven *PRSV CP* expression.

3.3. Detection of transgenic construct sequence in GM papaya

To confirm the presence of the identified transgenic construct sequence, a qualitative PCR primer set (p981 and p982, with p981 designed to the 5'-terminal sequence of Contig_1 and p982 to the 3'-terminus of Contig_344501) was used to amplify border sequences between the P35S and the *PRSV CP* sequences (Fig. 2B). Agarose-gel electrophoresis of the PCR product showed a 1482 bp amplicon, whereas no PCR products were amplified from the negative controls (water or non-GM papaya template DNA) (Fig. 3A). BLASTn analysis of the obtained sequence indicated 446 bp coding for P35S and 992 bp coding for *PRSV CP* (Fig. 3B).

Multiple cloning sites (*Bam*HI and *Nco*I restriction sites) and mutations coding for two amino acids, methionine and alanine, upstream of *PRSV CP* were also detected. The *PRSV CP* sequence was identical to the gene in *PRSV* strains isolated in Taiwan, such as YK (GenBank accession No. X97251) (Wang & Yeh, 1997). The sequences were identical to transgenic construct sequences of the GM papaya line PRSV-YK (line PRSV-YK) found in some processed papaya commodities (papaya-leaf tea, papaya jam, papaya pickles) in Japan in 2011 (Nakamura et al., 2011).

3.4. Identification of line PRSV-YK from whole genome sequence data

Using whole genome sequence data, we next searched for papaya genome sequences flanking the transgene integration site. *Agrobacterium* transformation is the most common method to develop GM papaya (Mendoza et al., 2008). Therefore, left- and right-border sequences in *Agrobacterium* T-DNA transformation vectors were used as known sequences to search for unknown border sequences.

A portion of the highly conserved left-border T-DNA sequence (17 bp, 5'-TGTTTACACCACAATAT-3') (Kim et al., 2003) was used to search the sequence read data. No matching sequence was obtained in the *de novo* assembled data, while one paired read (A5FPB:1:1114:21442:18611) in the unmapped reads matched (Supplementary Fig. 1). This read (419 bp) contained flanking sequence (1–132 bp), homologous to papaya genome sequence in Supercontig_16 (GenBank accession No. DS981536) in the database, and to transgenic sequences including T-DNA left-border sequence (133–419 bp). Event-specific sequences were also detected using qualitative PCR methods, and the pattern and sequence of the amplified PCR products indicates that the GM papaya is identical to that reported for event 16-0-1/17-0-5 described previously (Fan et al., 2009) (Supplementary Figs. 2 and 3, Supplementary Table 3).

Analysis of the T-DNA right-border sequence did not match any mapped or unmapped reads, unlike the T-DNA left-border sequence. Instead, a 19-bp event-specific sequence (5'-TCAGTGT GAATGAGATAG-3', Supplementary Fig. 3) matched Contig_446121 (1740 bp) with 100% homology (Supplementary Fig. 4). The contig had the T-DNA right-border sequence of the transgenic construct (1–836 bp) and papaya genome sequence (837–1,740 bp). The papaya genome sequence was similar (99% homology) to part of Supercontig_16 (GenBank accession No. DS981536). These results

