Compilation of Study Reports

-

Studies Submitted in Support of the Food Safety Assessment of Fusarium Wilt Tropical Race 4 Resistant Banana Event QCAV-4

Submitted to:

Standards Management Officer
Food Standards Australia New Zealand
Level 4, 15 Lancaster Place
Majura ACT 2609
AUSTRALIA

Submitting Institution:

[redacted]

Contact information:

s 22
[redacted]

[redacted]
[redacted] 2023

[redacted]

Good laboratory practice compliance statement

Although some of the studies presented in this document were conducted in laboratories accredited by the National Association of Testing Authorities (NATA) and in compliance with the OECD Good Laboratory Practice Standards (GLP), most were conducted in non-NATA-accredited laboratories at Queensland University of Technology. Although the QUT-based studies were not conducted in strict compliance with GLP, they were conducted according to accepted scientific methods and responsible practices in line with the QUT Code for responsible conduct of research (MoPP D/2.6). The QUT Research Code is consistent with the Australian Code for the Responsible Conduct of Research 2018. The raw data and study records have been retained consistent with the management of research data and primary materials (MoPP D/2.8).

Table of contents

QUT2023-1: Nucleotide sequence analysis of the inserted DNA and host genomic flanking regions in event QCAV-4

STUDY IDENTIFICATION

QUT2023-1

AUTHORS

[redacted]
s 22
[redacted]

-

PERFORMING LABORATORY

[redacted]

-

[redacted]

-

[redacted]

[redacted]

[redacted] 2023

1.1 Introduction

Banana event QCAV-4 (QUT-QCAV4-6) was created by *Agrobacterium tumefaciens*-mediated transformation of banana (*Musa acuminata* subgroup Cavendish cv Grand Nain) embryogenic cells with plasmid pSAN3. This plasmid contains the *neomycin phosphotransferase* II (*npt*II) selectable marker gene from the Tn5 transposon of *Escherichia coli* strain K12 and the *MamRGA2* disease resistance (R) gene from the wild banana *Musa acuminata* ssp. *malaccensis* which confers resistance to the devastating banana fungal pathogen *Fusarium oxysporum* f. sp. *cubense* tropical race 4 (TR4) (Dale *et al*., 2017).

Southern blot analysis (Southern, 1975) of banana event QCAV-4 initially revealed the presence of multiple T-DNA integrations in this event (Dale *et al*., 2017). The objective of this study was to determine the nucleotide sequence of the inserted DNA within QCAV-4, including a portion of the 5' and 3' flanking host genomic sequences, in order to characterise the insert, demonstrate the integrity and contiguity of the functional elements, and to document any unexpected changes. To achieve this and confirm the absence of plasmid backbone integration in event QCAV-4, a whole genome sequencing and bioinformatic analysis approach was adopted. Finally, an open reading frame (ORF) analysis was conducted to investigate the possibility of any new start-to-stop ORFs being created at junctional regions with either the host genome or at positions of rearrangement, that could potentially encode sequences homologous to known allergens or toxins.

1.2 Materials and Methods

1.2.1 DNA extraction

High molecular weight genomic DNA (gDNA, average fragment size > 50 kb) was isolated from young *in vitro* leaf tissue of QCAV-4 plants using the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma-Aldrich, USA) according to the manufacturer's protocol.

1.2.2 Bioinformatic analysis

Extracted gDNA was sequenced at 150 bp paired-end (PE) reads on an Illumina Novaseq 6000 platform (Macrogen, Korea), generating 707,616,882 reads and 106.9 Gbp of data. This represented an estimated 178x coverage of the ~600 Mbp haploid Cavendish banana genome. Additional long-reads were sequenced on a PacBio Sequel II platform s 22 generating 4,923,888 reads and 75.9 Gbp of data with a read N50 of 17,973 bp. This represented ~126x coverage of the ~600 Mbp haploid Cavendish banana genome. All analyses were done in Geneious Prime® version 2022.2.1. PacBio long-reads were assembled with the Flye assembler plugin available for Geneious and Illumina short-reads were mapped to the assembled insert sequence using the Geneious mapper in custom sensitivity settings at low stringency.

1.3 Results

1.3.1 Preliminary evidence of a large integration in event QCAV-4

Initially, a whole genome Illumina short-read sequencing approach, in combination with mappings of the generated reads separately against the pSAN3 plasmid, was thought to be sufficient to allow (i) confirmation of the absence of vector backbone sequences in QCAV-4, (ii) identification of the chromosomal locations of the inserts and (iii) characterisation of the organisational structure of the insert(s). Mapping of all 707,616,882 Illumina sequencing reads obtained against the sequence of pSAN3 resulted in 16,104 reads mapping between the T-DNA left border (LB) and right border (RB) repeats of pSAN3 (Figure SR1-1, A). Only five reads mapped to the pSAN3 backbone region which were deemed to

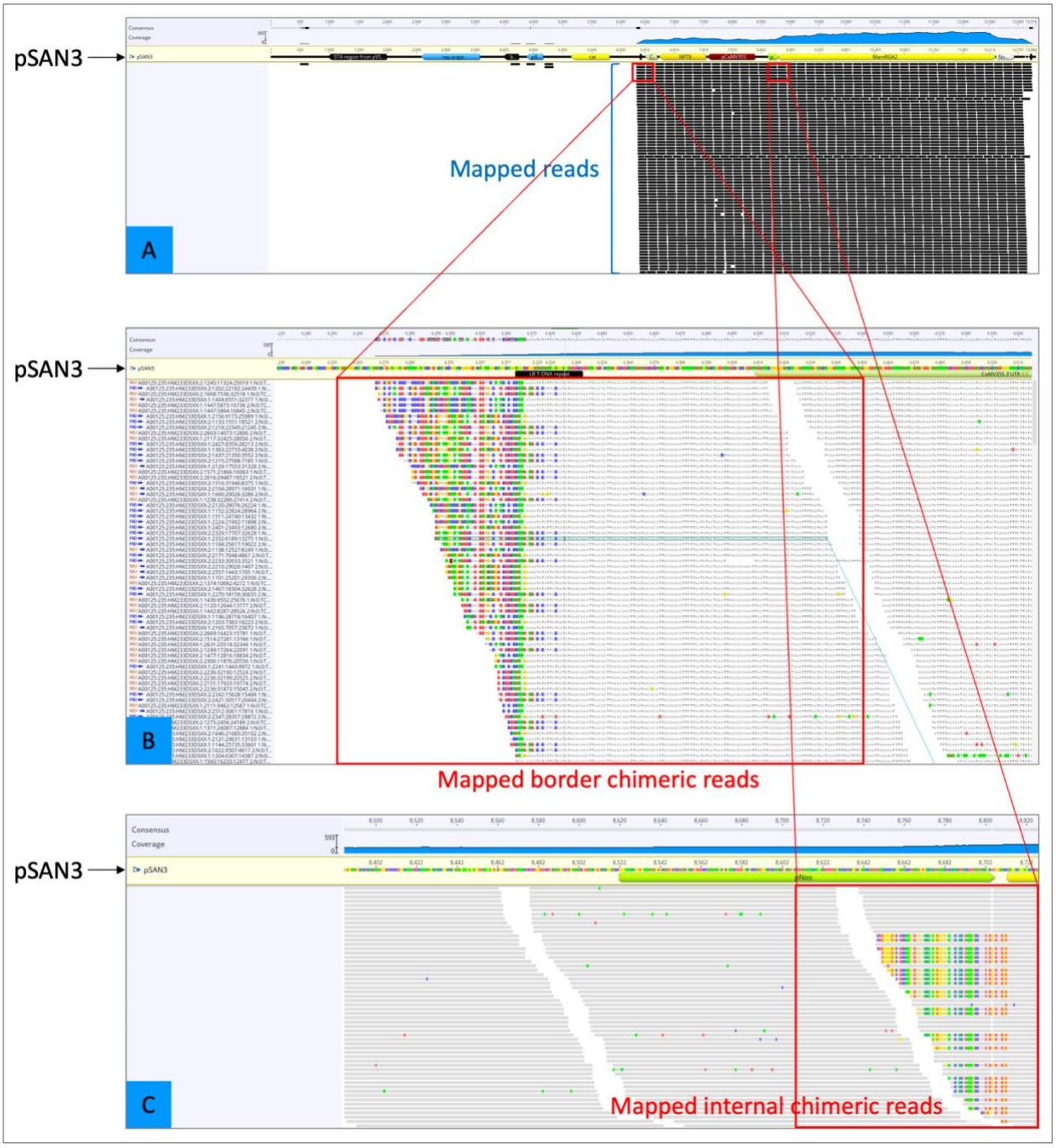


Figure SR1-1. Example of mapping results.

be due to the presence of endogenous bacteria in the original sample. This analysis also confirmed the absence of vector backbone sequences in QCAV-4.

By relaxing the mapping parameters in the previous step, chimeric reads containing sequences non-contiguous with the pSAN3 plasmid sequence such as genome-plasmid junctions and/or rearrangements of the plasmid could be identified (Figure SR1-1, B). A list of all non-redundant discontinuous reads around the LB and RB repeats of pSAN3 (Figure SR1-1, B) and all non-redundant internal discontinuous reads (Figure SR1-1, C) was made for further analysis. The discontinuous part (non-matching to pSAN3) of each of these reads was used as a query for BLAST searches into (i) the genome of *M. acuminata* DH-Pahang (version 4 available on Banana Genome Hub, http://banana-genome-hub.southgreen.fr/) to identify any genome-plasmid junctions and (ii) a local pSAN3 BLAST database created in Geneious to identify any plasmid-plasmid junctions/rearrangements. The presence of one LB-genome junction and one genome-RB junction as well as multiple T-DNA junctions (one RB/RB, one RB/LB) suggested the presence of a single and large integration site on chromosome 6 of the banana genome. Multiple attempts to decipher the structure of the insert using conventional PCR and sequencing methods were unsuccessful due to the nature of the large, inverted rearrangement (see 1.3.2 below).

1.3.2 Confirmation and characterisation of the insert sequence in QCAV-4

Long-read PacBio sequencing was then used to further characterise the integration site. More than 4.9 million reads and ~75 Gbp of data was obtained in Continuous Long Reads (CLR) mode with a read length N50 of 17,973 bp. This corresponded to ~42x coverage of the Cavendish genome. The T-DNA sequence of pSAN3 was used to filter out long-reads from the total genomic pool. Approximately 80 long-reads which mapped onto the T-DNA sequence were then assembled using the Flye plugin in Geneious. A single 26,849 bp T-DNA-containing insertion locus (referred to as the "insert") was assembled and together with 4,211 bp and 3,576 bp of 3' and 5' flanking chromosome 6 sequences, respectively, was further refined using the previously generated QCAV-4 Illumina genomic short-read sequences to correct all Flye assembly errors (short indels). Banana genome nucleotide BLAST analysis using the two flanking sequences either side of the insert confirmed its location between position 35,127,849 and 35,127,965 of chromosome 6 and in the intergenic region between two protein kinase domain-containing proteins (Macma4_06_g29410.1 and Macma4_06_g29420.1). Comparative analysis of the insertion locus in QCAV-4 with its parent non-GM line (GN212-12) showed that a 116 bp deletion resulted from the integration of the insert into one of the three chromosome 6 loci of QCAV-4. Deletions of this nature are a common feature of *Agrobacterium*-mediated transformation (Latham *et al*., 2006). Importantly, the insertion into the described chromosome 6 location did not interrupt any known open reading frame as can be seen on Figure SR1-2.

Comprehensive analysis of the insert sequence in event QCAV-4 revealed a structure comprised of three identical copies of the 6,702 bp pSAN3 T-DNA. These are identified by blue arrows labelled T-DNA1, T-DNA2 and T-DNA3 on Figure SR1-2. In addition, a 6,668 bp hybrid fragment of the *MamRGA2* expression cassette had recombined in opposite directions and inserted between T-DNA2 and T-DNA3 (identified by the green box on Figure SR1-2). Finally, several rearrangements have occurred at each of the following locations:

1. Between the 3' genome flanking region and T-DNA1 at position 4,212 to 4,273 (Figure SR1-2, G-LB orange box).

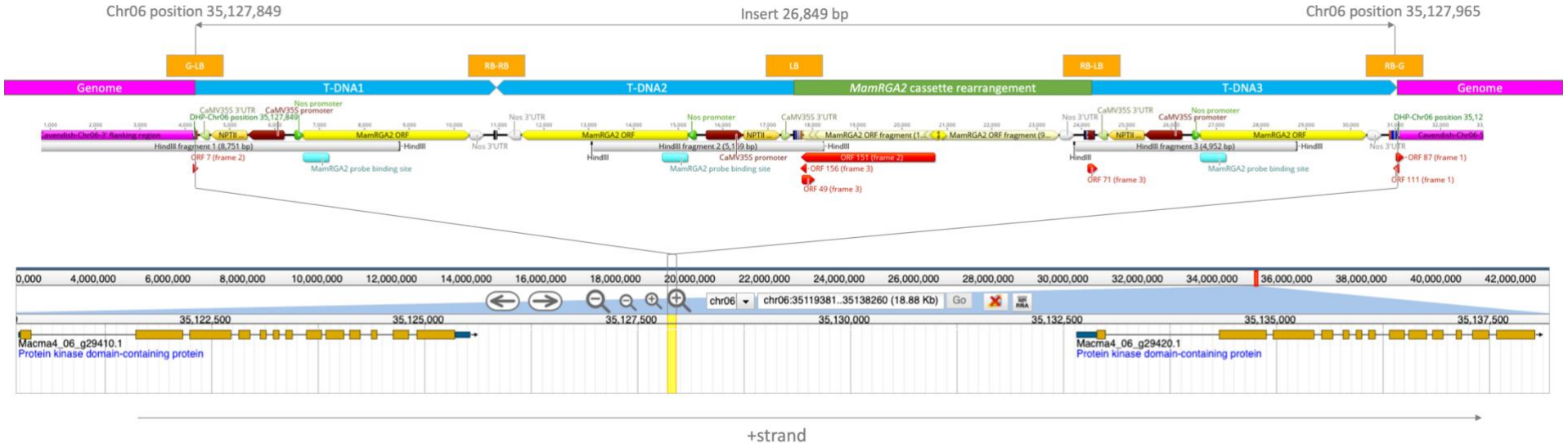


Figure SR1-2. General organisation of the insert in event QCAV-4. A T-DNA insertion of 26,849 bp was assembled from PacBio long-reads and corrected with Illumina short-reads and located on banana chromosome 6. The single insertion is located in the anti-sense direction between positions 35,127,849 and 35,127,965 creating a 116 bp deletion of the original chromosome 6 locus in the intergenic region between two protein kinase domain-containing proteins (Macma4_06_g29410.1 and Macma4_06_g29420.1) and therefore did not interrupt any known open reading frame. The insert contains three full, intact and functional copies of the 6,702 bp T-DNA (T-DNA 1 to 3, blue arrows) as well as two fragmented portions of the *MamRGA2* expression cassette recombined in opposite directions and inserted between T-DNA2 and T-DNA3 (green box). The two genome-T-DNA and three inter T-DNA junctions contain various levels of rearrangement and are indicated with orange boxes. Seven new ORFs larger than 30 amino acids were identified (red arrows) in the inter T-DNA regions only. No evidence of any vector backbone sequence was detected. **Please note:** the insert sequence and diagrams provided are oriented in the antisense of their real orientation on Chromosome 6.

2. Between T-DNA1 and T-DNA2 at position 10,905 to 10,942 (Figure SR1-2, RB-RB orange box).
3. Large *MamRGA2* expression cassette rearrangement at position 17,594 to 24,261 (Figure SR1-2, green box, LB and RB-LB orange boxes) including a 3,042 bp antisense portion of the *MamRGA2* ORF, a 2,672 bp sense portion of the *MamRGA2* ORF and its Nos 3'UTR and a 142 bp sense region of the CaMV35S promoter.
4. Between T-DNA3 and the 5' genome flanking region at position 30,868 to 31,060 (Figure SR1-2, RB-G orange box).

Combinations of the diagram presented on Figure SR1-2 and a detailed annotated sequence of the insert (Geneious and GenBank format) are provided.

1.3.3 Identification of new ORFs in event QCAV-4

To investigate the possibility of novel ORFs resulting from the presence of the insert in QCAV-4, an ORF analysis was conducted in Geneious to identify potential start-to-stop ORFs within the entire insert and including both the 3' 4,211 bp and 5' 3,576 bp sequences spanning the insert. This analysis examined each of three possible reading frames in both orientations for potential ORFs capable of encoding sequences of 30 or more amino acids. Seven new and unintended ORFs resulting from the presence of the insert in QCAV-4 were identified from this analysis (Figure SR1-2, red arrows and Table SR1-1).

1.4 Conclusions

Event QCAV-4 contains a single 26,849 bp insert in an intergenic region of Chromosome 6. It is comprised of three identical copies of the 6,702 bp pSAN3 T-DNA and a 6,668 bp hybrid fragment of the *MamRGA2* expression cassette inserted between T-DNA2 and T-DNA3. The insertion (i) deleted 116 bp from the original locus, (ii) did not interrupt any know ORFs in that location and (iii) created seven unintended ORFs in QCAV-4. The safety assessment of these ORFs including their potential for expression and any similarity to known allergenic and toxic proteins can be found in study report QUT2023-7.

Table SR1-1. Information regarding the seven newly identified ORFs in QCAV-4

ORF ID	Nucleotide location[a]	Frame	Strand	Length (bp)	Length (amino acid)	Molecular weight (kDa)	Deduced amino acid sequence
ORF 111	31,078 → 30,959	1	Sense (+)	120	39	4.5	MHVMLYSWIRRGREDDSGGSIRITHYYGQFKLKAGANSH*
ORF 87	31,004 → 31,174	1	Antisense (-)	171	56	6.5	MCYSDRSSRVVFPAPPNPTIEHHMHSGIIENKNLKFSTEKCFVIVRRLVHKTENVK*
ORF 7	4,188 → 4,286	2	Antisense (-)	99	32	4.1	MDRHLKSRIRFWFKQQWPRQLNNTLRCKQIDA*
ORF 151	20,724 → 17,749	2	Sense (+)	2976	991	112.8	MWVCVSDDFDVKRITREITEYATNGRFMDLTNLNMLQVNLKEEIRGTTFLLVLDDVWNEDPVKWESL LAPLDAGGRGSVVIVTTQSKKVADVTGTMEPYVLEELTEDDSWSLIESHSFREASCSSTNPRMEEIG RKIAKKISGLPYGATAMGRYLRSKHGESSWREVLETETWEMPPAASDVLSALRRSYDNLPPQLKLCF AFCALFTKGYRFRKDTLIHMWIAQNLIQSTESKRSEDMAEECFDDLVCRFFFRYSWGNYVMNDSVHD LARWVSLDEYFRADEDSPLHISKPIRHLSWCSERITNVLEDNNTGGDAVNPLSSLRTLLFLGQSEFR SYHLLDRMFRMLSRIRVLDFSNCVIRNLPSSVGNLKHLRYLGLSNTRIQRLPESVTRLCLLQTLLLE GCELCRLPRSMSRLVKLRQLKANPDVIADIAKVGRLIELQELKAYNVDKKKGHGIAELSAMNQLHGD LSIRNLQNVEKTRESRKARLDEKQKLKLLDLRWADGRGAGECDRDRKVLKGLRPHPNLRELSIKYYG GTSSPSWMTDQYLPNMETIRLRSCARLTELPCLGQLHILRHLHIDGMSQVRQINLQFYGTGEVSGFP LLELLNIRRMPSLEEWSEPRRNCCYFPRLHKLLIEDCPRLRNLPSLPPTLEELRISRTGLVDLPGFH GNGDVTTNVSLSSLHVSECRELRSLSEGLLQHNLVALKTAAFTDCDSLEFLPAEGFRTAISLESLIM TNCPLPCSFLLPSSLEHLKLQPCLYPNNNEDSLSTCFENLTSLSFLDIKDCPNLSSFPPGPLCQLSA LQHLSLVNCQRLQSIGFQALTSLESLTIQNCPRLTMSHSLVEVNNSSDTGLAFNITRWMRRRTGDDG LMLRHRAQNDSFFGGLLQHLTFLQFLKICQCPQLVTFTGEEEEKWRNLTSLQILHIVDCPNLEVLPA NLQSLCSLSTLYIVRCPRIHAFPPGGVSMSLAHLVIHECPQLCQHVPGTFGHP*
ORF 156	17,852 → 17,733	3	Sense (+)	120	39	4.4	MRFLPEVSACPWHIWSSMNALSCVSMSLAHLVIHECPQP*
ORF 71	24,139 → 24,339	3	Antisense (-)	201	66	7.9	MPLPTVVPKMDPHPRGASWKKKTFQPRLQSKWIDVNMLEQLWRIYCGVNKLTLRQLNNTLRTFLMY*
ORF 49	17,755 → 18,026	3	Antisense (-)	282	93	10.6	MTKCARDMLTQLRAFMDDQMCQGHADTSGRKRMDSWASDDVQGAEGAEALQVCRQYLQVWTINDVQN LKRSKVSPLLFLFAGEGYELWTLADL*

[a]Location on the supplied insert sequence including 5' and 3' flanking genomic sequence. **Please note:** the insert sequence and diagrams provided are oriented in the antisense of their real orientation on Chromosome 6.

1.5 Supporting documents

- No additional documents provided with study report QUT2023-1, however, raw data and mapping sequence files can be provided at FSANZ's request and can be visualised in Geneious.

QUT2023-2: Stability of the DNA insert in banana event QCAV-4 across multiple generations

STUDY IDENTIFICATION

QUT2023-2

AUTHORS

[redacted]
s 22
[redacted]

PERFORMING LABORATORY

[redacted]

[redacted] 2023

2.1 Introduction

Banana event QCAV-4 (QUT-QCAV4-6) was created by *Agrobacterium tumefaciens*-mediated transformation of banana (*Musa acuminata* subgroup Cavendish cv Grand Nain) embryogenic cells with plasmid pSAN3. This plasmid contains the *neomycin phosphotransferase* II (*npt*II) selectable marker gene from the Tn5 transposon of *Escherichia coli* strain K12 and the *MamRGA2* disease resistance (R) gene from the wild banana *Musa acuminata* ssp. *malaccensis* which confers resistance to the devastating banana fungal pathogen *Fusarium oxysporum* f. sp. *cubense* tropical race 4 (TR4) (Dale *et al*., 2017).

Southern blot analysis of banana event QCAV-4 initially revealed the presence of multiple T-DNA integrations in this event (Dale *et al*., 2017) and the insert was fully characterised in study report QUT2023-1. The objective of this study was to demonstrate stability of the insert in successive generations of event QCAV-4.

2.2 Materials and Methods

2.2.1 DNA extraction

Fresh (0.1 g) ground leaf tissue was resuspended in 0.7 mL of extraction buffer (100 mM Tris-HCl, pH 8.0; 50 mM EDTA; 1.4 M NaCl; 80 mM Na_2SO_3; 2% PVP-10; and 2% cetyltrimethylammonium bromide). Samples were incubated at 65°C for 30 min before mixing with an equal volume of chloroform:isoamylalcohol (24:1) and the mixture centrifuged at 18,000 x*g* for 5 min. The supernatant was mixed with an equal volume of isopropanol and incubated at room temperature for 5 min. Nucleic acids were pelleted by centrifugation as described above and the pellets washed with 70% ethanol, air dried, and resuspended in 100 µL of sterile distilled water.

2.2.2 Restriction endonuclease digestion

Samples of genomic DNA (5 µg) were digested overnight with 20 U of restriction enzyme *Hind*III (Cat. # R3104L, New England Biolabs) at 37°C. A further 20 U of *Hind*III was added and the digestion was continued for an additional hour. The digested DNA was precipitated by adding (per 100 µL volume), 10 µL 125 mM EDTA (pH 8.0), 10 µL 3M sodium acetate and 300 µL ethanol followed by incubation at room temperature (30 min) and centrifugation (21,000 x*g*, 30 min). The DNA pellet was washed with 70% ethanol, air-dried, and dissolved in 30 µL of sterile distilled water.

2.2.3 Electrophoretic separation and Southern blotting

Digested genomic DNA plus gel loading dye (Cat. #B7024S, New England Biolabs) was loaded onto a 0.8% agarose gel in 1x Tris-acetate-EDTA (TAE) buffer and electrophoresed at 40 V until the dye front reached the bottom of the gel. The DNA was depurinated by incubating the gel in 0.25 M HCl, followed by rinsing in water and DNA denaturation by incubating the gel in 0.5 M NaOH and 1.5 M NaCl for 30 min and rinsing again in water. The gel was then neutralised by submerging in neutralisation solution (0.5 M Tris-HCL, 1.5 M NaCl and 0.5 M EDTA, pH 7.5) and incubated with gentle agitation for 30 min, and the DNA transferred from the gel to a nylon membrane (Cat. #11417240001, Roche), rinsed twice in saline sodium citrate (SSC) for 2 min, and UV crosslinked. The nylon membrane was transferred to a hybridisation bottle with 20 mL of DIG Easy Hyb solution (Cat. #11603558001, Roche) and pre-hybridised at 42°C for 1 h.