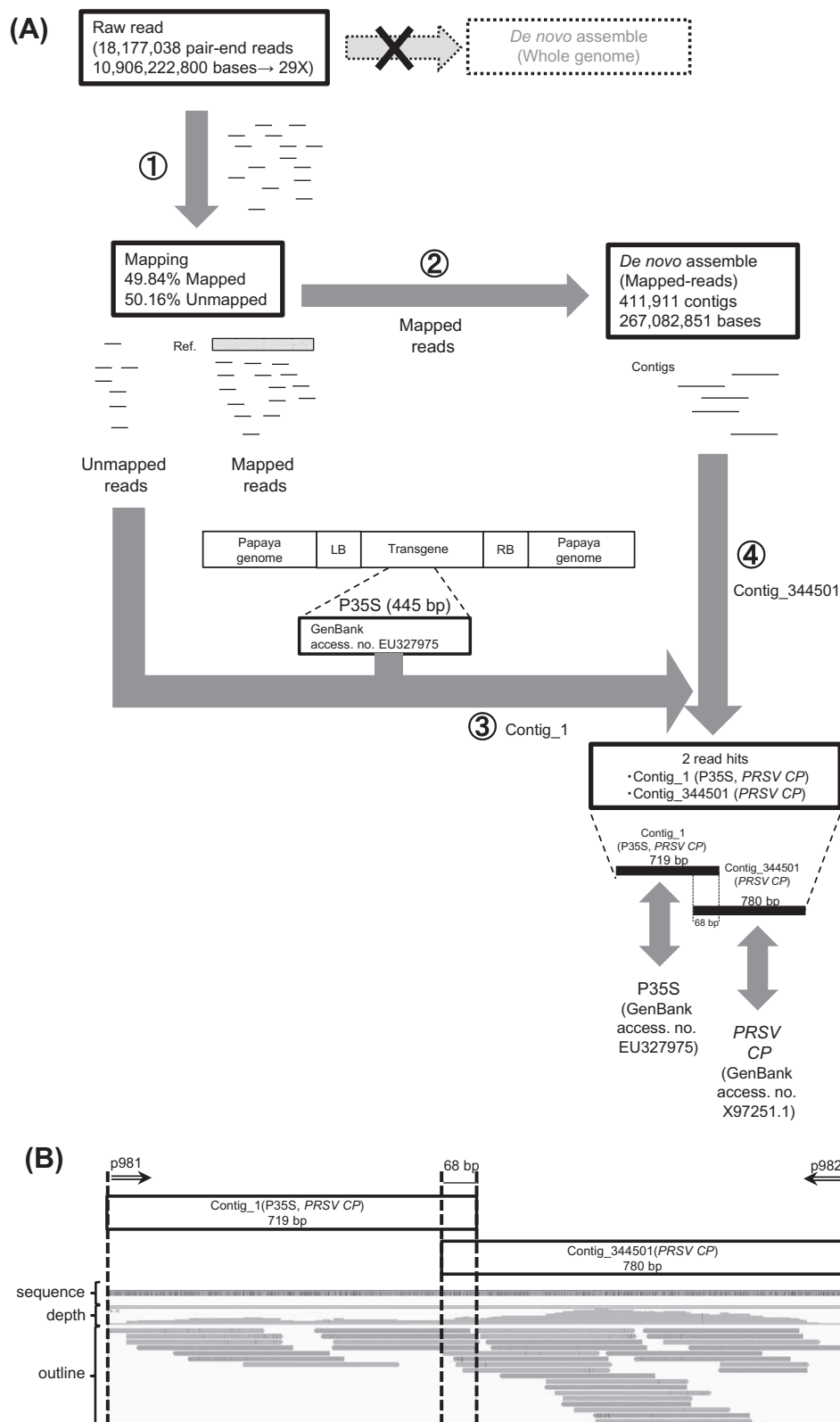


Fig. 2. Genome-wide sequence analysis of unknown GM papaya. (A) An Illumina Miseq sequencer was used for analysis of a DNA sample purified from fresh papaya fruit. A single run produced 18,177,038 M paired reads (total of 10,906,222,800 bases), which is equivalent to a whole papaya genomic DNA coverage estimate of about 29. ① The data output was mapped against the papaya genome reference sequence (Ming et al., 2008). Of all reads, 49.84% were grouped mapped reads and 50.16% were unmapped reads. ② *De novo* assembly of the mapped paired reads produced 411,911 contigs (267,082,851 bases). ③ BLASTn analysis of unmapped reads found one read (Contig_1) with 99.8% sequence homology to a 445-bp P35S sequence in a binary vector used for *Agrobacterium* transfection (GenBank access. No. EU327975). ④ Homology search of Contig_1 to the *de novo* assembled (mapped reads) data found one read (Contig_344501) having a 68-base sequence homologous to PRSV CP. (B) All output reads for Contig_1 and Contig_344501 were aligned. Sequences (68 bp) homologous between Contig_1 and Contig_344501 are shown. A qualitative PCR primer pair (p981/p982) that was used is shown by arrows.