2.2.4 DNA probe labelling and Southern blot hybridisation

The *MamRGA2*-specific probe was amplified by PCR using Taq DNA polymerase (Cat. #M712C, Promega) in reaction mixes containing the appropriate primers (Table SR2-1), 2 ng of plasmid template and DIG PCR

labelling mix (Cat. #11277065910, Roche). For hybridisation, 20 mL of freshly pre-heated (65°C) hybridisation buffer containing denatured DIG-labelled probe was added to the drained hybridisation bottle containing the membrane blot and hybridised overnight at 42°C. The membrane was washed and placed in blocking solution (3% skim milk powder in maleic acid buffer, 0.1 M maleic acid, 0.15 M NaCl, pH 7.5) for 1 h before incubation with anti-digoxigenin-AP (Cat. #11093274910, Roche), 1:20,000 in blocking solution for 30 min. Detection was achieved using CDP-Star (Cat. #11685627001, Roche), as per the manufacturer's instructions. Membranes were visualised using chemiluminescence on a ChemiDoc MP Imaging system (BioRad).

Table SR2-1. Primers used to amplify the *MamRGA2* probe

Primer name	Sequence (5' → 3')
RGA2-probe-F2	ATGGCTGGTGTCACATCACAGG
RGA2-probe-R2	GAACAACAGGAACCGCCCCATC

2.3 Results

Domesticated bananas are vegetatively propagated and, as such, the progeny from a parent plant is essentially genetically identical to each other and the parent plant. Bananas are also a perennial crop. When a banana is planted, the plant crop grows from meristematic tissue at the basal corm. After the development and maturity of a bunch on the plant crop, the plant crop pseudostem dies and another new pseudostem, known as the first ratoon, grows from a different meristem on the basal corm. This process can be repeated through numerous ratoons, and commercially through more than 10 ratoons before replanting (OGTR, 2023).

The stability of the insert in event QCAV-4 was investigated in several different "generations" of GM banana plants using Southern blot analysis. The samples used for this analysis are shown in Table SR2-2 and Figure SR2-1. The first-generation samples are derived from plants maintained in tissue culture and designated GN212-12 and 121-12. GN212-12 is a non-GM control plant derived from the banana cell line (GN212-12) used to generate GM banana event QCAV-4, while sample 121-12 is the original mother plant of event QCAV-4. Samples 20236, 20246 and 20265 represent three QCAV-4 plants (clones) in tissue culture that were initiated from meristems of plants growing in the field under OGTR licence DIR107 (2012-2015). Samples 20236, 20265 and 20267 represent three QCAV-4 banana plants (clones) currently growing in a field trial in the Northern Territory under OGTR licence DIR146. Samples were taken from each of these three plants growing in three successive "generations": the plant crop, first ratoon and second ratoon. Leaf tissue was collected from the banana plants, genomic DNA extracted and digested with the restriction enzyme *Hin*dIII. Evidence of genetic stability was indicated based on consistent banding patterns between the original QCAV-4 plants maintained in tissue culture and the three successive generations of QCAV-4 plants in the field.

A simplified organisation of the insert in QCAV-4 is shown in Figure SR2-2A with the expected *Hin*dIII restriction digestion products indicated along with the fragment sizes and *MamRGA2* probe hybridisation specificity. Although wild-type Grand Nain bananas contain three endogenous *RGA2* homologs (*MaRGA2*) to which the *MamRGA2* probe hybridises, these can be differentiated from *MamRGA2* due to their unique predicted *Hin*dIII restriction fragment sizes of 3,453 bp, 3,337 bp and 3,313 bp.

As shown in Figure SR2-2B, the *MamRGA2* probe hybridises to the three predicted *Hin*dIII restriction fragments of 8,751 bp, 5,169 bp and 4,952 bp in all QCAV-4 samples (lanes 2-14), as well as to undigested pSAN3 (lane 15). Further, the presence of the endogenous homologs of *RGA2* in Grand Nain bananas is

shown by hybridisation of the probe to the expected fragment sizes of 3,453 bp, 3,337 bp and 3,313 bp (the shorter two fragments co-migrate as a more intensely staining single band) in all banana samples.

The observation of consistent hybridisation patterns across each of the different banana "generations" confirmed stable integration of the inserted DNA in event QCAV-4.

Table SR2-2. Sample information for generational stability analysis

Sample ID	Sample Details	Origin of sample
GN212-12	Non-GM Grand Nain control	Tissue culture
121-12	QCAV-4 original mother plant	
20236	QCAV-4 (Ex-field trial 1 plants)	
20246		
20265		
20236	QCAV-4 - Plant crop	Field trial 2
20265		
20267		
20236	QCAV-4 - 1st ratoon	
20265		
20267		
20236	QCAV-4 - 2nd ratoon	
20265		
20267		

2.4 Conclusions

Southern blot analysis demonstrated that the insert was stably inherited over five generations in event QCAV-4.

2.5 Supporting documents

- No additional documents are provided with study report QUT2023-2.

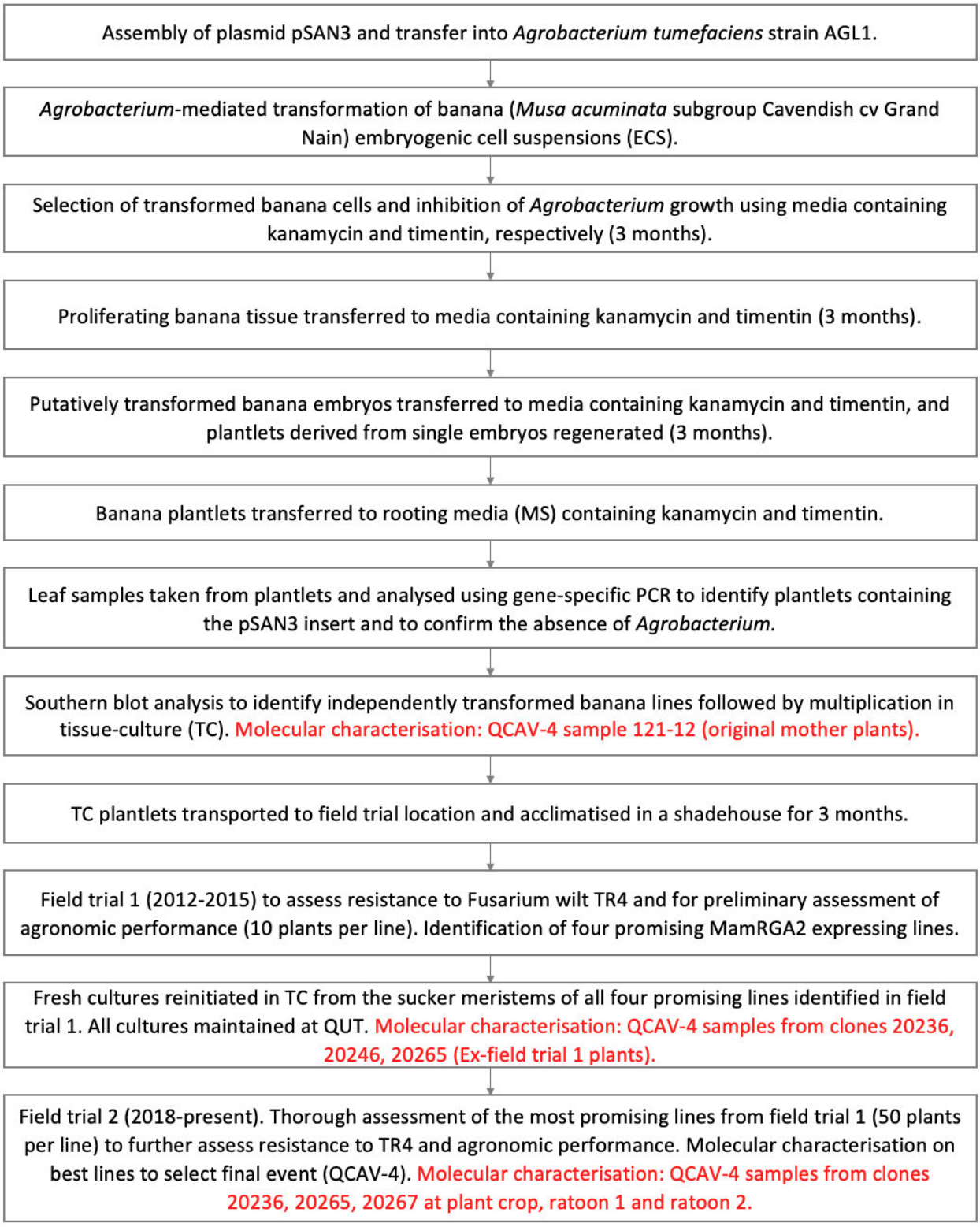


Figure SR2-1. Development process of banana event QCAV-4 and sample identification.

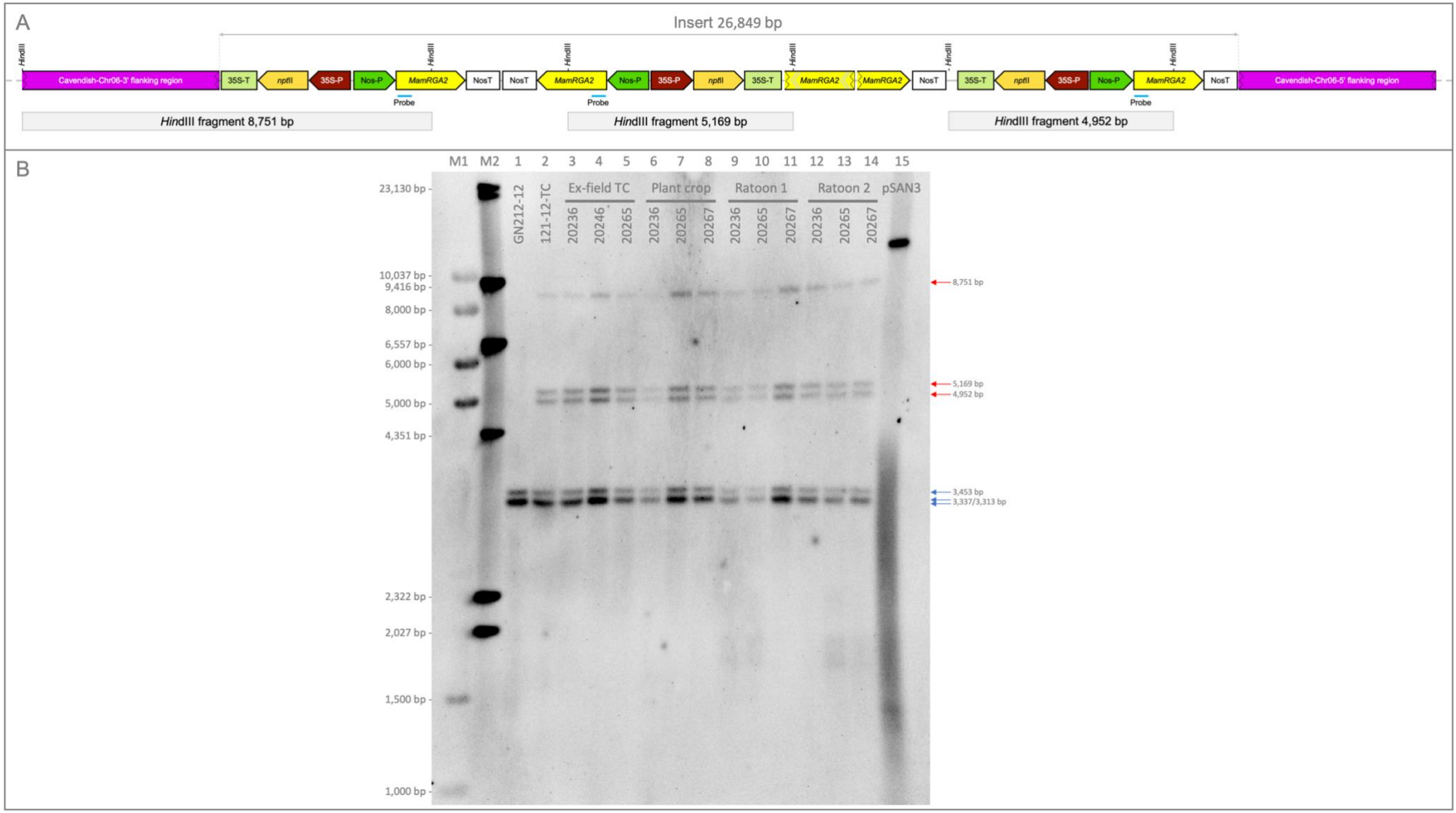


Figure SR2-2. Stability of the T-DNA insert in event QCAV-4 demonstrated by Southern blot analysis. A. Simplified representation of the organisation of the insert in QCAV-4. The relevant positions of *Hin*dIII restriction sites as well as predicted *Hin*dIII restriction fragments are shown (grey boxes) along with the hybridisation position of the chosen *MamRGA2* DIG-labelled DNA probe. B. Southern blot analysis of wild-type (GN212-12) and different generations of QCAV-4 (121-12, 20236, 20246, 20265 and 20267) plants. Genomic DNA from each sample was digested with *Hin*dIII and hybridised with the *MamRGA2* probe. Red arrows indicate *Hin*dIII DNA fragments originating from the insert while blue arrows indicate endogenous *MaRGA2* fragments. M1 = Bioline HyperLadder 1kb; M2 = Sigma-Aldrich DNA Molecular Weight Marker II, DIG-labeled; pSAN3 = linearised plasmid pSAN3.

QUT2023-3: Field performance of banana event QCAV-4

STUDY IDENTIFICATION

QUT2023-3

AUTHORS

[redacted]
[redacted]
s 22
[redacted]
[redacted]
[redacted]

PERFORMING LABORATORY

[redacted]
[redacted]
[redacted]

-

[redacted]
[redacted]

[redacted]

[redacted] 2023

3.1 Introduction

Banana event QCAV-4 (QUT-QCAV4-6) was created by *Agrobacterium tumefaciens*-mediated transformation of banana (*Musa acuminata* subgroup Cavendish cv Grand Nain) embryogenic cells with plasmid pSAN3. This plasmid contains the *neomycin phosphotransferase* II (*npt*II) selectable marker gene from the Tn5 transposon of *Escherichia coli* strain K12 and the *MamRGA2* disease resistance (R) gene from the wild banana *Musa acuminata* ssp. *malaccensis* which confers resistance to the devastating banana fungal pathogen *Fusarium oxysporum* f. sp. *cubense* tropical race 4 (TR4) (Dale *et al*., 2017).

The genetic modification resulting in banana event QCAV-4 was intended to confer resistance to TR4 and was not expected to affect agronomic performance or phenotypic characteristics. To confirm that QCAV-4 was otherwise agronomically equivalent to the parental Grand Nain control plants GN212-12, disease incidence, yield and other agronomic characteristics were assessed from plants growing in two confined field trials s 22

3.2 Materials and Methods

3.2.1 Experimental design and maintenance

The field phase of this study was carried out in two confined field trials at a single location in the Northern Territory with high TR4 disease pressure. The first field trial was conducted between 2012-2015, with the experimental design, methodology and outcomes from the trial published in Dale *et al*. (2017). The second field trial (2018-present; Figure SR3-1) tested a total of 300 plants which included event QCAV-4 as well as three other events generated with the pSAN3 construct namely, events RGA2-2, RGA2-3 and RGA2-5. For each event, 50 plants (replicates) were tested as well as 50 plants each of two types of non-transgenic controls, one derived from the original cell line from which the events were generated (GN212-12) and a conventional tissue culture control of the Cavendish cultivar Williams. The field trial design was based on plots of 10 identical clones of each event randomised across four blocks. Because each block could only accommodate 75 plants, two plots of events RGA2-2 and RGA2-5 were split into two smaller plots of 5 plants at the eastern end of the trial site. The field trial layout, which was generated in MusaBase (https://musabasequt.sgn.cornell.edu/), is provided in Figure SR3-2.

3.2.2 General maintenance and data recording

Dual-row raised beds, with an inter-row spacing of 5.5 m, were prepared incorporating lime and/or dolomite to achieve neutral pH. Prior to planting, a single superphosphate (SSP), potassium sulphate and NPK mix was added as a granular preparation and was rotary hoed into the beds. Plants were established at a spacing of 2.3 m. All plots were clearly identified using stake markers with individual identifiers for every plant.

Following establishment, the fertiliser regime included bi-monthly application of NPK (12%/5.2%/14% in granular form) and potassium sulphate (granular), alternating fortnightly applications of either potassium nitrate or calcium nitrate (fertigation) as well as monthly zinc, boron, manganese and copper fertigation. The levels of nutrient supplementation were adjusted in conjunction with leaf and soil analysis. Additional non-routine fertiliser applications included foliar sprays as required.

Weed control was achieved by regular spraying with Paraquat/Diquat (2.4 L/ha) or Glufosinate-ammonium (1 L/ha) until a closed canopy was achieved, then only when needed to reduce weeds on beds. Weed control between the rows was maintained by regular slashing of the inter-row spaces. For control of foliar diseases and insect pests, regular applications of fungicides/pesticides (Mancozeb (4 kg/ha); Chlorothalonil (2 L/ha)) was also carried out either by misting or helicopter.

Control of excess suckers was done manually by regular de-suckering with a knife. All suckers from the initial (“plant”) crop were killed until the plants were 4-5 months old. At that point, the most appropriate

Figure SR3-1. Aerial views of the DIR146 field trial site [redacted] **s 22** [redacted] The boxed area on the left panel shows the area planted corresponding to the field trial layout presented in Figure SR3-2. The boxed area on the right panel shows the entire field trial area approved under OGTR licence DIR146.

Shadehouse End			
Planting 1 - 21-03-2018			
BLOCK 1	BLOCK 2	BLOCK 3	BLOCK 4
20141	20243	20131	Williams 10
20142	20242	20132	Williams 9
20143	20241	20133	Williams 8
20144	20237	20134	Williams 7
20148	20236	20135	Williams 6
20149	20235	20136	Williams 5
20150	20234	20137	Williams 4
20151	20233	20138	Williams 3
20152	20232	20139	Williams 2
20153	20231	20140	Williams 1
20267	20077	20314	20313
20268	20076	20315	20312
20269	20075	20316	20311
20270	20074	20317	20310
20271	20073	20321	20309
20272	20072	20322	20308
20273	20071	20323	20307
20278	20070	20324	20306
20275	20069	20325	20305
20276	20068	20326	20304
20048	20216	20154	20253
20049	20215	20155	20252
20050	20214	20156	20251
20051	20213	20167	20250
20052	20212	20158	20249
20053	20211	20159	20248
20098	20210	20160	20247
20055	20209	20161	20246
20056	20208	20162	20245
20057	20207	20163	20244
20291	20206	20168	20117
20292	20205	20169	20116
20293	20219	20170	20145
20294	20203	20171	20114
20295	20202	20172	20113
20296	20201	20173	20112
20297	20345	20174	20111
20301	20344	20175	20110
20302	20343	20176	20109
20303	20341	20177	20108
20254	20067	20327	20200
20255	20066	20328	20185
20256	20065	20329	20198
20257	20064	20330	20197
20261	20063	20331	20346
20262	20062	20332	20195
20263	20061	20333	20194
20264	20060	20334	20193
20265	20059	20335	20192
20266	20058	20336	20191
20178	20097	20078	20230
20179	20096	20079	20229
20180	20095	20080	20228
20181	20094	20081	20227
20182	20093	20082	20226
20183	20092	20083	20225
20184	20091	20084	20224
20188	20090	20085	20223
20189	20089	20086	20222
20190	20088	20087	20221
Williams 21	Williams 50	Williams 11	Williams 40
Williams 22	Williams 49	Williams 12	Williams 39
Williams 23	Williams 48	Williams 13	Williams 38
Williams 24	Williams 47	Williams 14	Williams 37
Williams 25	Williams 46	Williams 15	Williams 36
Williams 26	Williams 45	Williams 16	Williams 35
Williams 27	Williams 44	Williams 17	Williams 34
Williams 28	Williams 43	Williams 18	Williams 33
Williams 29	Williams 42	Williams 19	Williams 32
Williams 30	Williams 41	Williams 20	Williams 31
20281	20290	20118	20130
20282	20289	20119	20129
20283	20288	20120	20128
20284	20287	20121	20124
20285	20286	20122	20123

Plot color
Non-GM control cv Grand Nain (GN212-12)
RGA2-2
RGA2-3
RGA2-4
RGA2-5
Non-GM control cv. Williams

Figure SR3-2. Field trial 2 (DIR146) layout.

sucker for the 1st ratoon plant was selected. De-suckering was then done at intervals of 5-6 weeks. Manual de-leafing for control of foliar diseases was also done on a regular basis. Plants were inspected weekly for bunch emergence and covering of bunches with reflective plastic was done to protect fruit and exclude frugivores. Bunch harvesting was also carried out on a weekly basis at maturity. Plants were irrigated daily via a micro-irrigation system unless sufficient rainfall had occurred. The irrigation schedule was adjusted monthly depending on the forecast weather conditions.

3.2.3 Agronomic parameters

Bunch weight and cycle time were examined over five generations (plant crop and four ratoons) in trial 2 for all events in the trial. Bunch weight was determined by removing the entire bunch at the base of the peduncle and weighing using a scale attached to an off-road vehicle. Cycle time was defined as the period from planting to harvest in the case of the plant crop, and the time between harvests for all subsequent (ratoon) generations and expressed in days. Yield calculations were made based on a conservative industry standard of 1670 plants/ha.

3.2.4 Statistical analysis

Statistical analysis (independent samples t-Test) was done on SPSS® Statistics Version 27 (IBM®) and significant differences with the non-GM control reported at $p<0.05$.

3.3 Results

3.3.1 Agronomic parameters

Two of the most important agronomic traits for banana production are bunch weight and cycle time. During the second field trial of QCAV-4 plants under licence DIR146, these two traits were examined over five generations (plant crop and four ratoons) for all events including QCAV-4 and the non-GM control line GN212-12 from which event QCAV-4 is derived. The raw data collected is presented in Figures SR3-4 and SR3-5 in section 3.5 of this document.

The average bunch weight in event QCAV-4 fitted well within the range of the non-GM control plants across all the five plant generations (Table SR3-1). However, when directly compared to the non-GM control plants, QCAV-4 had a significant ($p<0.001$) reduction in average bunch weight during the first two plant generations (plant crop and ratoon 1) (Table SR3-1). A slight reduction in productivity has commonly been observed in GM events following their initial establishment in the field and is therefore very unlikely to be the result of the expression of the *MamRGA2* transgene. Indeed, banana plants generated through biotechnology such as QCAV-4, unlike their non-GM counterparts, experience a physiologically and metabolically challenging tissue culture process which starts with the transformation of embryogenic cells using a plant pathogen (*Agrobacterium tumefaciens*). These cells then undergo a lengthy process of selection using antibiotics, regeneration, multiplication and rooting before being acclimatised to the natural environment in a shadehouse and subsequently planted into the field. Our data indicates that it took QCAV-4 three generations in the field to achieve comparable bunch weights to the non-GM control plants. Indeed, by ratoon 2 (generation 3), the average bunch weight obtained from QCAV-4 plants was marginally higher ($p= 0.743$) than that of the non-GM control plants while in ratoons 3 and 4 (generations 4 and 5), there were no significant differences in the average bunch weights obtained from QCAV-4 and non-GM control plants.

The ability of QCAV-4 to sustain a competitive performance under high TR4 incidence gives it a tremendous advantage over its non-GM counterpart. Indeed, a better measure of plant performance is to calculate its yielding capacity in a specified area. Yield comparisons between QCAV-4 and the non-GM control from field trial 2 are presented in Table SR3-2. Except for the plant crop, QCAV-4 outperformed its non-GM counterpart in all generations, with yield exceeding the non-GM controls by more than 50% in the 4th ratoon. This data clearly demonstrates that QCAV-4 can sustain the extreme disease pressure imposed by TR4 while maintaining an industry competitive yield performance. Although many factors determine yield from a banana plantation (including environmental conditions, agronomic practices, cultivar and ratooning

management), the average yield in Australia (of which the Cavendish banana accounts for around 95%) ranges from 20 to 30 t/ha (OGTR, 2023; Holligan *et al*., 2017)

From a banana farmers' perspective, a short cycle time is a preferred agronomical trait as it allows more harvests per unit of time and a faster return on investment. Cycle time is defined as the period from planting to harvest in the case of the plant crop, and the time between harvests for all subsequent (ratoon) generations. When comparing the cycle time between non-GM control and QCAV-4 plants, no distinct trend was observed (Table SR3-1). QCAV-4 had a slightly longer cycle time at plant crop (p= 0.106) and ratoon 2 (p= 0.959) although that difference was not significant. However, its cycle time was significantly shorter than the non-GM controls by approximately 6 and 15% in ratoon 1 (p<0.05) and ratoon 3 (p<0.01), respectively, while the ~9% reduction in cycle time in ratoon 4 was not significant (p= 0.119) (Table SR3-1).

In addition to yield and cycle time, agronomic characteristics such as plant height and girth were also recorded from plants bearing fruit in ratoons 6 and 7 (Table SR3-3). The raw data collected is also presented in Figure SR3-6 in section 3.5. Measurements between plants were standardised to the same developmental stage by only assessing plants with bunches (which reduced the number of plants available for assessment). Further, the high incidence of TR4 in the non-GM control plants limited data collection to 10 non-GM plants compared to 30 QCAV-4 plants. Statistical analysis revealed no significant differences between non-GM and QCAV-4 plants. In addition to plant height and girth, as a component of the US Plant Patent application for QCAV-4 (Patent No: PP34,398 P3), a broad array of botanical descriptors was recorded from the 5th generation (ratoon 4) of plants which showed that, in the absence of disease pressure, QCAV-4 appeared to be phenotypically identical to its non-GM counterpart (Tables SR3-7 and SR3-8 in section 3.5).