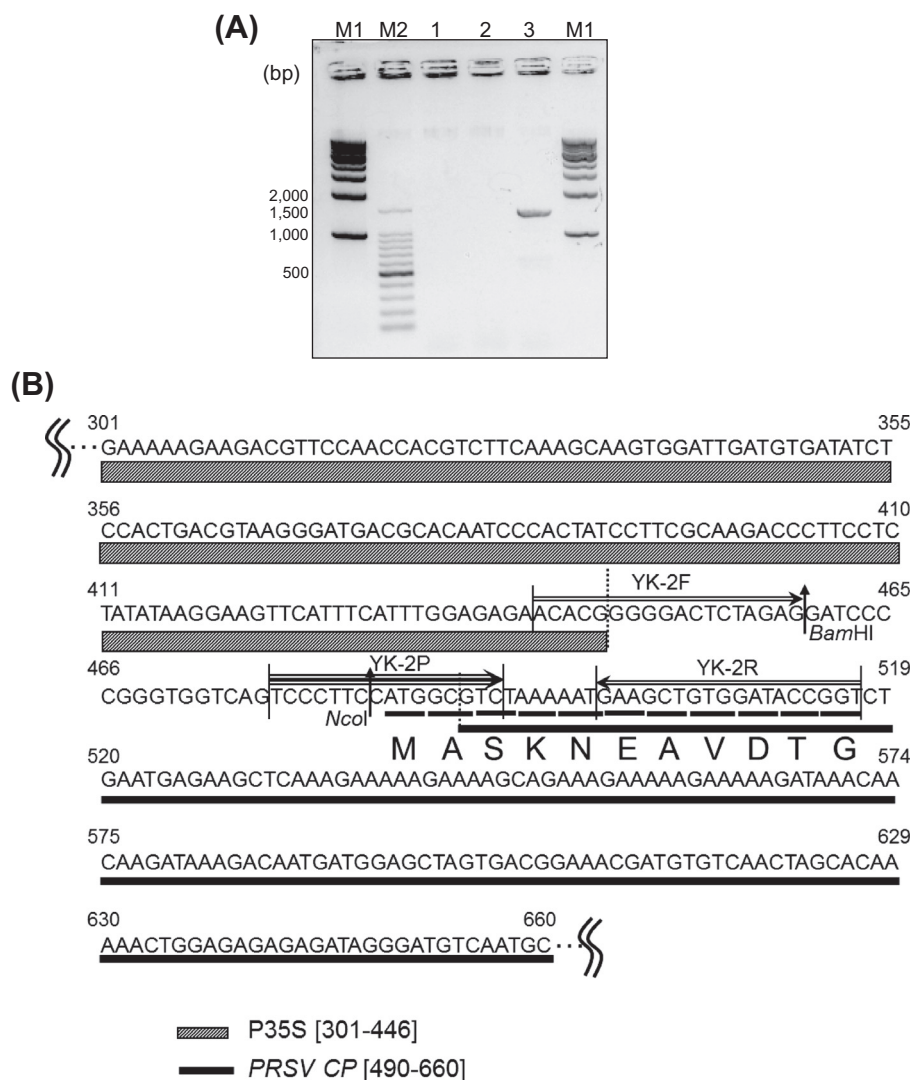


Fig. 3. Detection of line PRSV-YK in DNA purified from a papaya food commodity. (A) Agarose gel electrophoresis of PCR products obtained using a primer set (p981, forward primer; p982, reverse primer). Lane M1, 1 kbp DNA ladder marker; lane M2, 100 bp DNA ladder marker; lane 1, no-template control (water); lane 2, non-GM papaya (cultivar Sunset); lane 3, fresh papaya fruit commodity. (B) Part of the amplicon sequence (1482 bp) obtained from the fresh papaya fruit commodity. Position of real-time PCR primer/probe sets for detection is indicated by horizontal arrows above the DNA sequences. Amino acid sequences of the translated PRSV CP are indicated under the DNA sequences. Digestion sites of the restriction enzymes in transgenic construct are indicated by vertical arrows.

suggest that genome-wide sequencing enables analyses of transgenic constructs and event-specific sequences in GM papaya.

3.5. Development of detection method for GM papaya line PRSV-YK

Genome-wide sequence analysis revealed transgenic construct- and event-specific sequence in the line PRSV-YK. Event-specific sequences can be used to specifically detect a GM papaya line. Here, a detection method for all GM papaya lines having the transgenic construct of the one in line PRSV-YK was developed. We previously reported a transgenic construct-specific detection method for line PRSV-YK using PCR-based sequence data (Nakamura et al., 2011). In this study, we show the development of a reliable detection method for line PRSV-YK using whole genome sequence analysis. A set of primers and probe for the real-time PCR method was designed at the border sequence between P35S and PRSV-CP (Fig. 3B). This real-time PCR method consisted of detecting two target sequences, the transgenic construct-specific sequence for detection of line PRSV-YK (PRSV-YK detection method) and a reference control (*Chy* detection method) (Table 1). A BLASTn search of

the designed primers and probe using whole genome sequence data indicated the specificity for detection of line PRSV-YK. Neither the PRSV-YK nor the *Chy* method found any nonspecific amplification using DNA from rice, soybean, maize, potato, rapeseed, pineapple, peach or passion fruit, and only the *Chy* detection method showed specific amplification of papaya DNA (Fig. 1 and Table 1 in Nakamura et al., 2016).