Based upon the results from two confined field trials in Australia over a period of eight years, expression of the *MamRGA2* transgene in event QCAV-4 did not result in any unintended phenotypic or morphological effects that would preclude its cultivation.

3.3.2 Field performance of event QCAV-4

Event QCAV-4 (formerly referred to as a line of RGC2 in OGTR licence DIR107, a line of RGA2 in OGTR licence DIR146 and as RGA2-4 in Dale *et al*., (2017)) has been evaluated for resistance to TR4 in two OGTR-approved field trials conducted in Australia. The first trial was conducted from 2012-2015 (DIR107 issued in 2011) while the second trial commenced in 2018 and is ongoing (DIR146 issued in 2016). Both trials were conducted on a commercial banana farm in the Northern Territory with high TR4 disease pressure.

The first field trial included 10 replicates each of five independently transformed MamRGA2-expressing lines (QCAV-4 and four additional lines) as well as several non-GM control cultivars (Dale *et al*., 2017). Plants were assessed for TR4 infection during the 3-year duration of the trial by the presence of typical disease symptoms and by a combination of fungal isolation and PCR-based assays. By the end of the trial, 87.5% of the non-GM Grand Nain control plants were either dead or infected by TR4. In contrast, the infection rate in the five transgenic *MamRGA2* lines after 3 years ranged from 0-67% with QCAV-4 showing a 20% infection rate.

To investigate the basis of the resistance, quantitative RT-PCR (qRT-PCR) was used to assess the levels of transgene expression. Using primers that specifically amplified *MamRGA2* RNA, a strong correlation was observed between *MamRGA2* RNA expression and the degree of TR4 protection. Using primers that would amplify both *MamRGA2* RNA and endogenous *MaRGA2* RNA, the most resistant line was again the highest expressor of *RGA2* RNA (*MamRGA2* + endogenous *MaRGA2*) whereas the other three lines that were TR4 resistant also showed moderate to high levels of *RGA2* RNA expression (*MamRGA2* + endogenous *MaRGA2*). In contrast, the most susceptible transgenic *MamRGA2* line had the lowest expression levels of *RGA2* (*MamRGA2* + endogenous *MaRGA2*). The results of this field trial have been published in *Nature Communications* (Dale *et al*., 2017).

Table SR3-1. Bunch weight and cycle time in non-GM control and QCAV-4 plants

Harvest	Plant crop		Ratoon 1		Ratoon 2		Ratoon 3		Ratoon 4	
Bunch weight (kg)										
Genotype	**Average ± SD**	**Range**	**Average ± SD**	**Range**	**Average ± SD**	**Range**	**Average ± SD**	**Range**	**Average ± SD**	**Range**
non-GM	33.1 ± 4.7	18.5 - 42.0	29.9 ± 5.7	9.6 - 40.3	31.5 ± 8.8	10.0 - 49.8	29.8 ± 8.0	17.5 - 45.5	35.7 ± 5.4	27.5 - 45.5
n	46		37		32		25		15	
QCAV-4	28.1 ± 4.3***	**15.8** - 39.4	24.3 ± 5.1***	13.5 - 36.2	31.7 ± 6.0	15.0 - 42.3	28.6 ± 4.6	**14.0** - 40.5	34.8 ± 3.8	**25.5** - 42.0
n	50		49		42		45		34	
Cycle time (days)										
Genotype	**Average ± SD**	**Range**	**Average ± SD**	**Range**	**Average ± SD**	**Range**	**Average ± SD**	**Range**	**Average ± SD**	**Range**
non-GM	327.3 ± 11.0	317 - 362	212.4 ± 22.5	169 - 270	206.1 ± 13.1	183 - 224	206.2 ± 38.2	156 - 273	213.5 ± 33.9	147 - 257
n	46		36		28		18		13	
QCAV-4	331.1 ± 13.6	317 - **372**	199.2 ± 27.0*	**134** - **277**	207.3 ± 15.2	**179** - **249**	174.9 ± 27.3**	**141** - **275**	193.6 ± 39.5	**104** - **306**
n	50		49		42		40		31	

Independent samples t-Test, significant differences with the non-GM control asserted at 95%*, 99%**and 99.9%***.
Values in bold extend outside the range of the non-GM control dataset.
n = biological replicates. SD = standard deviation.

Table SR3-2. Yield comparisons between non-GM and QCAV-4 plants

Yield (t/ha)					
Genotype	**Plant crop**	**Ratoon 1**	**Ratoon 2**	**Ratoon 3**	**Ratoon 4**
non-GM	51	37	34	25	18
QCAV-4	47	40	44	43	40

Yield calculated based on 1670 plants per hectare.

Table SR3-3. Plant girth and height in non-GM and QCAV-4 plants

Genotype	n	Plant girth (mm)		Plant height (mm)	
		Average ± SD	**Range**	**Average ± SD**	**Range**
non-GM	10	774 ± 94	650 - 880	2,546 ± 173	2,350 - 2,810
QCAV-4	30	757 ± 53	660 - 870	2,615 ± 127	**2,280** - **2,840**

Data collected from a mixture of plants harbouring bunches from either ratoon 6 or 7.
Values in bold extend outside the range of the non-GM control dataset.
n = biological replicates. SD = standard deviation.

In the second (ongoing) field trial, 50 replicates of QCAV-4 and the three additional promising *MamRGA2* lines identified in field trial 1, in addition to 50 non-GM control banana plants, are currently being assessed for TR4 resistance. The presence of TR4 in plants showing typical disease symptoms is confirmed using PCR-based assays. The disease incidence in the QCAV-4 plants after five generations (plant crop and four ratoons) was 2% whereas 66% of non-GM control plants were infected (Figure SR3-3).

These results provide strong evidence, based on the differing levels of expression of the transgene across five different transgenic lines and controls, that MamRGA2 is providing resistance to TR4 in event QCAV-4 and that the phenotype trait was stable and inherited across multiple generations.

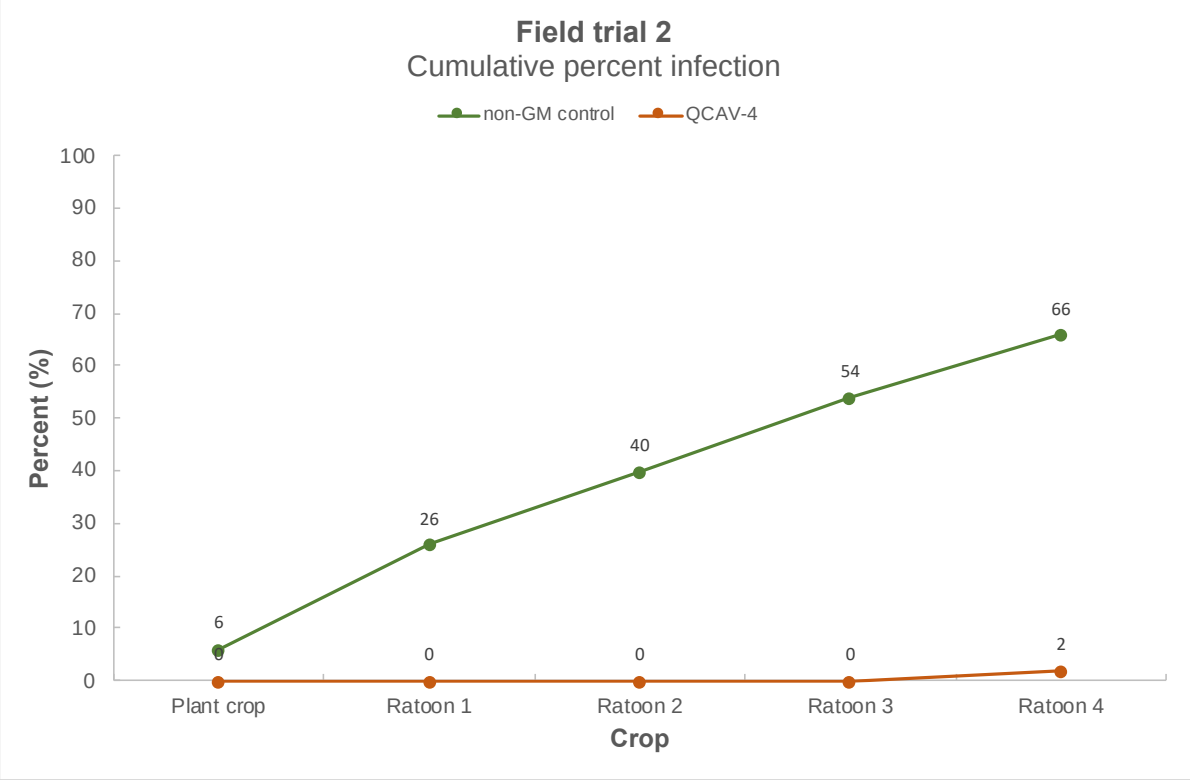


Figure SR3-3. Progression of TR4 infection in non-GM control plants and QCAV-4 over 5 crop cycles.

3.4 Conclusions

The resistance of QCAV-4 to TR4 was evaluated in two OGTR-approved field trials from 2012-2015 (DIR107) and from 2018-present (DIR146). Both trials were conducted in the Northern Territory on a commercial banana farm with high TR4 disease pressure. In both trials, the disease incidence in QCAV-4 plants was significantly lower than the non-genetically modified (GM) Grand Nain control plants. Further, except for TR4 resistance, both QCAV-4 plants and fruit were agronomically and phenotypically indistinguishable from the non-GM Grand Nain control plants and fruit.

3.5 Supporting documents

- Tables SR3-4 to SR3-8 below and associated Excel files supporting documents SD1 to SD3.

Table SR3-4. Bunch weight and cycle time from non-GM GN212-12 plants

Genotype	Plant ID	Bunch weight (kg)					Cycle 1 (days)	Cycle 2 (days)	Cycle 3 (days)	Cycle 4 (days)	Cycle 5 (days)
		Plant crop	Ratoon 1	Ratoon 2	Ratoon 3	Ratoon 4	Plant crop	Ratoon 1	Ratoon 2	Ratoon 3	Ratoon 4
GN212-12	20048	32.8	29 9	38 0			337	224	183		
GN212-12	20049	38.8			31 0		344				
GN212-12	20050	30.9					344				
GN212-12	20051	33.2	22 0	49 8			337	169	222		
GN212-12	20052	34.2	38 8				337	238			
GN212-12	20053	31.1					337				
GN212-12	20055			20 0							
GN212-12	20056	30.2					330				
GN212-12	20057	39.0	20 5	18 0	19 0	39 5	337	193	205	198	252
GN212-12	20098										
GN212-12	20068				23 5	31 0					175
GN212-12	20069	32.9	25.7	42 3	40 5	45 5	317	189	213	157	204
GN212-12	20070	32.2	29 9	34 8			317	213	214		
GN212-12	20071	33.3					323				
GN212-12	20072	31.6	26 2				330	190			
GN212-12	20073	30.5					323				
GN212-12	20074	31.2	32.6		17 5		323	207			
GN212-12	20075	25.4	25.7	25 5	22.7		323	183	222	156	
GN212-12	20076	26.4	28.4		25 0		323	238			
GN212-12	20077	29.0	28 8	29 3	33 0	38 5	337	224	183	175	199
GN212-12	20058	32.9	28.6				317	196			
GN212-12	20059	40.3	36 8	41 0			317	196	206		
GN212-12	20060	35.8					317				
GN212-12	20061	27.0			21 5		317				
GN212-12	20062	36.8	32 5	30 0		30 5	317	213	198		
GN212-12	20063	32.6	29.4	30 0	18 0		323	197	199	172	
GN212-12	20064	33.1		10 0			317				
GN212-12	20065	40.1	34 0	22 5			317	189	203		
GN212-12	20066		24 0	24 0							
GN212-12	20067	27.4	28.6	29 0			317	196	215		
GN212-12	20088	27.6	22.7	26 0	32 5		337	224	204	266	
GN212-12	20089	34.3	33 0	39 0			317	213	220		
GN212-12	20090	33.7	31 9	36 0	35 5	42 0	323	197	224	175	199
GN212-12	20091	28.2	30 5	30 5	25 0	31 3	317	189	222	198	257
GN212-12	20092	18.5	9.6		22 0	37 0	337	270			229
GN212-12	20093	35.0	31.1	39 0	29 0		317	196	215	205	
GN212-12	20094	37.4	37.6	48 5	39 8		323	207	214	175	
GN212-12	20095	42.0		38 5			317				
GN212-12	20096	34.3	34 0				323	207			
GN212-12	20097	41.7	32 3	18 5			323	217	210		
GN212-12	20078	38.0	32.7	32 0	37 5	42 5	323	238	183	182	190
GN212-12	20079	35.0	30 3	28 0	36 0	29 5	323	252	190	251	225
GN212-12	20080	37.0	34.1	36 5	27 5		337	211	196	203	
GN212-12	20081	30.6	32.7	27 0		33 5	323	238	218		
GN212-12	20082	32.0	33 0	39 0	33 0	37 5	344	217	197	266	245
GN212-12	20083	38.9	40 3	37 0	45 5		355	206	189	197	
GN212-12	20084	32.1	27.6	30 0	33 0	32 0	330	190	224	238	256
GN212-12	20085	32.6	33 8	32 0	34 0		323	238	197	224	
GN212-12	20086	27.8	25.7	25 5	21 0	27 5	323	238	204	273	147
GN212-12	20087	37.0	29 8		41 0	38 0	362	245			198
Average		33.1	29 9	31 5	29 8	35.7	327.3	212.4	206.1	206 2	213.5
SD		4.7	5.7	8.8	8.0	5.4	11 0	22 5	13.1	38.2	33.9

SD = standard deviation.

Table SR3-5. Bunch weight and cycle time from QCAV-4 plants

Genotype	Plant ID	Bunch weight (kg)					Cycle 1 (days)	Cycle 2 (days)	Cycle 3 (days)	Cycle 4 (days)	Cycle 5 (days)
		Plant crop	Ratoon 1	Ratoon 2	Ratoon 3	Ratoon 4	Plant crop	Ratoon 1	Ratoon 2	Ratoon 3	Ratoon 4
QCAV-4	20267	27.8	19 0	21 0	30 5	33 5	330	277	179	161	193
QCAV-4	20268	27.5	14 0		22.7	30 0	344	204			214
QCAV-4	20269	24.3	24 8	28.7	33 0	35 0	355	206	183	161	178
QCAV-4	20270	26.6	23 3	20 0	31 5		362	213	204	275	
QCAV-4	20271	32.7	27 5		33 8	36 0	355	193			178
QCAV-4	20272	28.0	25 9	42 3	28 5	32 0	344	176	199	157	207
QCAV-4	20273	31.4	20.7	30 5	25 5	29 0	330	231	189	189	186
QCAV-4	20275	31.6	21 9	41 0		35 5	372	134	238		
QCAV-4	20276	29.2	28 5	34 2	33 5	36 0	344	186	214	161	192
QCAV-4	20278	36.7	25 5	32.7	28 0	32 5	344	186	198	169	193
QCAV-4	20254	32.1	20.6	25 0	22.7	30 5	330	245	197	182	210
QCAV-4	20255	24.6	13 5				337	193			
QCAV-4	20256	28.6	14 8				330	277			
QCAV-4	20257	21.3	13.6		14 0		323	225			
QCAV-4	20261	30.8	16.6	29 5	27.1	40 0	323	207	220	141	175
QCAV-4	20262	30.1	23 3	36 0	28.7		330	183	215	156	
QCAV-4	20263	39.4	24 5	37 5	29 5		355	175	214	161	
QCAV-4	20264	34.4	29 0	34 3		42 0	330	200	198		
QCAV-4	20265	31.7	28 8	36 5	30 5	37 0	330	210	188	169	219
QCAV-4	20266	35.0	25 0	37 8	40 5	39 0	337	193	198	156	151
QCAV-4	20231	24.7	24.1		28 5		330	210	204	161	
QCAV-4	20232	25.6	24 8	34 0	27 5		330	190	208	156	
QCAV-4	20233	26.0	21 2		24 5	28 0	330	231			161
QCAV-4	20234	25.3	24.7	32 5	34 5		330	176	244	169	
QCAV-4	20235	25.8	25 2	27 0	21 0		330	218	196	175	
QCAV-4	20236	27.3	26 9	32 5	25 5		330	218	196	175	
QCAV-4	20237	23.6	24 5	31 0	24 5	35 0	323	217	179	165	182
QCAV-4	20241	21.7	20.6	40 0	28 5	33 5	317	189	213	150	204
QCAV-4	20242	25.4	24.1	25 5	36 2	40 5	317	189	203	167	183
QCAV-4	20243	20.6	17 8	35 0	30 0	31 0	355	151	213	143	140
QCAV-4	20244	24.4	19.7	24 0	26 0	39 0	344	217	197	181	216
QCAV-4	20245	26.2	34 5	33 8	34 8		330	168	211	182	
QCAV-4	20246	27.3	27 0	27 0	25 0	37 0	330	183	215	211	201
QCAV-4	20247	28.7	31.1	36 0	22 5		317	189	213	165	
QCAV-4	20248	28.3	28 3	34 5	27 8	37 0	317	189	222	163	225
QCAV-4	20249	27.2	24 3	31 2	28.1	30 0	317	189	203	153	176
QCAV-4	20250	30.1	32 5	38 0	33 5	39 0	317	203	199	165	104
QCAV-4	20251	29.9	32 0		33 3	37 5	317	160			306
QCAV-4	20252	28.4	26 5	28 3	30 0	36 0	323	175	221	165	199
QCAV-4	20253	15.8	27 8	28 5	25 5	25 5	317	196	196	153	192
QCAV-4	20221	28.9	22 9	15 0	23 5		323	207	249	245	
QCAV-4	20222	20.5	27 3	32 0	31 8		344	204	210	196	
QCAV-4	20223	27.4	26 0	39 0	28 5	39 0	317	181	221	178	281
QCAV-4	20224	30.6	25.6	37 0	30 0	36 0	317	181	221	165	206
QCAV-4	20225	33.1	21 0	22 0	31 3	34 0	317	244	197	224	108
QCAV-4	20226	31.7	36 2	30 0	27 0		317	203	208	211	
QCAV-4	20227	29.0	20.1	27 0	26 5	35 0	323	197	208	211	216
QCAV-4	20228	32.9	30 2	36 0		37 0	317	189	213		
QCAV-4	20229	26.3	23 9	34 0	31 5	32 0	323	183	213	165	206
QCAV-4	20230	28.0		33 0	29 0	33 5	330			165	199
Average		28.1	24 3	31.7	28.6	34 8	331.1	199.2	207 3	174 9	193.6
SD		4.3	5.1	6.0	4.6	3.8	13.6	27 0	15.2	27.3	39.5

SD = standard deviation.

Table SR3-6. Plant girth and height comparison between non-GM GN212-12 and QCAV-4 plants

non-GM GN212-12 plants		
Plant ID	**Girth (mm)**	**Height (mm)**
20052	675	2,350
20064	650	2,370
20068	660	2,370
20072	870	2,810
20074	870	2,630
20075	800	2,520
20083	880	2,810
20086	695	2,430
20091	805	2,640
20093	830	2,530
Average	**774**	**2,546**
SD	**94**	**173**
QCAV-4 plants		
Plant ID	**Girth (mm)**	**Height (mm)**
20221	680	2,280
20222	700	2,570
20223	760	2,450
20224	770	2,680
20225	720	2,450
20227	870	2,630
20228	780	2,510
20229	740	2,680
20230	810	2,830
20233	750	2,620
20235	690	2,550
20237	740	2,540
20242	795	2,555
20245	660	2,490
20248	830	2,840
20249	840	2,820
20250	770	2,780
20252	700	2,670
20253	830	2,660
20254	730	2,600
20264	810	2,690
20265	720	2,580
20268	740	2,490
20269	750	2,580
20271	710	2,610
20272	690	2,540
20273	830	2,770
20275	750	2,550
20276	750	2,730
20278	780	2,690
Average	**757**	**2,615**
SD	**53**	**127**

Data collected from either ratoon 6 or 7 based on availability.
SD = Standard deviation.

Table SR3-7. Plant botanical descriptors comparison between non-GN212-12 and QCAV-4 plant

Comparison of 'QCAV-4' to parent 'Cavendish Grand Nain'*

Organ/Plant Part: Context	'QCAV-4'	'Cavendish Grand Nain'
Ploidy:	triploid	triploid
Pseudostem: overlapping of leaf sheaths	weak	weak
Pseudostem: tapering	absent or weak	absent or weak
Pseudostem: colour	purple	purple
Pseudostem: anthocyanin colouration	medium to strong	medium
Pseudostem: colour of inner side of basal sheath	purple	purple
Plant: compactness of crown	compact	compact
Plant: growth habit	drooping	drooping
Petiole: attitude of wings at base	curved outwards	curved outwards
Leaf blade: colour of midrib on lower side	green	green
Leaf blade: shape of base	both sides acute	both sides acute
Leaf blade: waxiness on lower side	medium	weak to medium
Leaf blade: width	broad	broad
Leaf blade: glossiness of upper side	absent	absent
Peduncle: diameter	large	large
Peduncle: pubescence	present	present
Peduncle: curvature	medium to strong	medium to strong
Bunch: length	long	long
Bunch: shape	cylindrical	cylindrical
Bunch: attitude of fruits	moderately turned up	moderately turned up
Bunch: compactness	medium	medium
Bunch: number of hands	many	many
Rachis: attitude of male part	vertical	vertical
Rachis: prominence of scars	weak	weak
Rachis: persistence of bracts	absent or weak	absent or weak
Rachis: persistence of hermaphrodite flowers	present	present
Fruit: colour of peel (before maturity)	greenish yellow (RHS 141C)	greenish yellow (RHS 141C)
Fruit: persistence of floral organs	present	present
Male inflorescence: persistence	present	present
Male inflorescence: shape	narrow ovate	narrow ovate
Male inflorescence: opening of bracts	closed or slightly open	closed or slightly open
Bract: colour of inner side	orange red	orange red
Bract: shape of apex	broad acute	broad acute

*For most characteristics, 4-5 individual plants were assessed.

US Plant Patent No: PP34,398 P3.

Table SR3-8. QCAV-4 fruit characteristics measurements as part of US Plant Patent No: PP34,398 P3

Plant ID	Bunch characteristics			Fruit characteristics			
	Length (cm)	Diameter (cm)	Number of hands	Length (cm)	Width (cm)	Pedicel length (cm)	Peel thickness (mm)
20221	88.3	122.6	11.0	16.0	3.5	2.8	3.8
20223	72.0	123.0	8.0	14.6	3.2	3.4	4.1
20229	83.0	127.0	10.0	14.4	3.4	2.9	3.2
20231	75.0	113.0	10.0	14.2	3.4	3.0	3.2
20241	72.0	120.0	10.0	13.6	3.4	3.2	3.3
20242	77.5	110.0	10.0	14.0	3.5	2.6	3.0
20247	76.0	132.0	10.0	13.8	3.3	3.4	3.5
20248	82.0	142.0	10.0	15.4	3.9	3.3	3.5
20251	88.0	122.5	11.0	15.4	3.3	3.3	3.7
20255	59.0	118.0	8.0	13.7	3.4	2.2	3.2
20257	88.5	122.3	10.0	14.5	3.6	2.5	4.0
20263	83.0	128.0	10.0	13.8	3.3	3.2	4.5
20268	80.0	146.0	9.0	15.5	3.7	3.2	3.1
20269	88.5	126.0	11.0	14.0	3.4	3.8	3.2
20273	84.0	120.0	11.0	13.0	2.8	3.1	2.7
Average ± SD (n=15)	**79.8 ± 8.2**	**124.8 ± 9.6**	**9.9 ± 1.0**	**14.4 ± 0.8**	**3.4 ± 0.2**	**3.1 ± 0.4**	**3.5 ± 0.5**

SD = Standard deviation.

QUT2023-4: Concentrations of MamRGA2 and NPTII proteins in fruit and peel tissues from event QCAV-4 and dietary exposure estimations

STUDY IDENTIFICATION

QUT2023-4

AUTHORS

[redacted]
s 22

PERFORMING LABORATORY

[redacted]

[redacted] 2023

4.1 Introduction

Banana event QCAV-4 (QUT-QCAV4-6) was created by *Agrobacterium tumefaciens*-mediated transformation of banana (*Musa acuminata* subgroup Cavendish cv Grand Nain) embryogenic cells with plasmid pSAN3. This plasmid contains the *neomycin phosphotransferase* II (*npt*II) selectable marker gene from the Tn5 transposon of *Escherichia coli* strain K12 and the *MamRGA2* disease resistance (R) gene from the wild banana *Musa acuminata* ssp. *malaccensis* which confers resistance to the devastating banana fungal pathogen *Fusarium oxysporum* f. sp. *cubense* tropical race 4 (TR4) (Dale *et al*., 2017).

To estimate potential human and environmental exposure to the MamRGA2 and neomycin phosphotransferase II (NPTII) proteins expressed in QCAV-4, the concentrations of these proteins in edible parts of the plant was determined by a combination of techniques including Western blot analysis, densitometry and quantitative enzyme-linked immunosorbent assay (ELISA).

4.2 Materials and Methods

4.2.1 Protein standards

The NPTII protein standard was obtained from Agdia (Cat. #LST 73000) while the MamRGA2 protein was expressed and semi-purified "in house" using the pET28a(+) expression system (Cat.#69864, Novagen, Merck) and the expression host *E. coli* Rosetta 2 (DE3) (Cat. #71397, Novagen, Merck). The characterisation and equivalence of the MamRGA2 protein derived from this *E. coli* expression system to MamRGA2 expressed in banana event QCAV-4 are detailed in study report QUT2023-8.

4.2.2 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot analysis

For SDS-PAGE, 4-20% Mini-PROTEAN TGX Precast Protein gels (Cat. #4561095, Bio-Rad) were used followed by staining with Coomassie brilliant blue R-250 (Cat. #0472-25G, Amresco). Gels were de-stained in a solution 10% acetic acid and 50% methanol for 2-3 h followed by deionised water for 1-2 h and imaged on a Chemidoc MP imaging system (Bio-Rad). An estimate of the molecular weight was determined relative to the Colour Protein Standard markers, Broad Range (11-245 kDa or 10-250 kDa) (Cat. #P7712S, P7719S, New England Biolabs) included on all gels.

For Western immunoblot analysis, duplicate gels were used, and proteins transferred to nitrocellulose membrane (Cat. #1704158, Bio-Rad), blocked in 3% bovine serum albumin (BSA)/Tris-buffered saline with Tween® 20 (TBST) for 1 h prior to incubation with 1:1,000 diluted (in 3% BSA/TBST) primary antibodies for 1 h. The primary mouse anti-NPTII antibody was sourced from the NPTII quantification ELISA kit (Cat. #PSP 73000, Agdia). Since antibodies against MamRGA2 are not commercially available, a primary mouse anti-MamRGA2 monoclonal antibody designated 17F07 was custom made by Maine Biotechnology Services (MBS, BBI Solutions) and supplied at a concentration of 0.7 mg/mL. For details regarding the manufacture of 17F07, refer to study report QUT2023-8. Incubation for 1 h with 1:5,000 diluted (in 3% BSA/TBST) goat-anti-mouse IgG-HRP (Cat. #62-6520, ThermoFisher), was used for secondary detection. Chemiluminescence was visualised using the Clarity™ Western ECL Substrate (Cat. #170-5060, Bio-Rad) and imaged on the Chemidoc MP imaging system (Bio-Rad) at various durations of exposure.

4.2.3 Estimation of MamRGA2 protein concentration by densitometry

For densitometric analysis, BSA (A7906, Sigma) and semi-purified MamRGA2 were prepared in 2x Laemmli sample buffer (Laemmli, 1970). A standard curve consisting of a serial doubling dilution of BSA (5 - 0.3 µg) and triplicate samples of semi-purified MamRGA2 (5 µL) were subjected to SDS-PAGE and stained with

Coomassie. Gels were imaged on a Chemidoc MP imaging system (Bio-Rad) and densitometric analysis performed using the Image Lab Software, v6.1 (Bio-Rad). A linear regression was made from the intensity of the BSA bands and used to extrapolate the concentration of MamRGA2 within the semi-purified MamRGA2 extracts.

4.2.4 Limit of detection (LOD) of the 17F07 mouse anti-MamRGA2 monoclonal antibody

A doubling dilution series of a known amount of MamRGA2 (from 31.3 ng - 244 pg) was prepared in 2x Laemmli sample buffer from a semi-purified stock of *E. coli*-expressed MamRGA2 (0.4 µg/µL established by densitometry - study report QUT2023-8). Proteins were subjected to SDS-PAGE and Western blot analysis and the LOD visually established from the digital image generated.

4.2.5 Limit of quantitation (LOQ) of the MamRGA2 protein in banana tissue

The limit of MamRGA2 protein quantitation in banana tissue was determined using Western blot analysis and ripe non-GM Grand Nain banana fruit protein extracts spiked with various amounts of *E. coli*-expressed MamRGA2. Protein was extracted from 80 mg of freeze-dried ripe non-GM banana fruit tissue (WT GN 20062-G1), finely homogenized using the TissueLyser II (85300, QIAGEN), by resuspending in 300 µL protein extraction buffer (25 mM phosphate buffer pH 7.0, 100 mM NaCl, 1 mM EDTA, 1% Triton X-100, 50 µM ascorbic acid, 10 mM β-mercaptoethanol and cOmplete Mini Protease Inhibitor Cocktail (04693124001, Merck)). Samples were mixed on the TissueLyser II at 30 Hz for 30 s, followed by 20 Hz for 5 min and then centrifuged at 20,000 x*g* for 10 min at 4˚C to pellet the insoluble debris. Solubilised protein was transferred to a fresh tube and total soluble protein (TSP) concentration was measured at 4.4 µg/µL using the Bio-Rad Protein Assay (Cat. #500-0006, Bio-Rad), following the manufacturer's instructions (total amount of TSP from 80 mg of freeze-dried ripe banana was 4.4 µg/µL x 300 µL = 1,320 µg).

A volume equivalent to 35 µg of TSP (~8 µL) was spiked with 25, 12.5, 6.3, 3.1, 1.6, 0.8, 0.4, 0.2, 0.1 and 0 ng of *E. coli*-expressed MamRGA2. Samples were subjected to SDS-PAGE and Western blot analysis and the LOQ visually established from the digital image generated.

4.2.6 Detection of MamRGA2 and NPTII proteins in edible tissue of event QCAV-4

Ripe fruit and peel samples were obtained from three independent QCAV-4 plants in addition to non-GM GN212-12 control plants. Total protein extracts were prepared from these samples as described in 4.2.5 except for peel tissue which was resuspended in 600 µL protein extraction buffer. The TSP concentration was measured as described in 4.2.5. For the detection of each protein (MamRGA2 and NPTII), a duplicate set of samples was prepared at a concentration of 35 µg TSP (ripe fruit) and 15 µg TSP (ripe peel) and subjected separately to SDS-PAGE. One gel was stained with Coomassie to assess protein loading uniformity while the second gel was subjected to Western blot analysis.

4.2.7 Quantification of NPTII in edible tissue of event QCAV-4 by ELISA

TSP extracted from QCAV-4 and non-GM control GN212-12 banana fruit and peel tissue was analysed using a commercially available NPTII ELISA kit (Cat. #PSP 73000) following the manufacturer's instructions. Briefly, 80 mg of freeze-dried fruit and peel tissue was resuspended in 800 µL and 1200 µL of 1x Protein Extraction Buffer 1 (PEB1, supplied with kit), respectively. Samples were placed on a rotator at 4°C for 1 h with occasional tapping and vortexing to resuspend the powder, before centrifugation at 20,000 x*g* for 15 min at 4°C to pellet the debris. The concentration of TSP was determined using a commercial Protein Assay (Cat. #500-0006, Bio-Rad) with BSA as a standard. The supernatants from sample extraction were diluted 1/200 in 1x PEB1 to fall within the range of the standard curve. For each ELISA, a standard curve was generated with known amounts of NPTII protein standard (Cat. #LST 73000) prepared in 1/200 dilution of

fruit or peel TSP. NPTII expression was quantitatively evaluated at 650 nm using the GloMax Discover spectrophotometer (Promega).

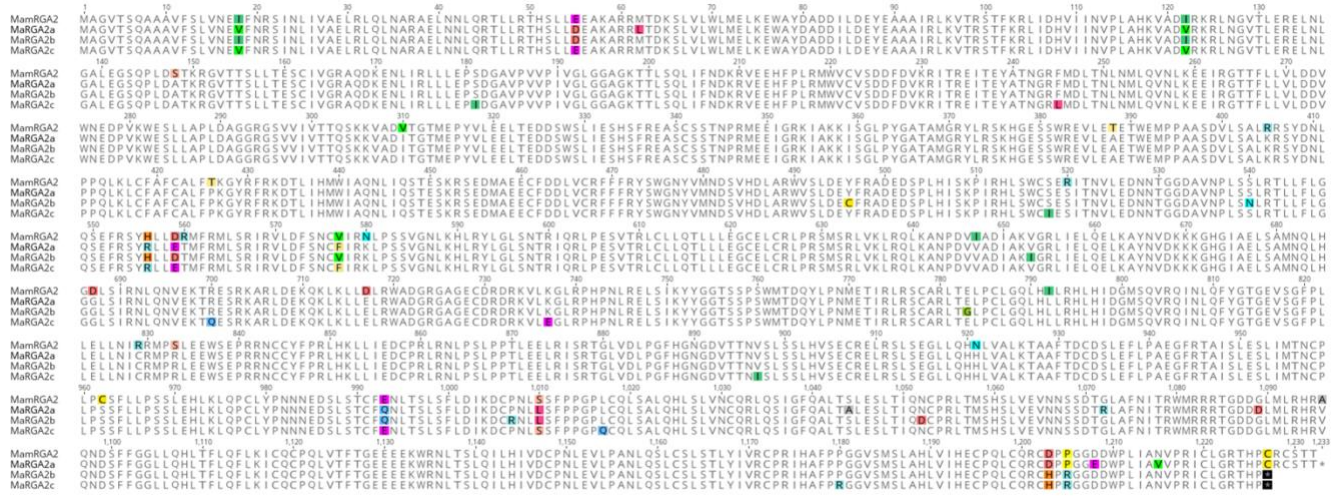
4.3 Results

4.3.1 Tissue specificity and concentrations of MamRGA2 in QCAV-4

MamRGA2 is a 1,232 amino acid protein with a predicted molecular weight of 139.6 kDa. The triploid Cavendish banana genome contains three alleles of endogenous *MaRGA2* which are translated into peptides with more than 97.3% amino acid identity to MamRGA2 (Dale *et al*., 2017). For the detection and quantification of MamRGA2 expression in edible parts of event QCAV-4, an antibody-based approach was used. Because of the large number of CC-NBS-LRR proteins in banana and the high likelihood of cross-reactivity, a mouse anti-MamRGA2 monoclonal antibody (designated 17F07) was generated for the purpose of specifically detecting MamRGA2 in event QCAV-4 while avoiding cross-detection of the endogenous MaRGA2 protein as well as other R proteins. Similarities between MamRGA2 and endogenous MaRGA2 amino acid sequences are shown in Figure SR4-1.

To assess the limit of detection (LOD) of the 17F07 antibody, Western blot analysis was done using semi-purified *E. coli*-expressed MamRGA2 (His_6-tag, ~142 kDa) which represented about 7% of the total protein concentration based upon densitometry (study report QUT2023-8). A serial doubling dilution of MamRGA2 protein solution was separated by SDS-PAGE, transferred to nitrocellulose membrane, and incubated with the 17F07 antibody. This analysis (Figure SR4-2A) revealed an MamRGA2 LOD between 0.98 and 1.9 ng (~1-2 ng, lanes 7 and 6, respectively). The limit of quantification (LOQ) was subsequently calculated to be around 45-88 ppm of total soluble protein (TSP) by Western blot using serially diluted MamRGA2 protein spiked into ripe banana fruit protein extracts (Figure SR4-2B). Indeed, the Western analysis presented in Figure SR4-2B demonstrates that MamRGA2 could still be detected between 1.6 - 3.1 ng (lanes 7 and 6, respectively). Based on 35 µg of TSP loaded in each lane, the LOQ_{TSP} is between 1.6 ng/35 µg and 3.1 ng/35 µg or 45-88 ppm TSP. To express the LOQ with respect to the mass of dry and fresh banana tissue, the dry weight equivalent of 35 µg of TSP was first calculated. The TSP yield from 80 mg of freeze-dried ripe banana (calculated at 1,320 µg in section 2.5) was used to calculate the dry weight equivalent of 35 µg of TSP loaded on each lane at 2.1 mg (80 mg x 35 µg/1320 µg). Since banana fruit tissue contains on average 75% moisture, this was equivalent to 8.4 mg of fresh ripe banana fruit. With this information, the LOQ based on dry tissue mass was calculated as 1.6-3.1 ng/2.1 mg or 0.8-1.5 ppm while the LOQ based on fresh tissue mass was calculated as 1.6-3.1 ng/8.4 mg or 190-370 ppb. In summary, the LOD of *E. coli*-expressed MamRGA2 with the 17F07 mouse anti-MamRGA2 monoclonal antibody was shown to be between 1 and 2 ng, while the LOQ of MamRGA2 in a banana fruit matrix was calculated to be 0.8-1.5 ppm or 190-370 ppb on a dry or fresh weight basis, respectively.

The tissue-specific expression of MamRGA2 in edible parts of event QCAV-4 was subsequently investigated. TSP extracts were made from ripe fruit flesh (35 µg) and peel tissues (15 µg) collected from non-GM GN212-12 plants as well as QCAV-4 and used for SDS-PAGE followed by Western immunoblot



Sequence	MamRGA2	MaRGA2a	MaRGA2b	MaRGA2c
MamRGA2		97.5	97.6	97.3
MaRGA2a	97.5		98.4	98.4
MaRGA2b	97.6	98.4		98.2
MaRGA2c	97.3	98.4	98.2	

Figure SR4-1. Protein alignment showing differences between MamRGA2 and endogenous MaRGA2 sequences. Top: Global alignment with Blosum62 cost matrix generated in Geneious, dissimilar amino acids are highlighted in color. Left: percent identity matrix.

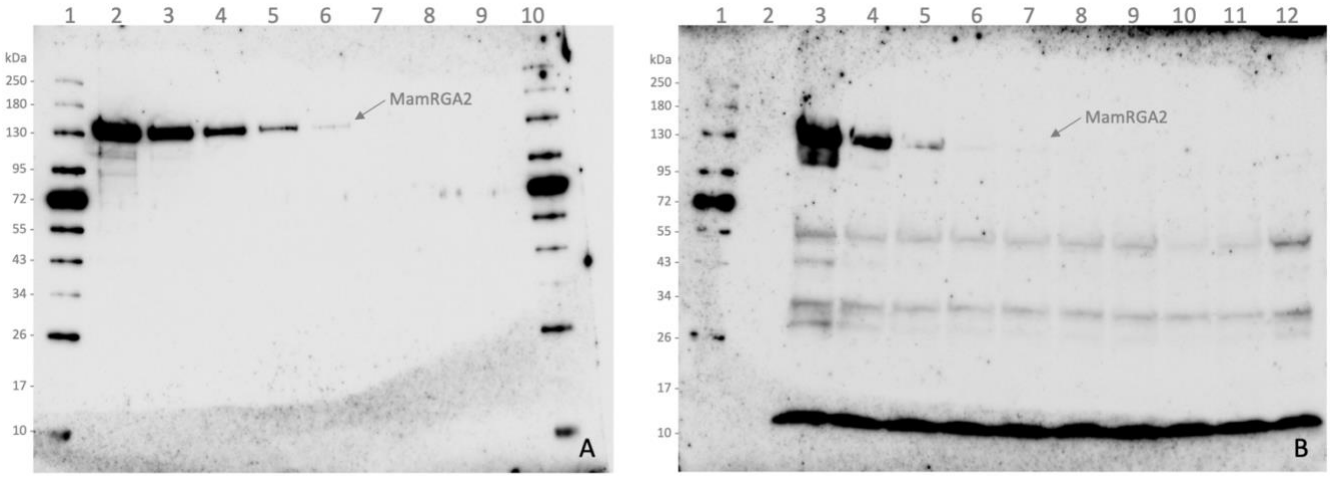


Figure SR4-2. Protein detection efficacy by Western immunoblotting. A: Limit of detection (LOD) of *E. coli*-expressed MamRGA2 (139.6 kDa). Lanes 1 and 10, Colour Pre-stained Protein Standard, Broad Range (10-250 kDa) (Cat. #P7719S, New England Biolabs) and lanes 2-9, *E. coli*-expressed MamRGA2 at 31.3 ng, 15.6 ng, 7.8 ng, 3.9 ng, 1.9 ng, 0.98 ng, 0.49 ng and 0.24 ng, respectively. B: Limit of quantitation (LOQ) of *E. coli*-expressed MamRGA2 spiked into ripe banana fruit total soluble protein (TSP). Lane 1, Colour Pre-stained Protein Standard, Broad Range (10-250 kDa); lane 2, empty; lanes 3-11, 35 µg TSP from wild-type GN212-12 fruit spiked with 25 ng, 12.5 ng, 6.3 ng, 3.1 ng, 1.6 ng, 0.8 ng, 0.4 ng, 0.2 ng and 0.1 ng, respectively, and lane 12, 35 µg TSP from wild-type GN212-12 fruit (unspiked). Primary mouse anti-MamRGA2 monoclonal antibody 17F07 (MBS, BBI Solutions), 1:1,000 followed by secondary goat-anti-mouse IgG-HRP (Cat. #62-6520, ThermoFisher), 1:5,000, chemiluminescent substrate development and exposure time of 2 h. Low molecular weight band (~15 kDa) due to non-specific binding of the secondary IgG.

analysis (Figure SR4-3). MamRGA2 could not be detected in any of the edible plant parts of event QCAV-4 tested and therefore its concentration is believed to be below the lower LOQ of 190 ppb in fresh banana fruit tissue.

Interestingly, a low molecular band of ~15 kDa was observed in all banana fruit protein extracts in Western blot analysis suggesting non-specific binding of the secondary goat-anti mouse IgG to an endogenous banana protein. To investigate this further, the Western blot analysis previously described for the detection of MamRGA2 in banana fruit tissue (Figure SR4-3) was repeated but using only the secondary goat-anti mouse IgG for detection (Figure SR4-4). This result confirmed the non-specific binding of the goat-anti mouse IgG to an endogenous ~15 kDa banana fruit protein of unknown identity.

4.3.2 Estimated maximum human dietary exposure to MamRGA2

In Australia, 95% of households purchase bananas with an average yearly per capita consumption of 16 kg for 2020/21 (Hort Innovation, 2022). Since bananas are almost exclusively consumed as fresh fruit and the amount of MamRGA2 protein present in edible parts of event QCAV-4 was established below the lower LOQ of 190 ppb (or 190 μg/kg), individuals consuming banana fruit in Australia would be exposed to less than 3.04 mg of MamRGA2 protein per year (or less than 8.3 μg/day), assuming that 100% of the Australian Cavendish market was replaced with event QCAV-4. Data on the consumption of banana peel in Australia is unavailable but is considered to be marginal in comparison resulting in even lower exposure to the MamRGA2 protein from QCAV-4 peel consumption.

4.3.3 Tissue specificity and concentrations of NPTII in QCAV-4

Since the NPTII amino acid sequence expressed in QCAV-4 is nearly identical (99.6%) to the one expressed in several GM events already assessed by FSANZ (Beck *et al*., 1982), its safety assessment was limited to two types of analysis, (i) an updated bioinformatics comparison of its amino acid sequence to known protein toxins and allergens (study report QUT2023-7) and (ii) the detection and quantification of the amount of NPTII protein present in edible parts of the plant by Western blot analysis and quantitative enzyme-linked immunosorbent assay (ELISA). Western immunoblot analysis using a commercially available NPTII-specific antibody revealed a band of the expected molecular weight for NPTII (29.1 kDa) in both fruit and peel samples collected from event QCAV-4 (Figure SR4-5B, lane 2-7) but absent from the wild-type GN212-12 (Figure SR4-5B, lane 8-9).

To estimate human dietary exposure to the NPTII enzyme expressed in QCAV-4, its concentration in edible parts of the plant was determined by quantitative ELISA. These analyses were performed on the same ripe fruit and peel samples analysed previously by Western blotting and presented in Figure SR4-5. On average, the concentration of NPTII in fresh ripe fruit and peel samples analysed was 3.1 and 4.5 ppm, respectively (Table SR4-1).

4.3.4 Estimated maximum human dietary exposure to NPTII

As per section 4.3.2, human dietary exposure to NPTII was calculated based on an average yearly per capita consumption of fresh banana fruit of 16 kg for 2020/21 (Hort Innovation, 2022) at 49.6 mg per year (or 136 μg/day). This scenario again assumes a 100% shift of the Australian Cavendish market to event QCAV-4. Dietary exposure to NPTII through the consumption of QCAV-4 banana peel is difficult to establish because of the lack of reliable data on the consumption of this tissue in Australia. If consumption of banana peel was similar to the fruit flesh (16 kg/per/year) then the exposure would be 72 mg per year (or 197 μg/day).

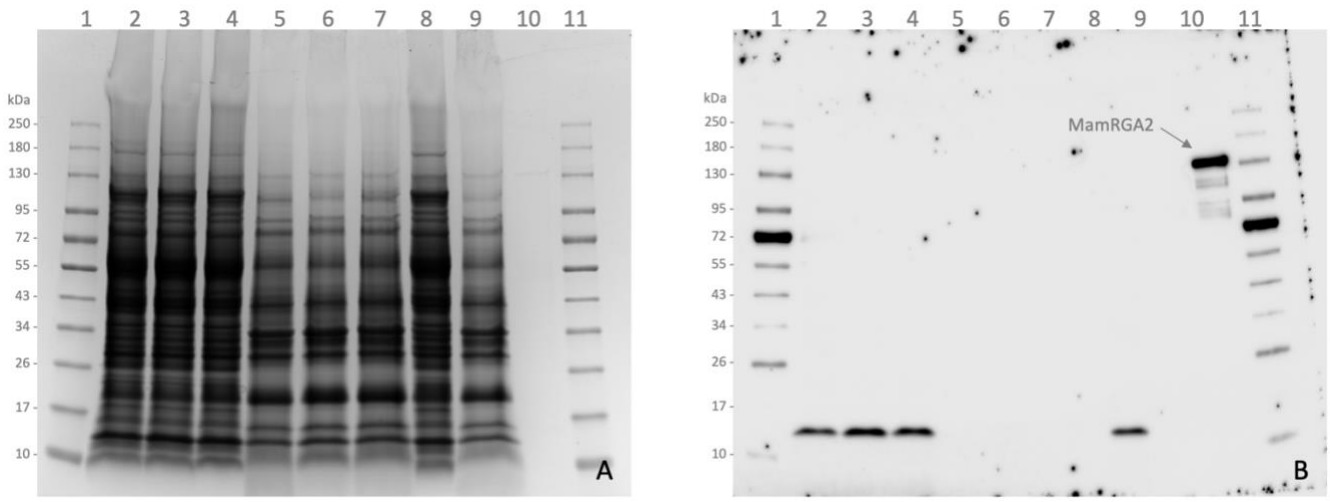


Figure SR4-3. Detection of MamRGA2 in banana fruit tissue of event QCAV-4. A: 4-20% Mini-PROTEAN TGX Precast Protein Gel, Coomassie brilliant blue R-250 stain (Cat. #0472-25G, Amresco) and B: Western immunoblot. Detection with primary mouse anti-MamRGA2 monoclonal antibody 17F07 (MBS, BBI Solutions), 1:1,000 followed by secondary goat-anti-mouse IgG-HRP (Cat. #62-6520, ThermoFisher), 1:5,000, chemiluminescent substrate development and auto optimal exposure time of 436.7 s (~7.3 min). Lane 1 and 11, Colour Pre-stained Protein Standard, Broad Range (10-250 kDa) (Cat. #P7719S, New England Biolabs); lanes 2-4, 35 µg total soluble protein (TSP) from ripe fruit of QCAV-4 plants 20224, 20229 and 20230; lane 5-7, 15 µg TSP from ripe peel of QCAV-4 plants 20234, 20236 and 20246; lane 8-9, TSP from ripe fruit (35 µg) and peel (15 µg) of wild-type GN212-12 (20077), respectively; lane 10, *E. coli*-expressed MamRGA2 (15.6 ng, 139.6 kDa). Low molecular weight band (~15 kDa) due to non-specific binding of the secondary IgG.