To evaluate the sensitivity of the method, test samples of DNA purified from line PRSV-YK were serially diluted in DNA from non-GM papaya. Primers and probe for both PRSV-YK and *Chy* detection were used to estimate the limit of detection (LOD) from 21 replicates of the real-time PCR run. The PRSV-YK detection method was quantitative ($R^2 = 0.99$) for Ct values on amplification plots in the range of 0.010–100% of PRSV-YK DNA (Fig. 2 in Nakamura et al., 2016). According to the international standard for qualitative PCR detection, LOD should have a false-negative rate of $\leq 5\%$ and the relative standard deviation (RSD) of reproducibility should be $\leq 33\%$ (Codex Alimentarius Commission, 2005; International Organization for Standardization, 2006). Primer/probe sets for the PRSV-YK and *Chy* detection methods

Table 2
Detection limit of PRSV-YK and *Chy* detection methods.

Line PRSV-YK DNA content (w/w [%])	Signal rate ^a		Mean Ct value		SD		RSD (%)	
	PRSV-YK	<i>Chy</i>	PRSV-YK	<i>Chy</i>	PRSV-YK	<i>Chy</i>	PRSV-YK	<i>Chy</i>
0.100	21/21	21/21	37.37	22.34	0.33	0.25	0.88	1.12
0.020	21/21	21/21	40.62	22.40	0.85	0.36	2.09	1.59
0.010	21/21	21/21	41.48	22.42	0.89	0.42	2.15	1.85
0.001	4/21	21/21		22.36		0.39		1.73

SD, standard deviation; RSD, relative standard deviation.

^a Signal rate is indicated by number of positive signals out of 21 experimental trials.

Table 3
Variance analysis of Ct values.

Line PRSV-YK DNA content (w/w [%])	Detection method				
	PRSV-YK		<i>Chy</i>		
	0.05	0.10	0.00	0.05	0.10
Mean	38.29	37.39	22.66	22.59	22.58
RSD (%)	1.60	1.63	0.84	0.90	0.83
RSD _r (%)	1.48	1.37	0.40	0.41	0.44
RSD _R (%)	2.17	2.62	1.92	2.08	1.84

allowed detection of line PRSV-YK in all 21 repeated tests when the concentration of line PRSV-YK is 0.010% (w/w). The highest RSD of the Ct values was 2.15% at 0.010% DNA. These results indicate that the LOD is 0.010% for detection of GM papaya line PRSV-YK (Table 2).

According to the internationally accepted harmonized protocol for interlaboratory collaborative study procedures (AOAC International, 2012), the repeatability and the reproducibility of the PRSV-YK detection method using real-time PCR were analyzed (Nakamura et al., 2016). Both the PRSV-YK and *Chy* detection methods had low RSD, repeatability RSD (RSD_r) and reproducibility RSD (RSD_R), which are equivalent to a Ct value of less than 1 (Table 3).

4. Discussion

In this study, papaya food commodities were screened for contamination with unauthorized GM food by a real-time PCR method targeting common elements (*NPTII*, *NOST* and *P35S*) in transgenic constructs for GM papaya. Whole genome sequence analysis was performed using DNA purified from one of the commodities (fresh papaya fruit) that was confirmed positive for GM papaya in a screening test. This GM papaya was detected as GM papaya (designated line PRSV-YK) using transgenic construct-specific real-time PCR method, which was described previously (Nakamura et al., 2011). This real-time PCR method was developed from the restricted DNA sequence information, and repeatability and reproducibility of the method had to be improved. An Illumina MiSeq sequencer was used to analyze the transgenic construct and event-specific sequences for the GM papaya. A single-run output of data was enough to search for low-copy transgenic construct and event-specific sequences. Line PRSV-YK was identified as GM papaya event 16-0-1/17-0-5 (Fan et al., 2009), which has not yet authorized in Japan. The DNA sequences obtained were used to develop a reliable real-time PCR detection method for line PRSV-YK. This developed real-time PCR method was transgenic construct-specific, repeatable and reproducible for detection of the GM papaya, as was confirmed by an interlaboratory validation study among 12 independent GM testing laboratories.