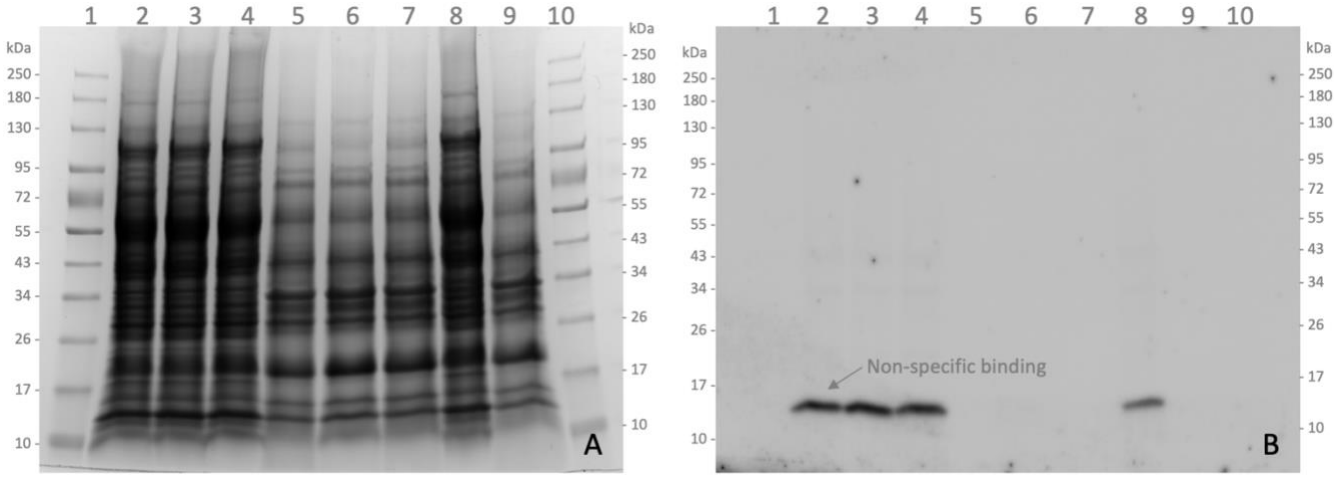
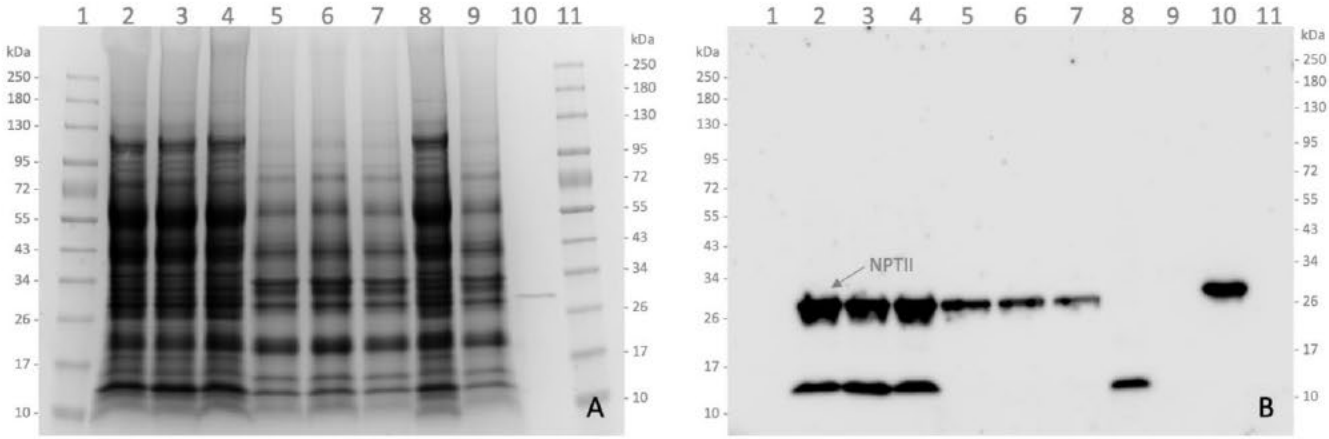


Figure SR4-4. Non-specific binding of secondary goat-anti mouse IgG. A: 4-20% Mini-PROTEAN TGX Precast Protein Gel, Coomassie brilliant blue R-250 stain (Cat. #0472-25G, Amresco) and B: Western immunoblot. Lanes 1 and 10, Colour Pre-stained Protein Standard, Broad Range (10-250 kDa) (Cat. #P7719S, New England Biolabs); lanes 2-4, 35 µg total soluble protein (TSP) from ripe fruit of QCAV-4 plants 20224, 20229 and 20230; lanes 5-7, 15 µg TSP from ripe peel of QCAV-4 plants 20234, 20236 and 20246; lanes 8-9, TSP from ripe fruit (35 µg) and peel (15 µg) of wild-type GN212-12 plant 20077, respectively. Detection with goat-anti-mouse IgG-HRP (Cat. #62-6520, ThermoFisher), 1:5,000, chemiluminescent substrate development and auto optimal exposure time of 30 min.



FigureSR4-5. Detection of NPTII in edible banana tissue of event QCAV-4. A: 4-20% Mini-PROTEAN TGX Precast Protein Gel, Coomassie brilliant blue R-250 stain (Cat. #0472-25G, Amresco) and B: Western immunoblot. Detection with primary anti-NPTII detection antibody (Cat. #PSP 73000, Agdia), 1:1,000 followed by secondary goat-anti-mouse IgG-HRP (Cat. #62-6520, ThermoFisher), 1:5,000, chemiluminescent substrate development and auto optimal exposure time of 30 min. Lanes 1 and 11, Colour Pre-stained Protein Standard, Broad Range (10-250 kDa) (Cat. #P7719S, New England Biolabs); lanes 2-4, 35 µg total soluble protein (TSP) from ripe fruit of QCAV-4 plants 20224, 20229 and 20230; lanes 5-7, 15 µg TSP from ripe peel of QCAV-4 plants 20234, 20236 and 20246; lanes 8-9, TSP from ripe fruit (35 µg) and peel (15 µg) of wild-type GN212-12 plant 20077, respectively; lane 10, 113 ng of NPTII protein standard (Cat. #LST 73000, Agdia). Low molecular weight band (~15 kDa) due to non-specific binding of the secondary IgG.

Table SR4-1. NPTII concentration measured in event QCAV-4 edible tissues by ELISA

Tissue	Sample	NPTII ng/mg (DW)			NPTII % TSP	NPTII (ng/mg FW)*
		Rep.	Average	SD		
Fruit	20224	11.5	12.4	0.6	0.08	3.1
	20229	12.6			0.07	
	20230	12.9			0.09	
Peel	20234	19.0	18.1	0.7	0.15	4.5
	20236	18.0			0.10	
	20246	17.2			0.15	

Average NPTII background readings recorded from respective non-GM GN212-12 plant 20077 tissues were subtracted from all QCAV-4 readings. DW = dry weight, FW = fresh weight, SD = standard deviation, Rep. = replicate, *Assume 25% dry matter content.

4.4 Conclusions

Assuming 100% of the Australian Cavendish market was replaced with event QCAV-4, the maximum human exposure to MamRGA2 and NPTII through the consumption of fresh banana fruit was calculated to be lower than 8.3 µg/day and 136 µg/day, respectively. Consumption of banana peel in Australia is almost certainly considerably less than fruit resulting in even lower exposure to the MamRGA2 protein than from QCAV-4 fruit consumption. Regarding NPTII, if consumption of banana peel was similar to the fruit, then the maximum human exposure to NPTII would be 72 mg per year (or 197 µg/day).

4.5 Supporting documents

- No additional documents are provided with study report QUT2023-4.

QUT2023-5: Differential messenger RNA (mRNA) expression of *MamRGA2*, *npt*II and the endogenous *MaRGA2* in event QCAV-4

STUDY IDENTIFICATION

QUT2023-5

AUTHORS

[redacted]
s 22
[redacted]

PERFORMING LABORATORY

[redacted]

-

Novogene
[redacted]

[redacted] 2023

5.1 Introduction

Banana event QCAV-4 (QUT-QCAV4-6) was created by *Agrobacterium tumefaciens*-mediated transformation of banana (*Musa acuminata* subgroup Cavendish cv Grand Nain) embryogenic cells with plasmid pSAN3. Analysis of the nucleotide sequence of the insert in QCAV-4 revealed a single large 26,849 bp insert on Chromosome 6 containing three intact copies of the CC-NBS-LRR resistance (R) gene *MamRGA2* and the *npt*II selectable marker gene. Details of this analysis can be found in study report QUT2023-1.

Part of the safety assessment of event QCAV-4 requires quantification of the amounts of newly expressed proteins produced in edible parts (fruit and peel) of the plant to estimate their human dietary exposure (study report QUT2023-4). Since the amount of MamRGA2 protein in edible portions of QCAV-4 remained below the limit of detection (LOD), this study report investigates the presence and amount of mRNA produced from the *MamRGA2* transgene (relative to that of the three endogenous *MaRGA2* alleles) as well as the *npt*II selectable marker gene.

5.2 Materials and Methods

5.2.1 RNA extraction

High quality RNA was isolated from banana leaf, root and fruit tissue. Leaf and root samples were collected from three 12-week-old QCAV-4 plants and three 12-week-old non-GM Grand Nain control (GN212-12) plants grown under controlled conditions (27°C, 16 h photoperiod) while fruit samples were collected from a single field-grown QCAV-4 and GN212-12 plant. To 100 mg of freeze-dried leaf and root tissue and 30 mg of freeze-dried fruit tissue was added 600 µL of PureLink™ Plant RNA Reagent (Invitrogen, 12322-012), vortexed thoroughly and allowed to stand for 5 min. Samples were centrifuged 13,000 x*g* for 2 min and supernatant removed, to which 120 µL of 5 M NaCl and 360 µL of chloroform was added and samples kept on ice for 10 min before centrifugation at 13,000 x*g* for 5 min at 4°C. RNA was then further purified from the supernatant using an RNeasy Mini Kit (Qiagen, 74104) as per the manufacturer's instructions and including on-column DNase digestion (Qiagen, 79254). RNA concentration and quality was measured using both Nanodrop spectrophotometry and electrophoresis through a 2% agarose gel. Purified RNA samples (> 4 µg) were stored in RNAstable® (Biomatrica, 93221-001) and sent to Novogene for analysis. Additional quality controls, library construction and sequencing parameters internal to Novogene are described in the final "QC Analysis Report" provided as SD5-1 (section 5.5).

5.2.2 Bioinformatic analysis

Extracted RNA was sequenced on an Illumina NovaSeq6000 S4 flow cell at 150 bp paired-end (PE) reads s 22 Illumina RNA-Seq datasets of non-GM control (GN212-12) and transgenic (QCAV-4) leaf (3 replicates), root (3 replicates) and fruit (1 replicate) were mapped to the *MamRGA2* and *npt*II ORFs from pSAN3 as well as the three endogenous *MaRGA2* alleles (alleles a, b and c) and a single copy house-keeping gene (*cyclophilin*; *CyP*). Mappings were done using the STAR aligner on QUT's High Performance Computing (HPC) cluster, with reads mapped as pairs with default parameters. The number of reads mapped to each sequence was recorded and compiled into Table SR5-1. Raw mapped read numbers were normalised for library size (data output size) using the sample with the largest data output (S_4L3) followed by normalisation for abundance of the *CyP* reference gene.

Table SR5-1. RNA-Seq data and STAR aligner mapping outputs

Genotype	Tissue	Sample ID	Replicate	Data output (number of reads)	Number of reads mapping to					
					MamRGA2	*npt* II	*MaRGA2a*	*MaRGA2b*	*MaRGA2c*	*CyP*
GN212-12	Leaf	CL1	1	270,677,278	0	8	0	0	0	87,514
		CL2	2	295,854,422	0	8	0	0	0	136,106
		CL3	3	270,312,958	0	0	0	0	2	126,868
	Root	CR1	1	300,942,562	0	6	204	76	148	244,448
		CR2	2	280,792,854	0	0	64	42	112	195,970
		CR3	3	268,404,728	0	10	152	50	154	202,482
	Fruit	S_20566	1	121,350,076	0	0	4	0	72	21,788
QCAV-4	Leaf	S_4L1	1	268,119,840	35,400	1,043,361	4	0	0	233,400
		S_4L2	2	306,157,394	59,678	919,668	6	0	4	198,616
		S_4L3	3	335,663,842	28,174	771,628	0	2	0	239,782
	Root	S_4R1	1	274,556,348	53,766	497,638	240	116	192	210,132
		S_4R2	2	270,628,374	43,326	444,776	204	84	270	225,484
		S_4R3	3	282,458,650	49,828	442,012	304	106	388	190,762
	Fruit	S_20221	1	123,121,508	178	182,400	0	0	56	18,346

5.3 Results

5.3.1 RNA-Seq analysis

An RNA-Seq approach was used to assess the relative amounts of leaf, root and fruit mRNA transcripts, originating from the *MamRGA2* and *npt*II transgenes and from the endogenous *MaRGA2* gene, in QCAV-4 and the non-GM Grand Nain control. Both leaf and root tissue were collected from three 12-week-old plants from both the non-GM control (GN212-12) and QCAV-4 as well as one ripe fruit sample collected from both a non-GM control and QCAV-4 plant in the field. High quality RNA was extracted from each of these tissues and Illumina RNA-Seq data generated. After mapping the raw RNA-Seq reads from each sample independently to the nucleotide sequences of *MamRGA2*, *npt*II, all three alleles of *MaRGA2* and the *CyP* reference gene, the number of reads was recorded and normalised as described in section 5.2.2. The expression of the endogenous *MaRGA2* alleles was negligible in both GN212-12 and QCAV-4 with the expression levels considered as "noise" in comparison to *MamRGA2* and *npt*II (Table SR5-2). As expected, no reads from the seven libraries of GN212-12 tissues mapped to *MamRGA2* (Table SR5-2). Although a very small number of reads (22) mapped to *npt*II, these were most likely due to the presence of endogenous bacteria known to be present in banana tissue (Martínez *et al.*, 2003). The highest relative level of expression of *MamRGA2* in QCAV-4 was in root tissue (24x), followed by leaf (20x) with negligible levels in fruit tissue (1x). In the fruit of QCAV-4, *MamRGA2* expression was about 3.2-fold higher than the expression of all three endogenous *MaRGA2* alleles combined. Interestingly, the expression of the *npt*II selectable marker gene was highest in the fruit of QCAV-4 and about a 1,000x higher than the expression of the *MamRGA2* transgene.

Table SR5-2. Tissue-specific RNA-Seq differential analysis in non-GM control GN212-12 versus QCAV-4

Genotype	Tissue	Normalised average number of reads mapped to		
		Transgene		Endogenous
		MamRGA2	*npt*II	*MaRGA2*
GN212-12 (non-GM control)	Leaf*	0	15	2
	Root*	0	7	450
	Fruit^	0	0	1,019
QCAV-4	Leaf*	55,482	1,199,830	7
	Root*	69,077	648,468	902
	Fruit^	2,835	2,905,092	892

*Averaged from data sets from three plants, ^single plant dataset

5.4 Conclusions

In summary, the abundance of *MamRGA2* transcripts in the fruit of QCAV-4 is relatively low compared to other tissues and is of a similar order of magnitude as its endogenous counterparts. These results are consistent with previous analysis presented in study report QUT2023-4 which demonstrated that the amount of MamRGA2 protein in edible portions of QCAV-4 is extremely low and, in fact, below the level of detection (LOD). In contrast, the relative amounts of transcripts originating from the *npt*II selectable marker gene were high, especially in the fruit of QCAV-4 and consistent with that expected of a transgene controlled by the strong, constitutive CaMV35S promoter in banana.

5.5 Supporting documents

- SD5-1-RNA-Seq-[redacted] QC reports.pdf
- SD5-2-RNA-Seq analysis on HPC.xlsx (contains raw mapping output data and calculations).

QUT2023-6: Safety assessment of the seven new ORFs identified in event QCAV-4

STUDY IDENTIFICATION

QUT2023-6

AUTHORS

[redacted]
s 22

PERFORMING LABORATORY

[redacted]

-

[redacted]

[redacted] 2023

6.1 Introduction

Banana event QCAV-4 (QUT-QCAV4-6) was created by *Agrobacterium tumefaciens*-mediated transformation of banana (*Musa acuminata* subgroup Cavendish cv Grand Nain) embryogenic cells with plasmid pSAN3. Analysis of the nucleotide sequence of the insert in QCAV-4 revealed a single large 26,849 bp insert on Chromosome 6. To investigate the possibility of novel open reading frames (ORFs) resulting from the presence of the insert in QCAV-4, an ORF analysis was conducted to identify potential start-to-stop ORFs within the entire insert and including both the 3' 4,211 bp and 5' 3,576 bp chromosome 6 sequences spanning the insert. This analysis assessed each of three possible reading frames in both orientations for potential ORFs capable of encoding sequences of 30 or more amino acids. Seven new and unintended ORFs resulting from the presence of the insert in QCAV-4 were identified from this analysis the details of which can be found in study report QUT2023-1.

The presence of unwanted and uncharacterised ORFs can potentially pose a biosafety issue that is addressed in this study. The transcriptional potential of each of these new ORFs was initially assessed using an *in silico* analysis to identify putative upstream and downstream *cis*-regulatory elements, followed by the analysis of RNA-Seq data from various tissues. Finally, in cases where these ORFs were potentially translated, the deduced amino acid sequences were analysed for homology to known allergenic or toxic proteins.

6.2 Materials and Methods

6.2.1 *In silico* regulatory motif searches

For each of the seven newly identified ORFs, 1,000 bp of upstream ("promoter area") and 300 bp of downstream ("3'UTR area") nucleotide sequences were extracted from Geneious Prime® version 2022.2.1 in ".fasta" format and used separately for *in silico* analysis. Searches for plant promoter-like sequences and transcription factor binding sites were done using PlantCARE (Lescot *et al*., 2002) and the TSSP Prediction of PLANT Promoters algorithm accessible from Softberry (Solovyev *et al.*, 2010). Searches for 3' UTR-like sequences were done using the POLYAH algorithm available in Softberry (Salamov and Solovyev, 1997).

6.2.2 RNA extraction

Leaf and root samples were collected from three 12-week-old QCAV-4 plants and three 12-week-old non-GM control GN212-12 plants grown under controlled conditions (27°C, 16 h photoperiod) while fruit samples were collected from a single field grown QCAV-4 and GN212-12 plant. To extract high quality RNA, 100 mg of freeze-dried leaf and root tissue and 30 mg of freeze-dried fruit tissue was mixed with 600 µL of PureLink™ Plant RNA Reagent (Invitrogen, 12322-012), vortexed thoroughly and allowed to stand for 5 min. Samples were centrifuged 13,000 x*g* for 2 min and the supernatant removed, to which 120 µL of 5 M NaCl and 360 µL of chloroform was added and samples kept on ice for 10 min before centrifugation at 13,000 x*g* for 5 min at 4°C. RNA was then further purified from the supernatant using an RNeasy Mini Kit (Qiagen, 74104) as per the manufacturer's instructions and including on-column DNase digestion (Qiagen, 79254). RNA concentration and quality was measured using both Nanodrop spectrophotometry and electrophoresis through a 2% agarose gel. Purified RNA samples (> 4 µg) were stored in RNAstable® (Biomatrica, 93221-001) and sent to s 22 for analysis. Additional quality controls, library construction and sequencing parameters internal to [redacted] are described in the final "QC Analysis Report" provided as SD5-1 (section 5.5 in study report QUT2023-5).

6.2.3 Bioinformatic analysis

Extracted RNA was sequenced on an Illumina NovaSeq6000 S4 flow cell at 150 bp paired-end (PE) reads s 22 Illumina RNA-Seq datasets of non-GM control (GN212-12) and QCAV-4 leaf (3 replicates each), root (3 replicates each) and ripe fruit (1 replicate each) were mapped to the nucleotide sequence of the 26,849 bp insert in QCAV-4 together with its 3' 4,211 bp and 5' 3,576 bp flanking chromosome 6 sequences. Mappings were done in Geneious with the Geneious mapper using default parameters in "medium sensitivity".

6.2.4 Allergenicity searches

In silico analyses were performed to compare the predicted amino acid (Figure SR6-1) sequences of the seven new ORFs to known allergenic proteins in the Food Allergy Research and Resource Program (FARRP) dataset, which is available through www.AllergenOnline.org (University of Nebraska). Full-length sequence searches were done using the FASTA36 alignment tool and significant homology determined when greater than 50% homology was found between the query sequence and a database entry and the E-value was less than 1×10^{-4}. Sliding window searches (80-mer) were done using FASTA36 and significant homology determined when there was greater than 35% homology between the query sequence and a database entry and an E-value lower than 1×10^{-4}. Exact 8-mer match searches were done using FASTA36 and reported at 100% identity.

6.2.5 Toxicity searches

Potential structural similarities shared between the predicted amino acid sequences of the seven new ORFs and proteins known or suspected to be toxins were evaluated in Geneious using the Basic Local Alignment Search Tool (BLAST) and a toxin protein BLAST database created in Geneious. The database was generated from a subset of sequences derived from the UniProt Knowledgebase, comprised of 568,002 manually annotated and reviewed sequences from Swiss-Prot and 226,771,948 automatically annotated, un-reviewed sequences from TrEMBL, that were selected using a keyword search on toxins (KW-0800). At the time of its creation (August 29, 2022), the collection contained a total of 92,851 sequences, including 7,523 reviewed sequences from Swiss-Prot and 85,328 un-reviewed sequences from TrEMBL. Matches between the query sequence and a database entry were considered significant at an E-value lower than 1×10^{-5}.

6.3 Results

To address the transcriptional potential (mRNA expression) of the seven newly identified ORFs, a two-stage approach was undertaken and described in sections 6.3.1 and 6.3.2 below. The allergenicity and toxicity potential of each of the new ORFs was also assessed in sections 6.3.3 and 6.3.4, respectively.

6.3.1 *In silico* functional motifs analysis

The upstream and downstream sequences of each of these sequences were scrutinised for the presence of both 5' and 3' regulatory elements which could result in the transcription of these ORFs. *In silico* analyses of *cis*-elements upstream (1,000 bp) and downstream (300 bp) of each newly identified ORF using tools available on PlantCARE and Softberry were inconclusive because of the high number of small motif sequences returned from each of these searches. Examples of search results outputs are given in Figure SR6-15 to SR6-25 in section 6.5. The significance of these small motifs in the context of gene expression is difficult to assess.

>ORF_111_(frame_1)
MHVMLYSWIRRGREDDSGGSIRITHYYGQFKLKAGANSH*

>ORF_87_(frame_1)
MCYSDRSSRVVFPAPPNPTIEHHMHSGIIENKNLKFSTEKCFVIVRRLVHKTENVK*

>ORF_7_(frame_2)
MDRHLKSRIRFWFKQQWPRQLNNTLRCKQIDA*

>ORF_151_(frame_2)
MWVCVSDDFDVKRITREITEYATNGRFMDLTNLNMLQVNLKEEIRGTTFLLVLDDVWNEDPVKWESLLAPLDAGGRGSVVIVTTQSKKVADVTGTMEPYVLEELTED
DSWSLIESHSFREASCSSTNPRMEEIGRKIAKKISGLPYGATAMGRYLRSKHGESSWREVLETETWEMPPAASDVLSALRRSYDNLPPQLKLCFAFCALFTKGYRFRKDTL
IHMWIAQNLIQSTESKRSEDMAEECFDDLVCRFFFRYSWGNYVMNDSVHDLARWVSLDEYFRADEDSPLHISKPIRHLSWCSERITNVLEDNNTGGDAVNPLSSLRTL
LFLGQSEFRSYHLLDRMFRMLSRIRVLDFSNCVIRNLPSSVGNLKHLRYLGLSNTRIQRLPESVTRLCLLQTLLLEGCELCRLPRSMSRLVKLRQLKANPDVIADIAKVGRLIE
LQELKAYNVDKKKGHGIAELSAMNQLHGDLSIRNLQNVEKTRESRKARLDEKQKLKLLDLRWADGRGAGECDRDRKVLKGLRPHPNLRELSIKYYGGTSSPSWMTDQ
YLPNMETIRLRSCARLTELPCLGQLHILRHLHIDGMSQVRQINLQFYGTGEVSGFPLLELLNIRRMPSLEEWSEPRRNCCYFPRLHKLLIEDCPRLRNLPSLPPTLEELRISRT
GLVDLPGFHGNGDVTTNVSLSSLHVSECRELRSLSEGLLQHNLVALKTAAFTDCDSLEFLPAEGFRTAISLESLIMTNCPLPCSFLLPSSLEHLKLQPCLYPNNNEDSLSTCF
ENLTSLSFLDIKDCPNLSSFPPGPLCQLSALQHLSLVNCQRLQSIGFQALTSLESLTIQNCPRLTMSHSLVEVNNSSDTGLAFNITRWMRRRTGDDGLMLRHRAQNDSFF
GGLLQHLTFLQFLKICQCPQLVTFTGEEEEKWRNLTSLQILHIVDCPNLEVLPANLQSLCSLSTLYIVRCPRIHAFPPGGVSMSLAHLVIHECPQLCQHVPGTFGHP*

>ORF_156_(frame_3)
MRFLPEVSACPWHIWSSMNALSCVSMSLAHLVIHECPQP*

>ORF_71_(frame_3)
MPLPTVVPKMDPHPRGASWKKKTFQPRLQSKWIDVNMLEQLWRIYCGVNKLTLRQLNNTLRTFLMY*

>ORF_49_(frame_3)
MTKCARDMLTQLRAFMDDQMCQGHADTSGRKRMDSWASDDVQGAEGAEALQVCRQYLQVWTINDVQNLKRSKVSPLLFLFAGEGYELWTLADL*

Figure SR6-1. Amino acid sequences of the seven newly identified ORFs in QCAV-4.

In a 2014 review, Hernandez-Garcia and Finer state that "… the small size of motif sequences that are recognized during genome-wide analyses using current prediction algorithms, frequently leads to the identification of a tremendous number of putative elements. Moreover, the presence of DNA sequence motifs alone is not sufficient to identify functional protein binding sites, which is highly influenced by other several factors. Many copies of a short sequence motif can be present in large genomes; however, pending position and accessibility, only a small portion of those copies may be functional and enable *in vivo* protein binding." Further, the likelihood of the presence and adequate functional location of upstream and downstream *cis*-regulatory elements working in conjunction to promote the expression of these new ORFs is very remote (Porto *et al*., 2014).

6.3.2 RNA-Seq analysis

Based on the results from the functional motifs analysis, an RNA-Seq approach was used to definitively assess the expression of these new ORFs. Both leaf and root tissue were collected from three 12-week-old plants from both the non-GM control (GN212-12) and QCAV-4 as well as one ripe fruit sample from both a non-GM control and QCAV-4 fruit in the field. High quality RNA was extracted from each of these tissues and Illumina RNA-Seq data generated. Each data set was mapped separately in Geneious to a 34,636 bp sequence containing the 26,849 bp sequence of the insert and the 3' 4,211 and 5' 3,576 bp chromosome 6 flanking sequences, to "trap" any 150 bp PE reads originating from a sequence transcribed from the insert. The interpretation of these mapping results is largely visual and needs to be assessed in the context of the arrangement of the insert. Therefore, supporting raw data and mapping files can be supplied at FSANZ's request. For demonstration purposes and to give some context to the sequences in the vicinity of the new ORFs, the Geneious mapping output (number of reads mapped to the insert and Chr06 flanks) is summarised in Table SR6-1.

Table SR6-1. Output of the Geneious mapping of RNA-Seq samples to the insert in QCAV-4

Genotype	Tissue	Sample ID	Replicate	Data output (number of reads)	Number of reads mapping to					% of reads mapping to	
					QCAV-4 insert with flanks	QCAV-4 insert only	*nptII*	*MamRGA2*	*Macma4_06_g29410.1* *	*npt* II	*MamRGA2*
Non-GM control GN212-12	Leaf	CL1	1	270,677,278	412	21	8	0	391	38.1	0.0
		CL2	2	295,854,422	432	23	10	0	409	43.5	0.0
		CL3	3	270,312,958	360	14	0	2	346	0.0	14.3
	Root	CR1	1	300,942,562	710	455	6	439	255	1.3	96.5
		CR2	2	280 792 854	593	235	0	219	358	0.0	93.2
		CR3	3	268,404,728	849	405	10	383	444	2.5	94.6
	Fruit	S_20566	1	121,350,076	130	84	0	81	46	0.0	96.4
QCAV-4	Leaf	S_4L1	1	268,119,840	1,029,853	1,029,257	987,696	37,682	596	96.0	3.7
		S_4L2	2	306,157,394	1,383,543	1,383,179	1,314,699	63,092	364	95.0	4.6
		S_4L3	3	335,663,842	1,166,313	1,165,314	1,130,085	30,391	999	97.0	2.6
	Root	S_4R1	1	274,556,348	771,808	771,466	711,129	57,683	342	92.2	7.5
		S_4R2	2	270,628,374	679,912	679,558	630,459	46,891	354	92.8	6.9
		S_4R3	3	282,458,650	710,540	710,088	652,255	55,165	452	91.9	7.8
	Fruit	S_20221	1	123,121,508	244,678	244,493	242,777	279	185	99.3	0.1

**Macma4_06_g29410.1* is located 2,309 bp upstream of the insertion site on the Chr06 5' flanking sequence

In all tissues of the non-GM control GN212-12, the presence of *npt*II transcripts was virtually undetectable. Using "low sensitivity" settings, the Geneious mapper could not discriminate between reads originating from the endogenous *MaRGA2* or the insert *MamRGA2* sequences. Therefore, endogenous *MaRGA2* reads were barely detectable in root and fruit tissue of GN212-12 and not detected from leaf tissue (Table SR6-1). As an example, a visualisation of the mapping obtained from a leaf, root and fruit control sample (CL1, CR1 and S_20566) is shown in Figure SR6-2 and SR6-3 where all reads originating from the endogenous *MaRGA2* can be seen randomly assigned to a region of *MamRGA2*. In contrast, large numbers of reads were mapped to the insert from all samples of QCAV-4 with, on average, 96.0, 92.3 and 99.3% of the reads originating from the *npt*II sequence (Table SR6-1).

By focussing on the areas of the QCAV-4 insert where each of the seven new ORFs were predicted, it was possible to demonstrate that none of these ORFs were transcribed in any of three types of QCAV-4 tissue examined (leaf, root and ripe fruit). A mapping result convincingly demonstrating the expression of any ORF is validated by two criteria, (i) depth of mapped reads below the sequence commensurate to the level of expression in that tissue and (ii) contiguity of mapped reads across and beyond the sequence. As an example, Figure SR6-4 shows the mapping depth and contiguity (and therefore expression) of the endogenous Macma4_06_g29410.1 located 2,309 bp upstream of the insertion site on the Chr06 5' flanking sequence. For ORFs 7 and 71, not all of the criteria are met with only a few reads fully or partially mapping to these sequences as opposed to the large number of reads 5' of these sequences originating from the expression of the *npt*II transgene and into the CaMV35S 3' UTR (Figure SR6-5 and SR6-6). ORFs 49, 151 and 156 originate from a truncated region of the *MamRGA2* transgene. The mapping visualisation presented in Figure SR6-7 shows that in the leaf of QCAV-4 but not in the fruit, a large number of reads mapped to the area without contiguity across the lengths of all three ORFs. This indicates that all these reads originate from the expression of the other three intact copies of *MamRGA2* present on the insert and have been wrongly (randomly) allocated in this area by Geneious. Finally, no reads mapped to either ORFs 87 or 111 in both the leaf and fruit of QCAV-4 as shown in Figure SR6-8.

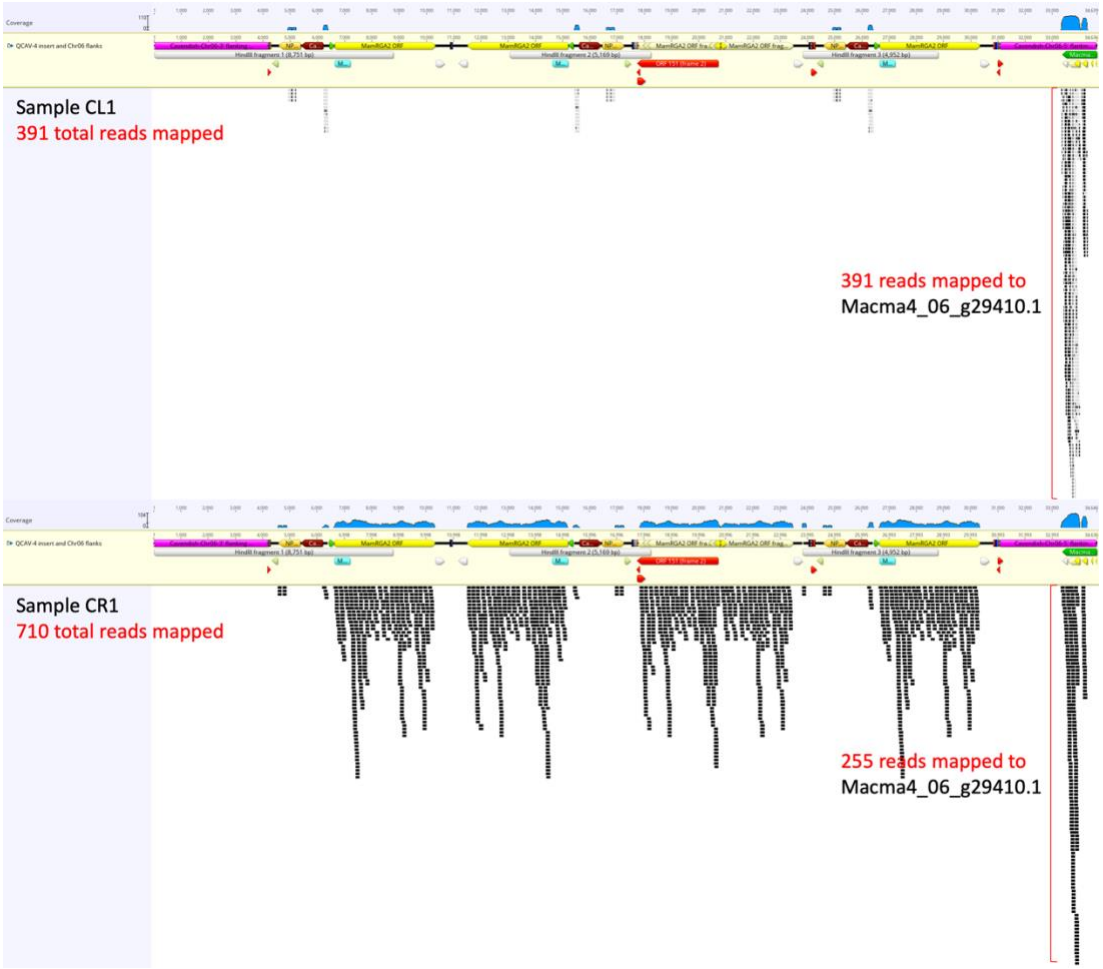


Figure SR6-2. Geneious mapping output for samples CL1 (GN212-12 leaf) and CR1 (GN212-12 root).

Figure SR6-3. Geneious mapping output for samples S_4L1 (QCAV-4 leaf), S_4R1 (QCAV-4 root), S_20221 (QCAV-4 fruit) and S_20566 (GN212-12 fruit).

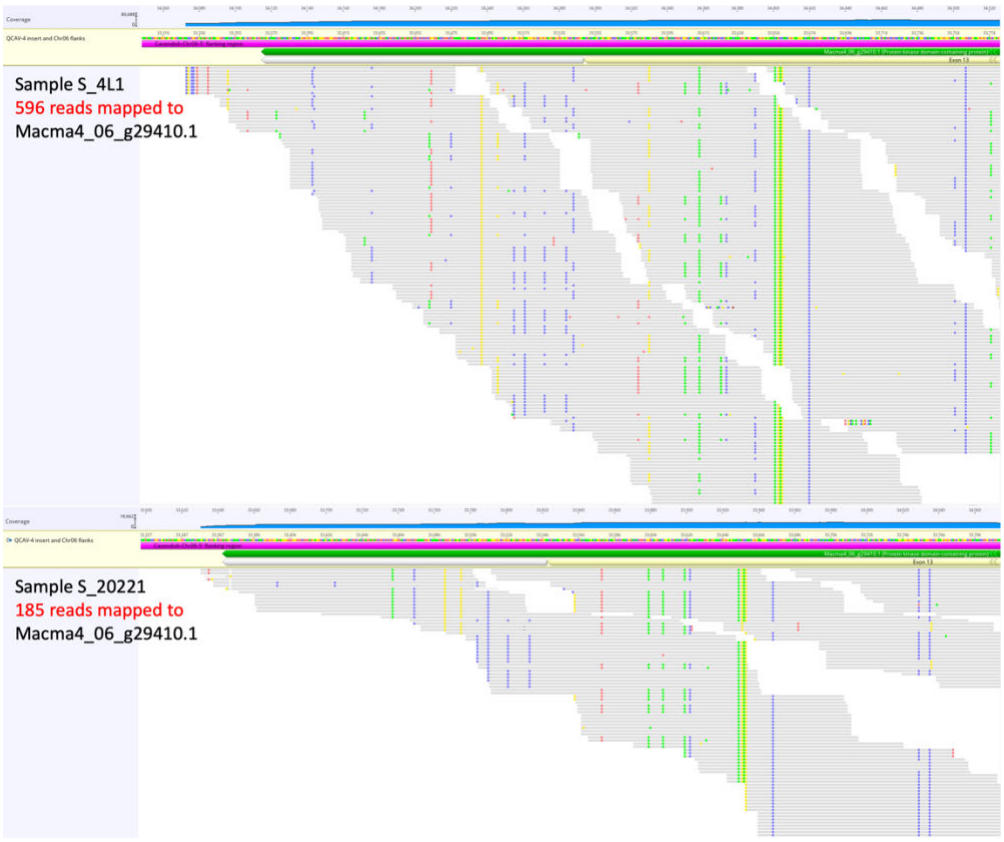


Figure SR6-4. Geneious mapping output focused on ORF Macma4_06_g29410.1 in samples S_4L1 (QCAV-4 leaf) and S_20221 (QCAV-4 fruit).

Figure SR6-5. Geneious mapping output focused on ORF 7 in samples S_4L1 (QCAV-4 leaf) and S_20221 (QCAV-4 fruit).

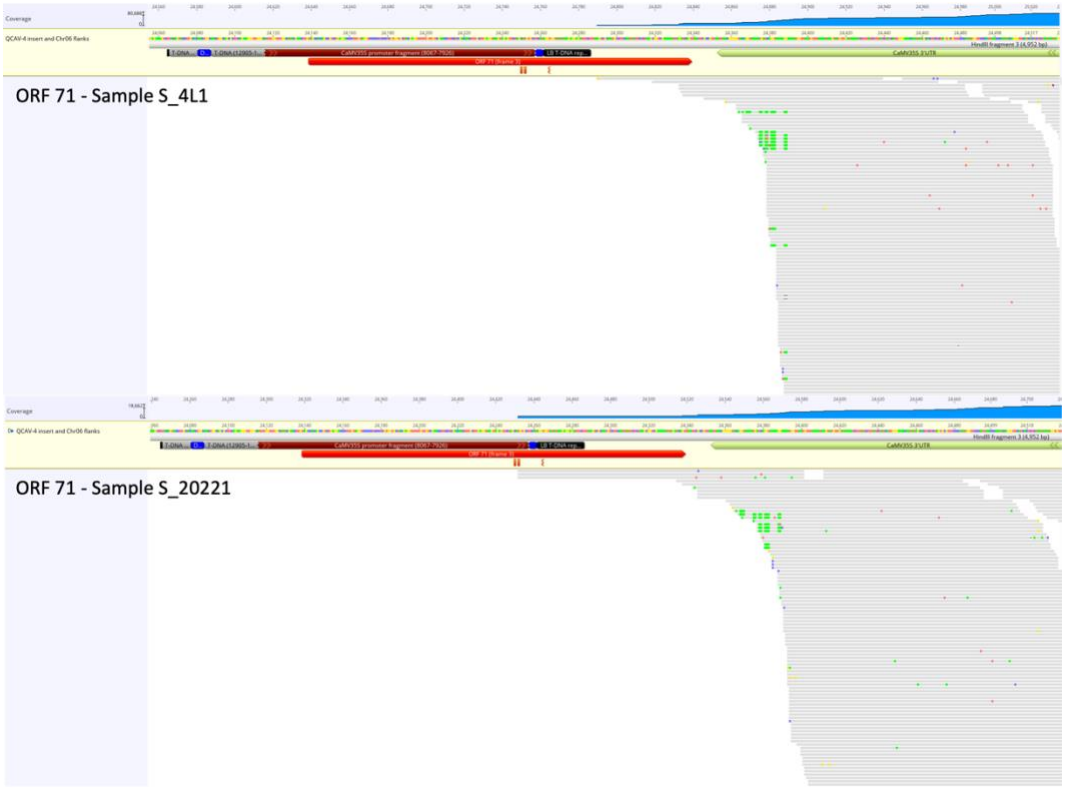


Figure SR6-6. Geneious mapping output focused on ORF 71 in samples S_4L1 (QCAV-4 leaf) and S_20221 (QCAV-4 fruit).

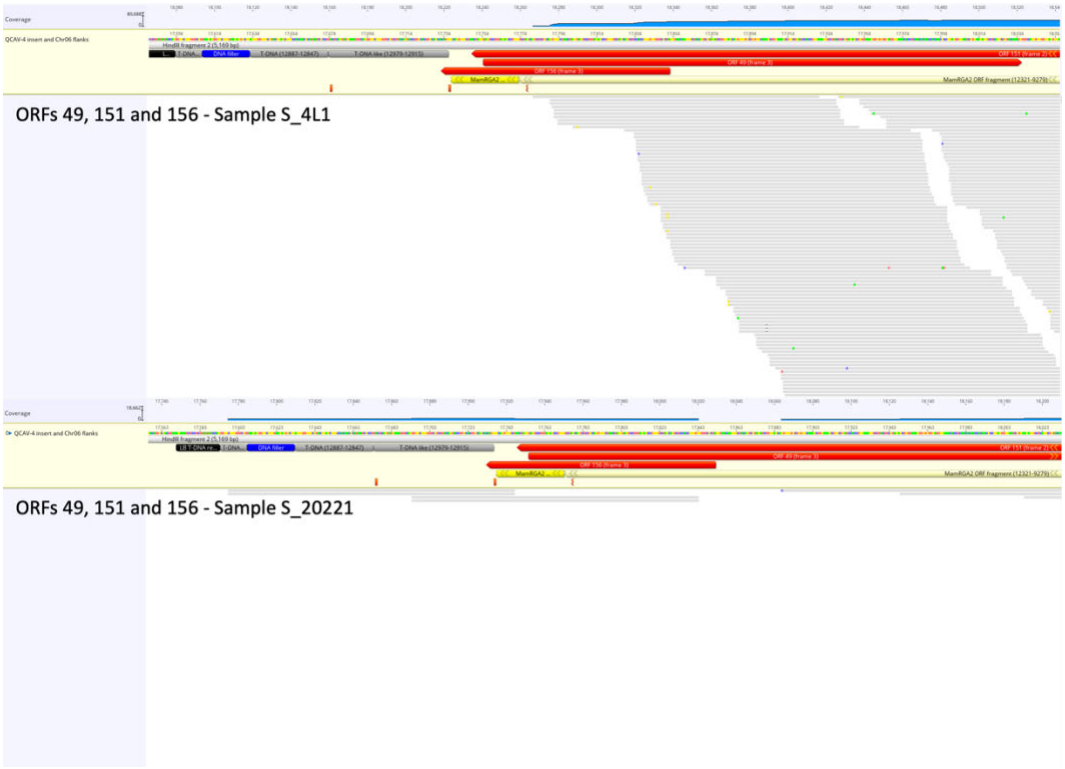


Figure SR6-7. Geneious mapping output focused on ORFs 49, 151 and 156 in samples S_4L1 (QCAV-4 leaf) and S_20221 (QCAV-4 fruit).

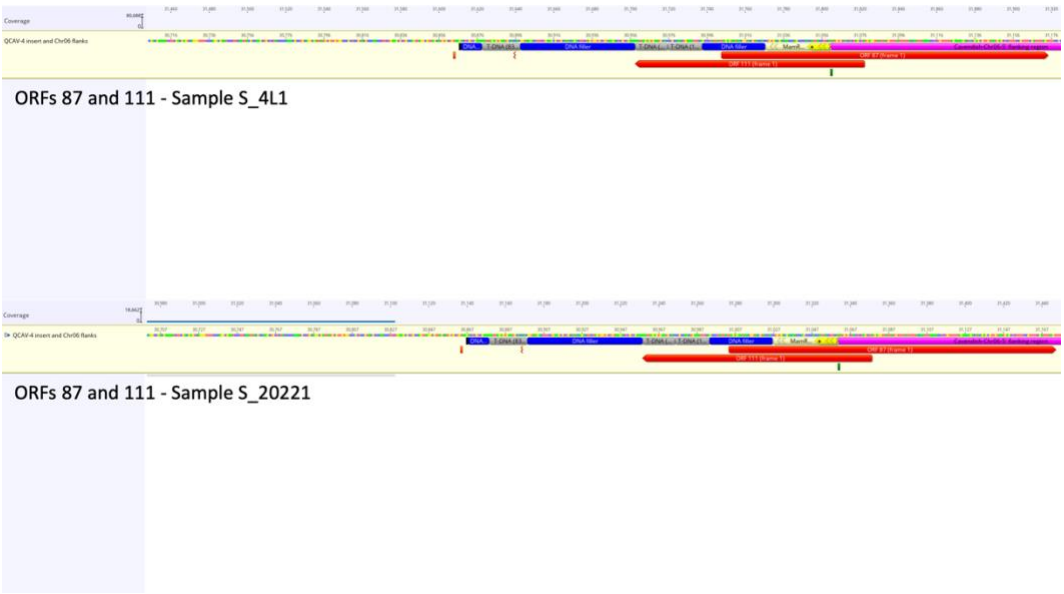


Figure SR6-8. Geneious mapping output focused on ORF 87 and 111 in samples S_4L1 (QCAV-4 leaf) and S_20221 (QCAV-4 fruit).

6.3.3 Assessment of the allergenicity potential of the seven new ORFs

Protein characterisation (study report QUT2023-4) and RNA-Seq analysis of *MamRGA2* (study report QUT2023-5) both indicated minimal exposure to this protein from the consumption of fruit and peel of event QCAV-4. Although exposure to MamRGA2 is likely to be negligible, *in silico* analyses were still performed to compare the predicted amino acid sequences of the seven new ORFs (Figure SR6-1) to known allergenic proteins in the FARRP dataset. The version of the database used (v21) contains 2,233 protein (amino acid) sequence entries that are categorised into 913 taxonomic-protein groups of unique proven or putative allergens (food, airway, venom/salivary and contact) and celiac protein sequences from 430 species. Three types of analyses were performed for this comparison on October 27, 2022 and as described in materials and methods section 2.1: the full-length sequence (E-value < 10^{-4}), 80-mer sliding window (35% homology with E-value < 10^{-4}) and 8-mer exact match searches did not identify similarities between any of the new ORFs and known allergens in the FARRP database (Table SR6-2-8) indicating that exposure to proteins resulting from the unlikely expression of any of these new ORFs is not of allergenicity concern.

Full result outputs from the allergenicity assessment of the seven new ORFs are presented in section 6.5 of this study report.

Table SR6-2. Search results using ORF 7 query sequence against the FARRP database on October 27, 2022

Database	AllergenOnline Database v21 (February 14, 2021)
Amino acid query	>ORF_7_(frame_2) MDRHLKSRIR FWFKQQWPRQ LNNTLRCKQI DA
Length (aa)	32
Full FASTA hits (E-value < 10^{-4})	0
Number of 80mers	1
80mers hits (35% homology, E-value < 10^{-4})	0
Number of 8mers	25
8mers with exact match	0

Table SR6-3. Search results using ORF 49 query sequence against the FARRP database on October 27, 2022

Database	AllergenOnline Database v21 (February 14, 2021)
Amino acid query	>ORF_49_(frame_3) MTKCARDMLT QLRAFMDDQM CQGHADTSGR KRMDSWASDD VQGAEGAEAL QVCRQYLQVW TINDVQNLKR SKVSPLLFLF AGEGYELWTL ADL
Length (aa)	93
Full FASTA hits (E-value < 10^{-4})	0
Number of 80mers	14
80mers hits (35% homology, E-value < 10^{-4})	0
Number of 8mers	86
8mers with exact match	0

Table SR6-4. Search results using ORF 71 query sequence against the FARRP database on October 27, 2022

Database	AllergenOnline Database v21 (February 14, 2021)
Amino acid query	>ORF_71_(frame_3) MPLPTVVPKM DPHPRGASWK KKTFQPRLQS KWIDVNMLEQ LWRIYCGVNK LTLRQLNNTL RTFLMY
Length (aa)	66
Full FASTA hits (E-value < 10^{-4})	0
Number of 80mers	1
80mers hits (35% homology, E-value < 10^{-4})	0
Number of 8mers	59
8mers with exact match	0

Table SR6-5. Search results using ORF 87 query sequence against the FARRP database on October 27, 2022

Database	AllergenOnline Database v21 (February 14, 2021)
Amino acid query	>ORF_87_(frame_1) MCYSDRSSRV VFPAPPNPTI EHHMHSGIIE NKNLKFSTEK CFVIVRRLVH KTENVK
Length (aa)	56
Full FASTA hits (E-value < 10^{-4})	0
Number of 80mers	1
80mers hits (35% homology, E-value < 10^{-4})	0
Number of 8mers	49
8mers with exact match	0

Table SR6-6. Search results using ORF 111 query sequence against the FARRP database on October 27, 2022

Database	AllergenOnline Database v21 (February 14, 2021)
Amino acid query	>ORF_111_(frame_1) MHVMLYSWIR RGREDDSGGS IRITHYYGQF KLKAGANSH
Length (aa)	39
Full FASTA hits (E-value < 10^{-4})	0
Number of 80mers	1
80mers hits (35% homology, E-value < 10^{-4})	0
Number of 8mers	32
8mers with exact match	0

Table SR6-7. Search results using ORF 151 query sequence against the FARRP database on October 27, 2022

Da abase	AllergenOnline Database v21 (February 14, 2021)
Amino acid query	>ORF_151_(frame_2) MWVCVSDDFD VKRITREITE YATNGRFMDL TNLNMLQVNL KEEIRGTTFL LVLDDVWNED PVKWESLLAP LDAGGRGSVV IVTTQSKKVA DVTGTMEPYV LEELTEDDSW SLIESHSFRE ASCSSTNPRM EEIGRKIAKK ISGLPYGATA MGRYLRSKHG ESSWREVLET ETWEMPPAAS DVLSALRRSY DNLPPQLKLC FAFCALFTKG YRFRKDTLIH MWIAQNLIQS TESKRSEDMA EECFDDLVCR FFFRYSWGNY VMNDSVHDLA RWVSLDEYFR ADEDSPLHIS KPIRHLSWCS ERITNVLEDN NTGGDAVNPL SSLRTLLFLG QSEFRSYHLL DRMFRMLSRI RVLDFSNCVI RNLPSSVGNL KHLRYLGLSN TRIQRLPESV TRLCLLQTLL LEGCELCRLP RSMSRLVKLR QLKANPDVIA DIAKVGRLIE LQELKAYNVD KKKGHGIAEL SAMNQLHGDL SIRNLQNVEK TRESRKARLD EKQKLKLLDL RWADGRGAGE CDRDRKVLKG LRPHPNLREL SIKYYGGTSS PSWMTDQYLP NMETIRLRSC ARLTELPCLG QLHILRHLHI DGMSQVRQIN LQFYGTGEVS GFPLLELLNI RRMPSLEEWS EPRRNCCYFP RLHKLLIEDC PRLRNLPSLP PTLEELRISR TGLVDLPGFH GNGDVTTNVS LSSLHVSECR ELRSLSEGLL QHNLVALKTA AFTDCDSLEF LPAEGFRTAI SLESLIMTNC PLPCSFLLPS SLEHLKLQPC LYPNNNEDSL STCFENLTSL SFLDIKDCPN LSSFPPGPLC QLSALQHLSL VNCQRLQSIG FQALTSLESL TIQNCPRLTM SHSLVEVNNS SDTGLAFNIT RWMRRRTGDD GLMLRHRAQN DSFFGGLLQH LTFLQFLKIC QCPQLVTFTG EEEEKWRNLT SLQILHIVDC PNLEVLPANL QSLCSLSTLY IVRCPRIHAF PPGGVSMSLA HLVIHECPQL CQHVPGTFGH P
Length (aa)	991
Full FASTA hits (E-value < 10^{-4})	0
Number of 80mers	912
80mers hits (35% homology, E-value < 10^{-4})	0
Number of 8mers	984
8mers with exact match	0

Table SR6-8. Search results using ORF 156 query sequence against the FARRP database on October 27, 2022

Database	AllergenOnline Database v21 (February 14, 2021)
Amino acid query	>ORF_156_(frame_3) MRFLPEVSAC PWHIWSSMNA LSCVSMSLAH LVIHECPQP
Length (aa)	39
Full FASTA hits (E-value < 10^{-4})	0
Number of 80mers	1
80mers hits (35% homology, E-value < 10^{-4})	0
Number of 8mers	32
8mers with exact match	0

6.3.4 Assessment of the toxicity potential of the seven new ORFs

To assess the potential for toxicity of any proteins expressed from the seven new ORFs, structural similarities shared between their predicted amino acid sequence and sequences in a local Geneious protein toxin BLAST database were evaluated as described in section 6.2.5. BLAST (blastp) searches using the BLOSUM45 similarity scoring matrix and the amino acid sequence of ORF 7 and 87 returned no matches (Figures SR6-9). Although ORFs 49, 71, 111 and 156 returned 9, 8, 9 and 1 match, respectively, none were of biological significance with an E-value acceptance criteria lower than 1 x 10^{-5} (Figures SR6-10 to SR6-13). The analysis of ORF 151 matched 28 protein accessions that contain the keyword “toxin” in the UniProt Knowledgebase, three of which showed an E-value lower than 1 x 10^{-5} (Figure SR6-14). Of these three proteins, two were from *Triticum aestivum* (wheat) and one was from *Hordeum vulgare* (barley). To assess the significance of these findings, an identical search into the same toxin database was made with the amino acid sequence of SvVNT1, another CC-NBS-LRR type resistance gene (*Rpi-vnt1*) product derived from *Solanum venturi* and that provides protection against foliar late blight in potato (*Solanum tuberosum*). As predicted, the blastp search using SvVNT1 as query returned the identical three sequences as the blastp search using MamRGA2 as query (Table SR6-9). SvVNT1 is expressed in the Innate® Hibernate (event Y9) and Innate® Acclimate (event X17) potatoes developed by J.R. Simplot Co. that have been approved for food, feed and cultivation in the USA and Canada since 2017. Both events were also approved for food that same year in Australia and New Zealand and for both food and feed in 2020 in the Philippines.