Detection of *NOST* and *P35S* is widely used to monitor food commodities for unauthorized GM food contamination in the EU

and Japan (Akiyama et al., 2009; Barbau-Piednoir et al., 2010; Waiblinger, Ernst, Anderson, & Pietsch, 2008). Identification of GM papaya and the development of a specific detection method require additional sequence information obtained from methods such as chromosome walking and inverse PCR. In this study, transgenic construct- and event-specific sequences were searched in whole genome sequence data, and the GM papaya line was identified.

A short known sequence could be used to identify unknown sequences. Our results indicate that transgenic construct-specific sequences such as the most common promoter for transgenic constructs for GM papaya, *P35S* (445 bp), can be used to search whole genome sequence data. Also, in the sequencing data, we identified event-specific sequences of line PRSV-YK using 17-bp border regions from T-DNA used for transformation. From the obtained genome sequence data, we developed a specific detection method for the GM papaya.

There are several next-generation sequencers available (Pabinger et al., 2014). The Illumina MiSeq sequencer generated 18,177,038 paired reads (total of 10,906,222,800 bases), which are equivalent to approximately 29 times' coverage of the papaya genome sequence available in the database. Higher throughput next-generation sequencers that provide longer base reads may be more advantageous to identify transgenic construct- and event-specific sequences.

From the identified transgenic construct sequence, a method for detection of line PRSV-YK was developed using real-time PCR. This method can detect all unauthorized GM papaya lines having a transgenic construct like line PRSV-YK. The interlaboratory validation study, which followed the international harmonized protocol for developed methods, indicated that the method is specific, repeatable and reproducible. The method detects line PRSV-YK at a concentration >0.010% (w/w). These results indicate that the method is suitable for specific detection of line PRSV-YK in various papaya food commodities. Using a duplicate preparation of the DNA purified from 14 papaya food commodities, we identified six independently packaged commodities (two pickled papaya samples, two papaya-leaf teas, one jam, and one fresh papaya fruit) positive for *Chy*, *NPTII*, *NOST* and *P35S*, and the developed PRSV-YK detection method confirmed that all of these commodities contained the transgenic construct of line PRSV-YK (Supplementary Table 2). Thus, this real-time PCR method can be adopted by laboratories for GM papaya regulatory purposes.

In conclusion, we showed that whole genome sequence data enable identification of unknown transgenic construct-specific and event-specific sequences in GM papaya using common sequences >17 bp as known sequences. Analysis of processed papaya commodities is underway to verify that whole genome sequencing of degraded short papaya DNA fragments in various food ingredients can be used for identification of transgenic constructs and GM event-specific sequences. The applicability of this method to processed foods will help in the immediate identification of unknown GM foods.

Financial support

This study was supported by a Grant (H25-Shokuhin-Ippan-015) from the Ministry of Health, Labour and Welfare of Japan and JSPS KAKENHI Grant No. (26860091).

Appendix A. Supplementary data

The sequence data of this article have been deposited in the DDBJ Sequenced Read Archive under accession No. PRJDB3976. Nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under the accession Nos. LC073698 (for Contig_1), LC073699 (for Contig_344501), LC073700 (for Contig_446121), LC073701 (for M0088466000000000-A5FPB11142144218611_1N01), LC073702 (for M0088466000000000-A5FPB11142144218611_2N01) and LC073703 (for amplicon of p981 & p982). Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.foodchem.2016.02.157>.