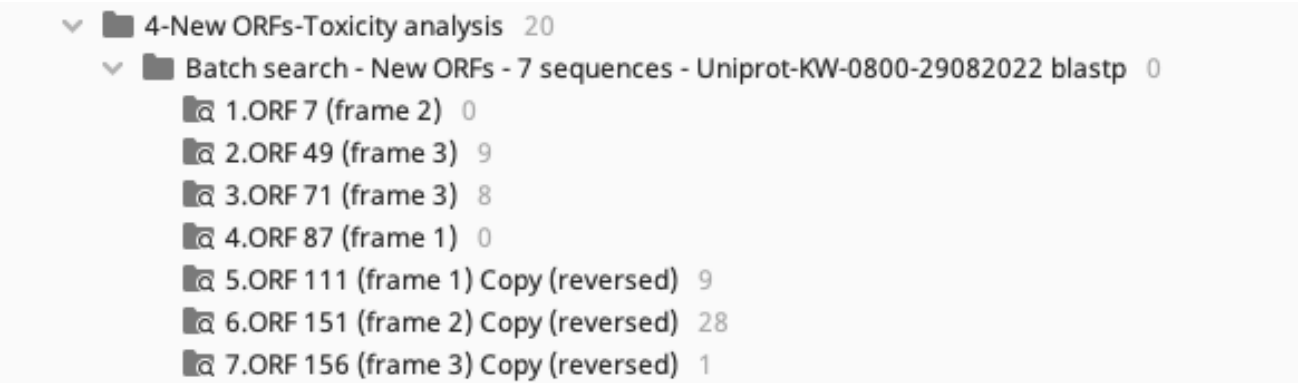


Figure SR6-9. Toxin database Blast search output using the amino acid sequences of the new ORFs as query in Geneious.

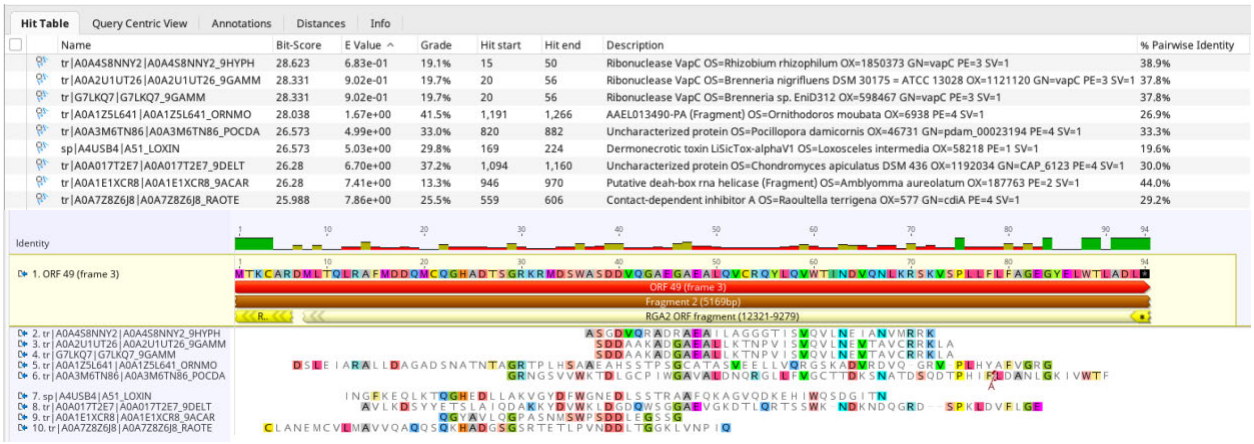


Figure SR6-10. Search results using ORF 49 as query sequence against the Geneious toxin database. Top, hit table and bottom, query-centric alignment.

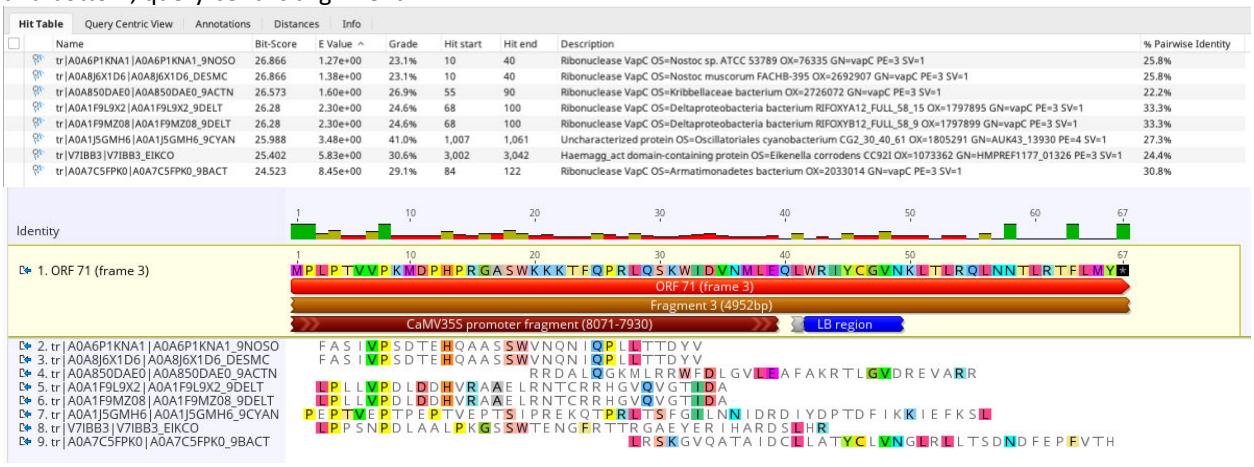


Figure SR6-11. Search results using ORF 71 as query sequence against the Geneious toxin database. Top, hit table and bottom, query-centric alignment.

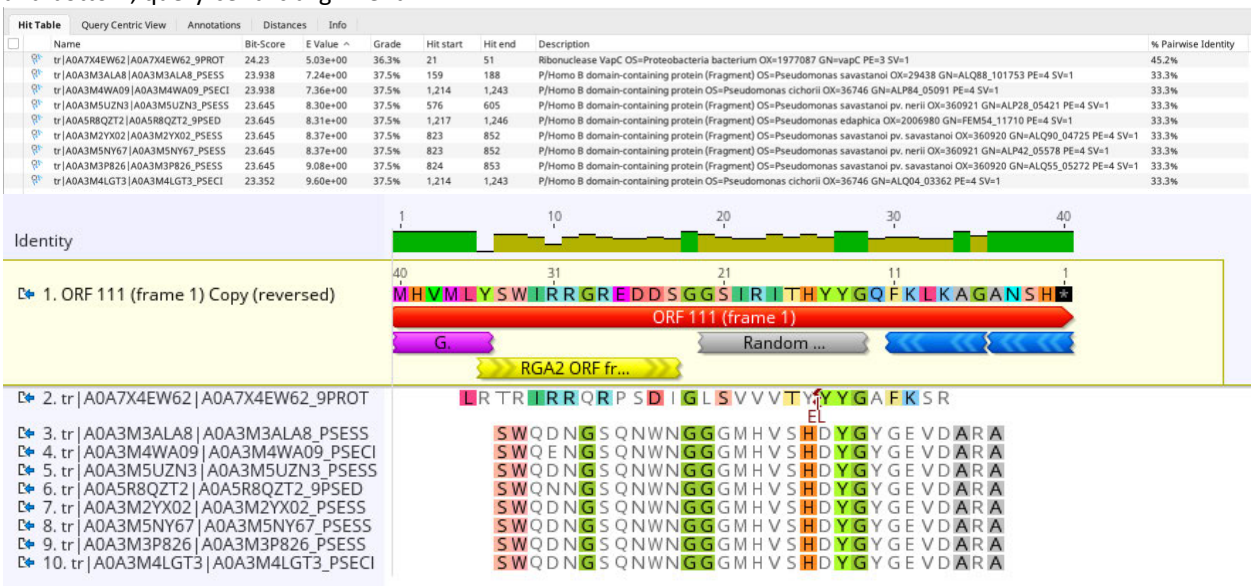


Figure SR6-12. Search results using ORF 111 as query sequence against the Geneious toxin database. Top, hit table and bottom, query-centric alignment.

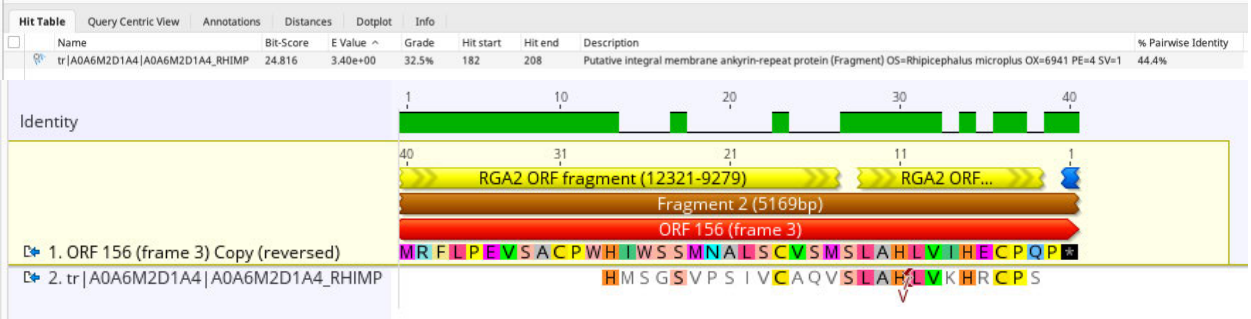


Figure SR6-13. Search results using ORF 156 as query sequence against the Geneious toxin database. Top, hit table and bottom, query-centric alignment.

Figure SR6-14. Search results using ORF 151 as query sequence against the Geneious toxin database. Top, hit table and bottom, query-centric alignment.

Table SR6-9. BLAST results using MamRGA2 and SvVNT1 as query sequence against the toxin database

Event	Query	Query match					
		Start-end	Identity	E-value	Accession	Description	Organism
QCAV-4	MamRGA2	303-695	24.7% in 393 aa overlap	$1.0x10^{-17}$	A0A3B6FZY2	NB-ARC domain-containing protein	*Triticum aestivum* (wheat)
		175-640	25.2% in 466 aa overlap	$4.8x10^{-16}$	A0A287K383	Uncharacterized protein	*Hordeum vulgare* (barley)
		1,239-1,813	24.9% in 575 aa overlap	$4.2x10^{-11}$	A0A3B6SRU1	rRNA N-glycosidase	*Triticum aestivum* (wheat)
Innate®	SvVNT1	175-630	28.7% in 456 aa overlap	$6.0 x10^{-39}$	A0A287K383	Uncharacterized protein	*Hordeum vulgare* (barley)
		306-674	28.5% in 369 aa overlap	$5.5x10^{-30}$	A0A3B6FZY2	NB-ARC domain-containing protein	*Triticum aestivum* (wheat)
		1,267-1,753	23.6% in 487 aa overlap	$1.4x10^{-11}$	A0A3B6SRU1	rRNA N-glycosidase	*Triticum aestivum* (wheat)
		139-260	36.8% in 122 aa overlap	$3.1x10^{-9}$	A0A3B6FZY2	NB-ARC domain-containing protein	*Triticum aestivum* (wheat)
QCAV-4	ORF 151	299-695	24.4% in 397 aa overlap	$7.7x10^{-18}$	A0A3B6FZY2	NB-ARC domain-containing protein	*Triticum aestivum* (wheat)
		250-640	24.7% in 391 aa overlap	$5.1x10^{-13}$	A0A287K383	Uncharacterized protein	*Hordeum vulgare* (barley)
		1,367-1,813	25.5% in 447 aa overlap	$6.7x10^{-7}$	A0A3B6SRU1	rRNA N-glycosidase	*Triticum aestivum* (wheat)

6.4 Conclusions

Using conservative search criteria, the amino acid sequences of all seven newly identified ORFs in QCAV-4 showed no biologically significant similarity to any protein known, or suspected, to be of allergenicity or toxicological concern.

6.5 Supporting documents

- RNA-Seq raw data and mapping sequence files can be provided at FSANZ's request and can be visualised in Geneious.
- SD6-1-AllergenOnline analysis of new ORFs.pdf

6.5.1 Results from the *in silico* functional motifs analysis searches

6.5.1.1 Search for promoter elements and transcription factor binding sites on PlantCARE, a database of plant cis-acting regulatory elements

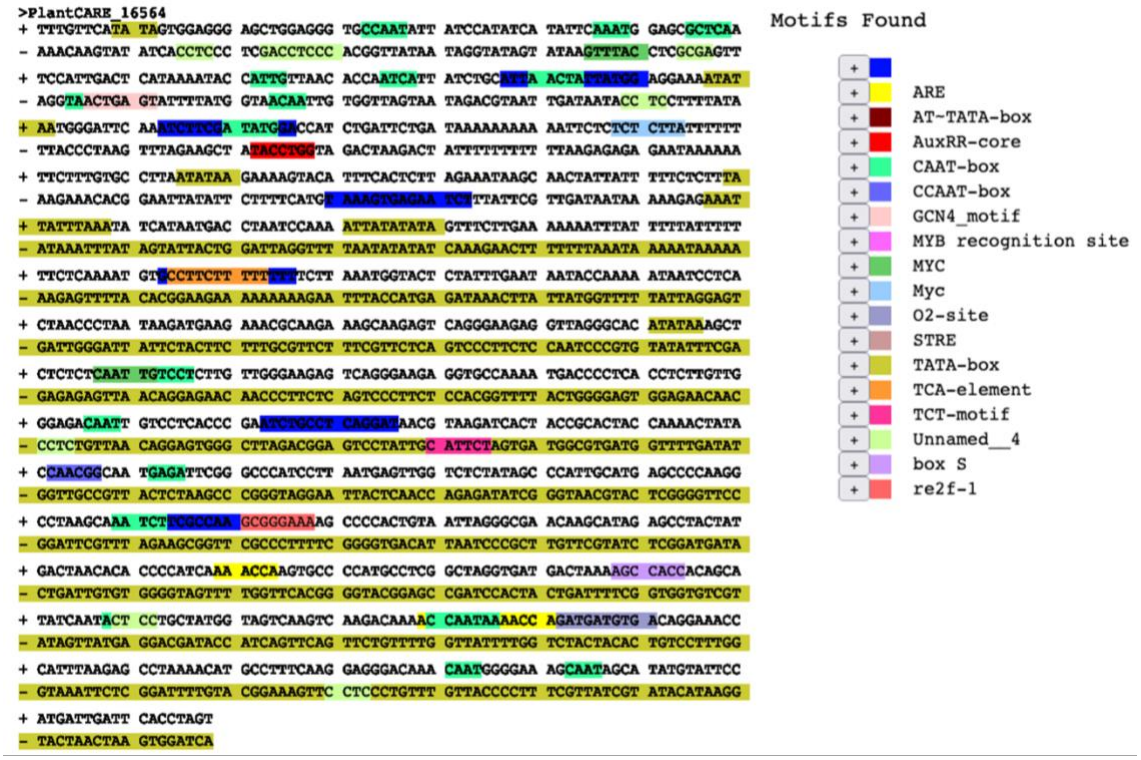


Figure SR6-15. PlantCARE search result for new ORF 7.

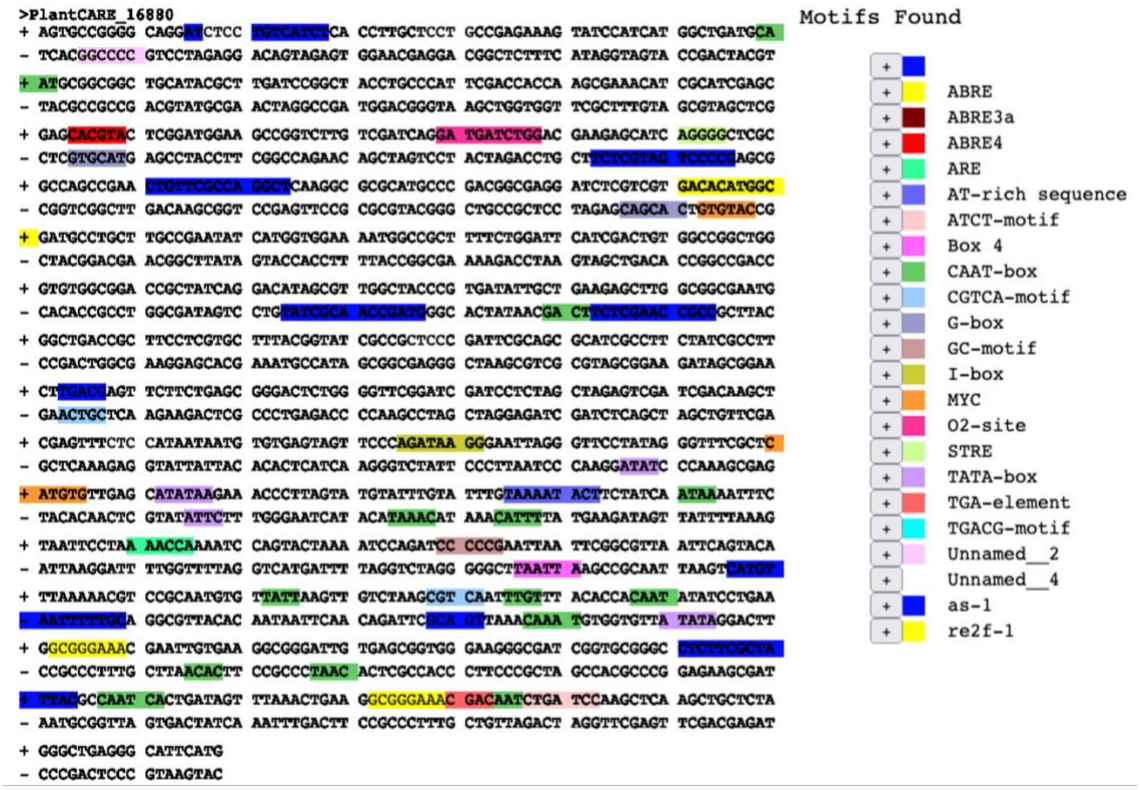


Figure SR6-16. PlantCARE search result for new ORF 49.

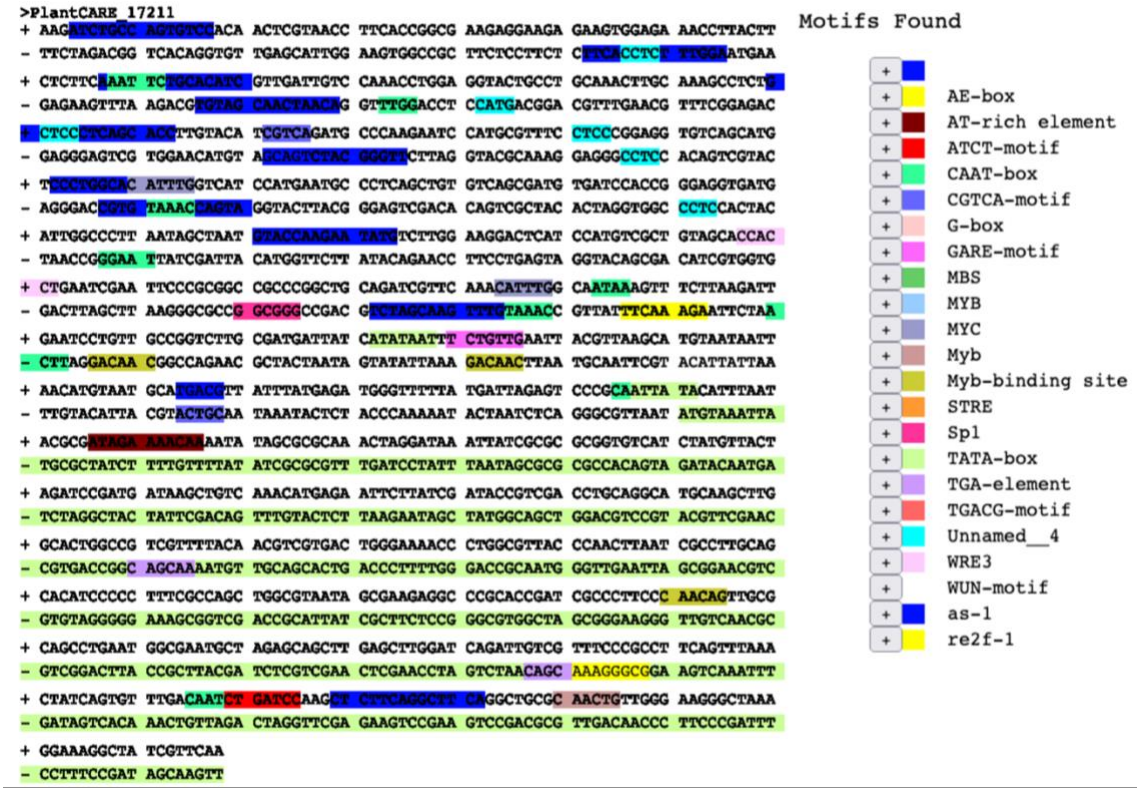


Figure SR6-17. PlantCARE search result for new ORF 71.