References

- Akiyama, H., Nakamura, F., Yamada, C., Nakamura, K., Nakajima, O., Kawakami, H., et al. (2009). A screening method for the detection of the 35S promoter and the nopaline synthase terminator in genetically modified organisms in a real-time multiplex polymerase chain reaction using high-resolution melting-curve analysis. *Biological & Pharmaceutical Bulletin*, 32, 1824–1829.
- Akiyama, H., Sugimoto, K., Matsumoto, M., Isuzugawa, K., Shibuya, M., Goda, Y., & Toyoda, M. (2002a). A detection method of recombinant DNA from genetically modified potato (NewLeaf Plus potato) and detection of NewLeaf Plus potato in snack. *Shokuhin Eiseigaku Zasshi*, 43, 24–29.
- Akiyama, H., Watanabe, T., Wakui, C., Chiba, Y., Shibuya, M., Goda, Y., & Toyoda, M. (2002b). A detection method of recombinant DNA from genetically modified potato (NewLeaf Y potato). *Shokuhin Eiseigaku Zasshi*, 43, 301–305.
- AOAC International (2012). Official methods of analysis of AOAC International. In G. W. Latimer (Ed.), *Appendix D: Guidelines for collaborative study procedures to validate characteristics of a method of analysis*. Gaithersburg, MD, USA.
- Barbau-Piednoir, E., Lievens, A., Mbongolo-Mbella, G., Roosens, N., Sneyers, M., Leunda-Casi, A., & Van den Bulcke, M. (2010). SYBR Green qPCR screening methods for the presence of “35S promoter” and “NOS terminator” elements in food and feed products. *European Food Research and Technology*, 230, 383–393.
- Codex Alimentarius Commission (2005). Consideration of the methods for the detection and identification of foods derived from biotechnology, general approach and criteria for the methods. In *Codex committee on methods of analysis and sampling (CX/MAS 05/26/9)*. Budapest, Hungary.
- Fan, M. J., Chen, S., Kung, Y. J., Cheng, Y. H., Bau, H. J., Su, T. T., & Yeh, S. D. (2009). Transgene-specific and event-specific molecular markers for characterization of transgenic papaya lines resistant to *Papaya ringspot virus*. *Transgenic Research*, 18, 971–986.
- Fermin, G., Keith, R. C., Suzuki, J. Y., Ferreira, S. A., Gaskill, D. A., Pitz, K. Y., Manshardt, R. M., et al. (2011). Allergenicity assessment of the *Papaya ringspot virus* coat protein expressed in transgenic rainbow papaya. *Journal of Agricultural and Food Chemistry*, 59, 10006–10012.
- Grohmann, L., Reiting, R., Made, D., Uhlig, S., Simon, K., Frost, K., et al. (2015). Collaborative trial validation of cry1Ab/Ac and Pubi-cry TaqMan-based real-time PCR assays for detection of DNA derived from genetically modified Bt plant products. *Accreditation and Quality Assurance*, 20, 85–96.
- Guo, J., Yang, L., Liu, X., Guan, X., Jiang, L., & Zhang, D. (2009). Characterization of the exogenous insert and development of event-specific PCR detection methods for genetically modified Huanong No. 1 papaya. *Journal of Agricultural and Food Chemistry*, 57, 7205–7212.
- International Organization for Standardization (2006). *Foodstuffs-Methods of analysis for the detection of genetically modified organisms and derived products-general requirements and definitions (ISO 24276)*. Geneva: Switzerland.
- James, C. (2014). Global Status of Commercialized Biotech/GM Crops: 2014. ISAAA Briefs No. 49. ISAAA: Ithaca, NY, USA.
- Kim, S. R., Lee, J., Jun, S. H., Park, S., Kang, H. G., Kwon, S., & An, G. (2003). Transgene structures in T-DNA-inserted rice plants. *Plant Molecular Biology*, 52, 761–773.
- Kim, J. H., Park, S. B., Roh, H. J., Park, S., Shin, M. K., Moon, G. L., et al. (2015). A simplified and accurate detection of the genetically modified wheat MON71800 with on calibrator plasmid. *Food Chemistry*, 176, 1–6.
- Kitagawa, M., Nakamura, K., Kondo, K., Ubukata, S., & Akiyama, H. (2014). Examination of processed vegetable foods for the presence of common DNA sequences of genetically modified tomatoes. *Shokuhin Eiseigaku Zasshi*, 55, 247–253.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357–359.
- Lin, H. T., Yen, G. C., Huang, T. T., Chan, L. F., Cheng, Y. H., Wu, J. H., et al. (2013). Toxicity assessment of transgenic *Papaya ringspot virus* of 823–2210 line papaya fruits. *Journal of Agricultural and Food Chemistry*, 61, 1585–1596.
- Lin, H. T., Yen, G. C., Lee, W. C., Tsai, Y. T., Wu, J. H., Yeh, S. D., et al. (2015). Repeated dose 90-day feeding study of whole fruits of genetically modified papaya resistant to *Papaya ringspot virus* in rats. *Journal of Agricultural and Food Chemistry*, 63, 1286–1292.
- Mano, J., Shigemitsu, N., Futo, S., Akiyama, H., Teshima, R., Hino, A., et al. (2009). Real-time PCR array as a universal platform for the detection of genetically modified crops and its application in identifying unapproved genetically modified crops in Japan. *Journal of Agricultural and Food Chemistry*, 57, 26–37.
- Mendoza, E. M. T., Laurena, A. C., & Botella, J. R. (2008). Recent advances in the development of transgenic papaya technology. *Biotechnology Annual Review*, 14, 423–462.
- Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J. H., et al. (2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature*, 452, 991–996.
- Nakamura, K., Akiyama, H., Kawano, N., Kobayashi, T., Yoshimatsu, K., Mano, J., et al. (2013a). Evaluation of real-time PCR detection methods for detecting rice products contaminated by rice genetically modified with a CpTI-KDEL-T-nos transgenic construct. *Food Chemistry*, 141, 2618–2624.
- Nakamura, K., Akiyama, H., Ohmori, K., Takahashi, Y., Takabatake, R., Kitta, K., et al. (2011). Identification and detection method for genetically modified papaya resistant to *Papaya ringspot virus* YK strain. *Biological & Pharmaceutical Bulletin*, 34, 1648–1651.
- Nakamura, K., Akiyama, H., Takahashi, Y., Kobayashi, T., Noguchi, A., Ohmori, K., et al. (2013b). Application of a qualitative and quantitative real-time polymerase chain reaction method for detecting genetically modified papaya line 55-1 in papaya products. *Food Chemistry*, 136, 895–901.
- Nakamura, K., Kondo, K., Akiyama, H., Ishigaki, T., Noguchi, A., Katsumata, H., et al. (2016). Interlaboratory validation study of real-time polymerase chain reaction detection method for unauthorized genetically modified papaya line PRSV-YK. *Data in Brief*, submitted for publication.
- Nakamura, K., Kondo, K., Kobayashi, T., Noguchi, A., Ohmori, K., Takabatake, R., et al. (2014). Identification and detection of genetically modified papaya resistant to *Papaya ringspot virus* strains in Thailand. *Biological & Pharmaceutical Bulletin*, 37, 1–5.
- Ohmori, K., Nakamura, K., Kasahara, M., Takabatake, R., Kitta, K., Fujimaki, T., et al. (2013). A DNA extraction and purification method using an ion-exchange resin type kit for the detection of genetically modified papaya in processed papaya products. *Food Control*, 32, 728–735.
- Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., et al. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in Bioinformatics*, 15, 256–278.
- RASFF Portal. Unauthorised genetically modified papaya from unknown origin. Notification details-2015.0589 (Notification date: 13/05/2015).
- Regulation (EC) No 1829/2003 of the European Parliament and of the Council. (September 22, 2003).
- Reiting, R., Broll, H., Waiblinger, H. U., & Grohmann, L. (2007). Collaborative study of a T-nos real-time PCR method for screening of genetically modified organisms in food products. *Journal für Verbraucherschutz und Lebensmittelsicherheit*, 2, 116–121.
- Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29, 24–26.
- Ruby, J. G., Bellare, P., & DeRisi, J. L. (2013). PRICE: software for the targeted assembly of components of (meta) genomic sequence data. *G3: Genes, Genomes, Genetics*, 3, 865–880.
- Standards for the Safety Assessment of GM Foods (Seed Plants), the Food Safety Commission of Japan. (January 29, 2004).
- Waiblinger, H. U., Ernst, B., Anderson, A., & Pietsch, K. (2008). Validation and collaborative study of a P35S and T-nos duplex real-time PCR screening method to detect genetically modified organisms in food products. *European Food Research and Technology*, 226, 1221–1228.
- Wang, C. H., & Yeh, S. D. (1997). Divergence and conservation of the genomic RNAs of Taiwan and Hawaii strains of *Papaya ringspot potyvirus*. *Archives of Virology*, 142, 271–285.
- Young, L., Hamerlindl, J., Babic, V., McLeod, J., Sharpe, A., Matsalla, C., et al. (2015). Genetics, structure, and prevalence of FP967 (CDC Triffid) T-DNA in flax. *Springerplus*, 4, 146.
- Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18, 821–829.