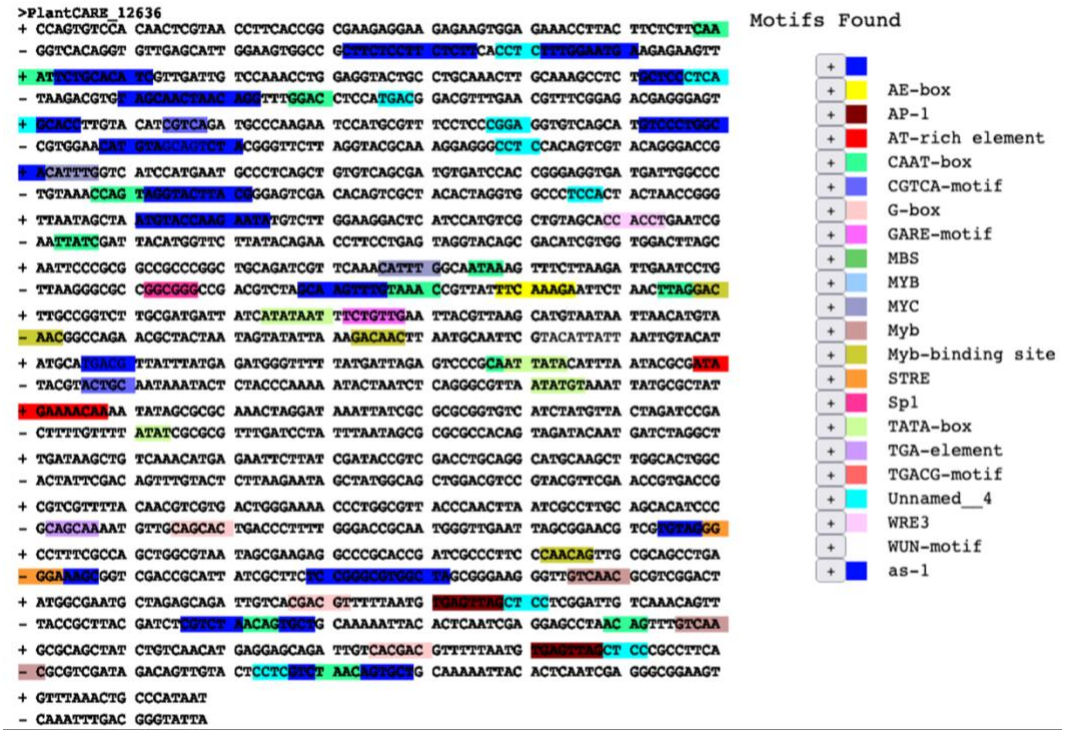


Figure SR6-18. PlantCARE search result for new ORF 87.

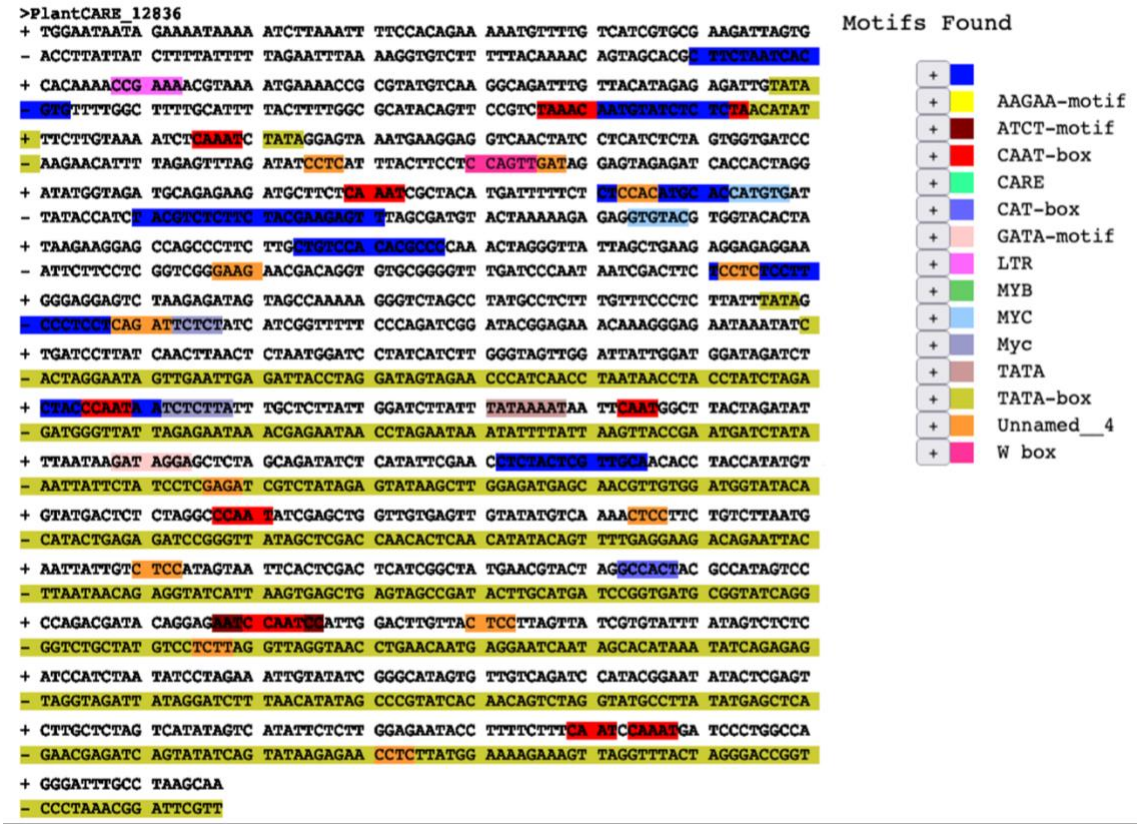


Figure SR6-19. PlantCARE search result for new ORF 111.

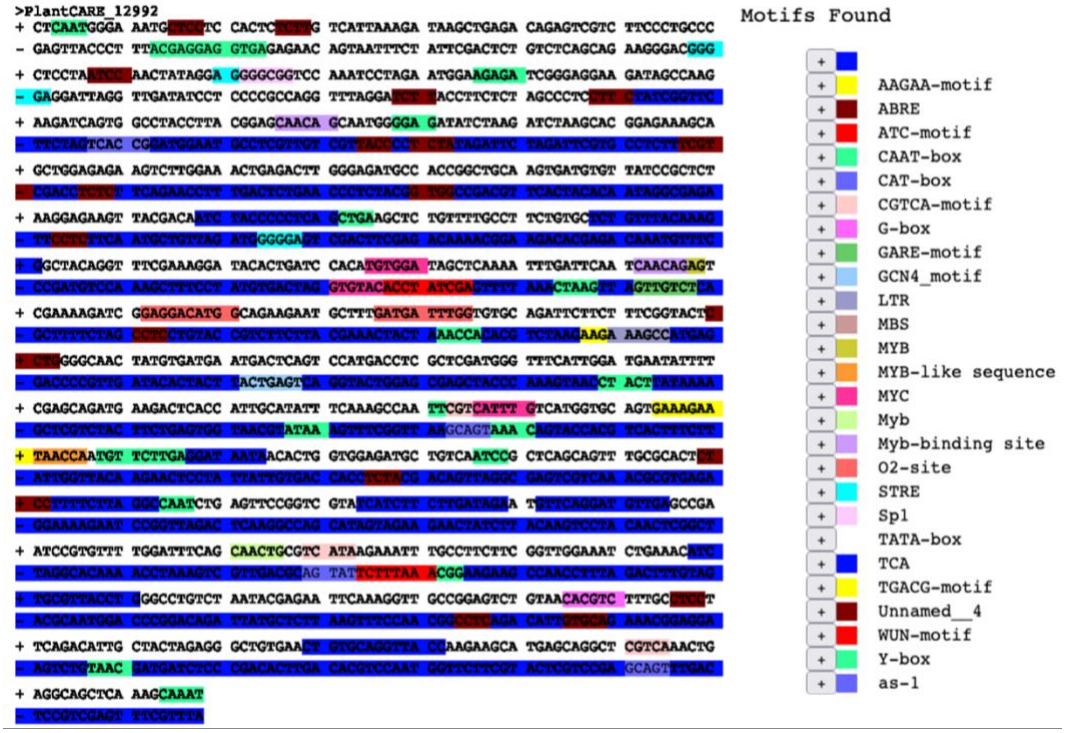


Figure SR6-20. PlantCARE search result for new ORF 151.

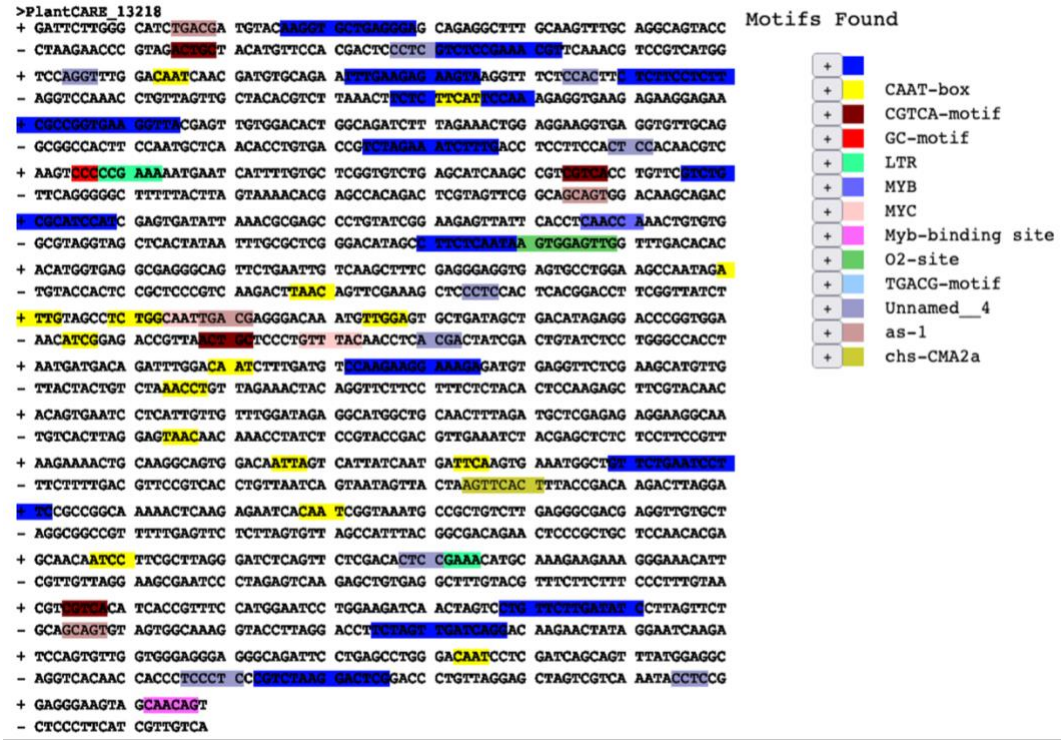


Figure SR6-21. PlantCARE search result for new ORF 156.

6.5.1.2 Search for promoter elements and transcription factor binding sites on TSSP, Softberry

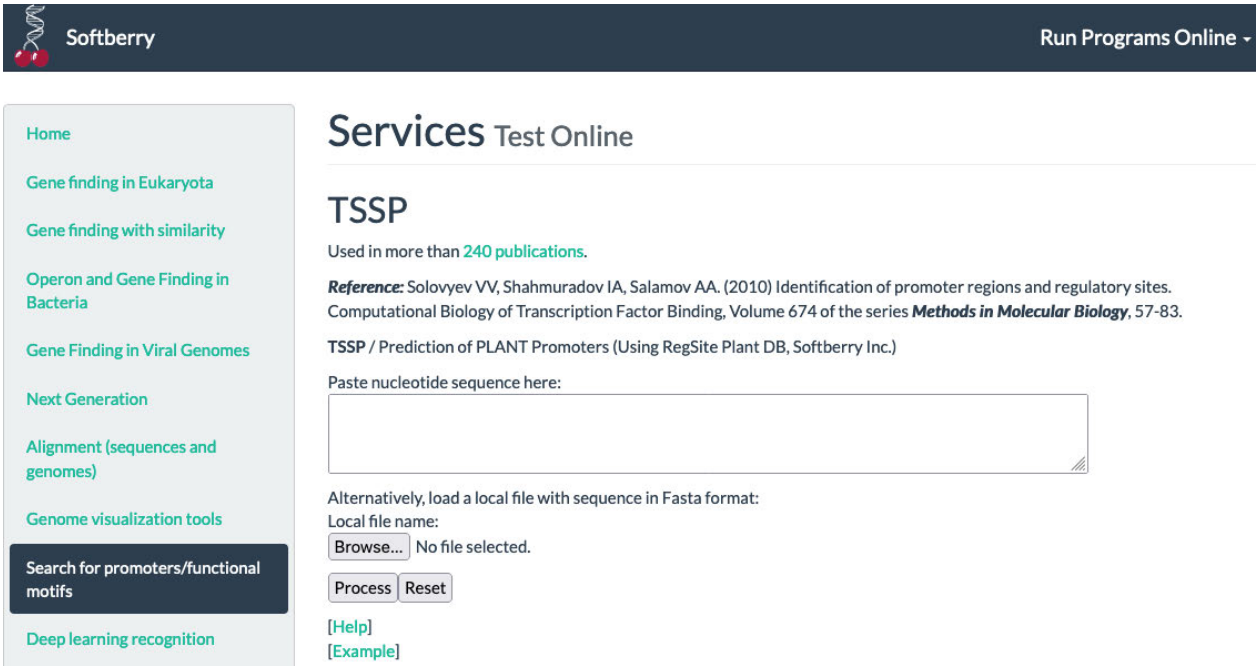


Figure SR6-22. TSSP search page.

```
>ORF_7_(frame_2)-5'-1000bp_extraction
 Length of sequence-      1000
 Thresholds for TATA+ promoters -  0.02, for TATA-/enhancers -  0.04
     1 promoter/enhancer(s) are predicted
 Promoter Pos:    511 LDF-  0.04 TATA box at    479    19.07
 Transcription factor binding sites/RegSite DB:
for promoter at position -     511
   489 (-) RSP00005     CTWWWWWWGT
   265 (+) RSP00024     TATTATTT
   290 (+) RSP00026     gcttttgaTGACtTcaaacac
   475 (-) RSP00087     TtACgcaagcaaTGACatct
   332 (+) RSP00098     cttaatatATTTTTAattattttattctcttaa
   336 (+) RSP00117     aattATTTTTAaa
   337 (+) RSP00125     aatATTTTTAtt
   328 (+) RSP00126     gaaaattttaatATTTTTAtttagtatt
   337 (+) RSP00135     aatATTTTTAtc
   336 (+) RSP00140     aattATTTTTAtt
   233 (+) RSP00161     WAAAG
   485 (+) RSP00161     WAAAG
   373 (-) RSP00161     WAAAG
   281 (-) RSP00161     WAAAG
   281 (+) RSP00337     ATATTTAWW
   292 (-) RSP00337     ATATTTAWW
   340 (+) RSP00339     RTTTTTR
   413 (-) RSP00339     RTTTTTR
   395 (+) RSP00398     TTTGAA
   284 (+) RSP00477     TTTAA
   384 (-) RSP00477     TTTAA
   289 (-) RSP00477     TTTAA
   267 (+) RSP00508     gcaTTTTTatca
   338 (+) RSP00508     gcaTTTTTatca
   344 (+) RSP00508     gcaTTTTTatca
   345 (+) RSP00508     gcaTTTTTatca
   346 (+) RSP00508     gcaTTTTTatca
   367 (+) RSP00508     gcaTTTTTatca
   368 (+) RSP00508     gcaTTTTTatca
   369 (+) RSP00508     gcaTTTTTatca
   370 (+) RSP00508     gcaTTTTTatca
   371 (+) RSP00508     gcaTTTTTatca
   415 (-) RSP00508     gcaTTTTTatca
   339 (-) RSP00508     gcaTTTTTatca
   338 (-) RSP00508     gcaTTTTTatca
   337 (-) RSP00508     gcaTTTTTatca

© 1999 - 2023 www.softberry.com

>ORF_49_(frame_3)-5'-1000bp_extraction
 Length of sequence-      1000
 Thresholds for TATA+ promoters -  0.02, for TATA-/enhancers -  0.04
     0 promoter/enhancer(s) are predicted

© 1999 - 2023 www.softberry.com

>ORF_71_(frame_3)-5'-1000bp_extraction
 Length of sequence-      1000
 Thresholds for TATA+ promoters -  0.02, for TATA-/enhancers -  0.04
     0 promoter/enhancer(s) are predicted

© 1999 - 2023 www.softberry.com

>ORF_87_(frame_1)-5'-1000bp_extraction
 Length of sequence-      1000
 Thresholds for TATA+ promoters -  0.02, for TATA-/enhancers -  0.04
     0 promoter/enhancer(s) are predicted

© 1999 - 2023 www.softberry.com

>ORF_111_(frame_1)-5'-1000bp_extraction
 Length of sequence-      1000
 Thresholds for TATA+ promoters -  0.02, for TATA-/enhancers -  0.04
     0 promoter/enhancer(s) are predicted

© 1999 - 2023 www.softberry.com

>ORF_151_(frame_2)-5'-1000bp_extraction
 Length of sequence-      1000
 Thresholds for TATA+ promoters -  0.02, for TATA-/enhancers -  0.04
     0 promoter/enhancer(s) are predicted

© 1999 - 2023 www.softberry.com

>ORF_156_(frame_3)-5'-1000bp__extraction
 Length of sequence-      1000
 Thresholds for TATA+ promoters -  0.02, for TATA-/enhancers -  0.04
     0 promoter/enhancer(s) are predicted

© 1999 - 2023 www.softberry.com
```

Figure SR6-23. TSSP search results.

6.5.1.3 Search for 3'-end cleavage and polyadenylation regions on POLYAH, Softberry

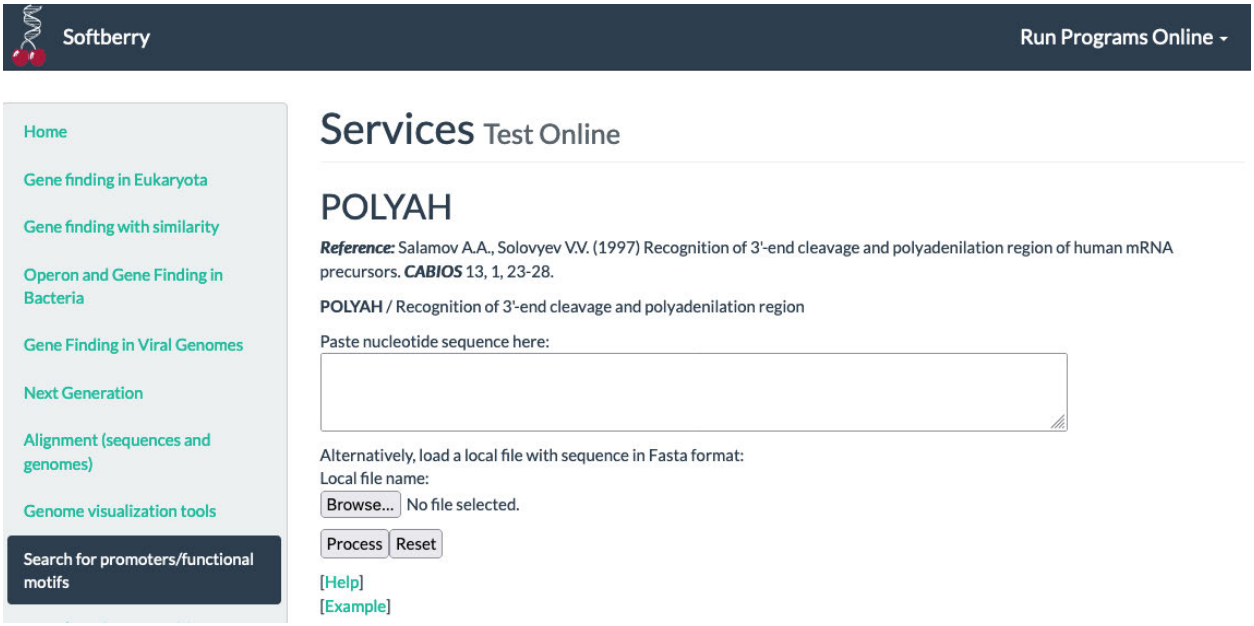


Figure SR6-24. POLYAH search page.

Figure SR6-25. POLYAH search results.

QUT2023-7: *In silico* assessment of MamRGA2 and NPTII for allergenicity and toxicity

STUDY IDENTIFICATION

QUT2023-7

AUTHORS

s 22

PERFORMING LABORATORY

[redacted]

-

[redacted]

[redacted]

[redacted]2023

7.1 Introduction

Banana event QCAV-4 (QUT-QCAV4-6) was created by *Agrobacterium tumefaciens*-mediated transformation of banana (*Musa acuminata* subgroup Cavendish cv Grand Nain) embryogenic cells with plasmid pSAN3. Analysis of the nucleotide sequence of the insert in QCAV-4 revealed a single large 26,849 bp insert on chromosome 6 containing three intact copies of both the CC-NBS-LRR resistance (R) gene *MamRGA2* and the *npt*II selectable marker gene. Details of this analysis can be found in study report QUT2023-1.

Assessing the toxicity and allergenicity potential of newly-expressed proteins produced in the edible portions of a genetically engineered food crop is an integral component of the overall safety assessment. As there is currently no single criterion that is sufficiently predictive of potential toxicity or allergenicity, a "weight-of-evidence" approach is recommended by Codex (2003) for hazard identification that considers a number of criteria including the *in silico* evaluation of the amino acid (aa) sequence similarity to known toxins or allergens (Codex, 2003; Delaney *et al.*, 2008). This document reports the assessment of the amino acid sequences of MamRGA2 and NPTII for similarity to known or suspected allergenic and toxic proteins.

7.2 Materials and Methods

7.2.1 Allergenicity searches

In silico analyses were performed to compare the predicted amino acid sequences of MamRGA2 and NPTII to known allergenic proteins in the Food Allergy Research and Resource Program (FARRP) dataset, which is available through www.AllergenOnline.org (University of Nebraska). Full-length sequence searches were done using the FASTA36 alignment tool and significant homology determined when greater than 50% homology was found between the query sequence and a database entry and the E-value was less than 10^{-4}. Sliding window searches (80-mer) were done using FASTA36 and significant homology determined when there was greater than 35% homology between the query sequence and a database entry and an E-value lower than 1 x 10^{-4}. Exact 8-mer match searches were done using FASTA36 and reported at 100% identity.

7.2.2 Toxicity searches

Potential structural similarities shared between the predicted amino acid sequences of MamRGA2 and NPTII and proteins known or suspected to be toxins were evaluated in Geneious Prime® version 2022.2.1. using the Basic Local Alignment Search Tool (BLAST) and a toxin protein BLAST database created in Geneious. The database was generated from a subset of sequences derived from the UniProt Knowledgebase, comprised of 568,002 manually annotated and reviewed sequences from Swiss-Prot and 226,771,948 automatically annotated, un-reviewed sequences from TrEMBL, that were selected using a keyword search on toxins (KW-0800). At the time of its creation, August 29, 2022, the collection contained a total of 92,851 sequences, including 7,523 reviewed sequences from Swiss-Prot and 85,328 un-reviewed sequences from TrEMBL. Matches between the query sequence and a database entry were considered significant at an E-value lower than 1 x 10^{-5}.

7.3 Results

7.3.1 Assessment of the allergenicity potential of MamRGA2 and NPTII

Protein characterisation (study report QUT2023-4) and RNA-Seq analysis of *MamRGA2* (study report QUT2023-5) both indicated minimal exposure to this protein from the consumption of fruit and peel of

event QCAV-4. Although exposure to MamRGA2 is likely to be negligible, *in silico* analyses were still performed to compare the predicted amino acid sequences of MamRGA2 (accession #ACF21694) to known allergenic proteins in the FARRP dataset. The version of the database used (v21) contains 2,233 protein (amino acid) sequence entries that are categorised into 913 taxonomic-protein groups of unique proven or putative allergens (food, airway, venom/salivary and contact) and celiac protein sequences from 430 species. Three types of analyses were performed for this comparison on October 27, 2022 and as described in materials and methods section 2.1: the full-length sequence (E-value < 10^{-4}), 80-mer sliding window (35% homology with E-value < 10^{-4}) and 8-mer exact match searches did not identify similarities between MamRGA2 and known allergens in the FARRP database (Table SR7-1) indicating that exposure to the MamRGA2 protein is unlikely to be of allergenicity concern.

Similar analyses were done to assess the allergenicity potential of NPTII (accession #AAF65391). These analyses did not identify any known allergens with homology to NPTII (Table SR7-2), indicating that, as previously demonstrated by other studies (OGTR, 2017), the NPTII protein is not allergenic.

Full result outputs from the allergenicity assessment of MamRGA2 and NPTII are presented in SD7-1 and SD7-2 (section 7.5).

Table SR7-1. Search results using MamRGA2 query sequence against the FARRP database on October 27, 2022

Database	AllergenOnline Database v21 (February 14, 2021)
Amino acid query	>MamRGA2 MAGVTSQAAA VFSLVNEIFN RSINLIVAEL RLQLNARAEL NNLQRTLLRT HSLLEEAKAR RMTDKSLVLW LMELKEWAYD ADDILDEYEA AAIRLKVTRS TFKRLIDHVI INVPLAHKVA DIRKRLNGVT LERELNLGAL EGSQPLDSTK RGVTTSLLTE SCIVGRAQDK ENLIRLLLEP SDGAVPVVPI VGLGGAGKTT LSQLIFNDKR VEEHFPLRMW VCVSDDFDVK RITREITEYA TNGRFMDLTN LNMLQVNLKE EIRGTTFLLV LDDVWNEDPV KWESLLAPLD AGGRGSVVIV TTQSKKVADV TGTMEPYVLE ELTEDDSWSL IESHSFREAS CSSTNPRMEE IGRKIAKKIS GLPYGATAMG RYLRSKHGES SWREVLETET WEMPPAASDV LSALRRSYDN LPPQLKLCFA FCALFTKGYR FRKDTLIHMW IAQNLIQSTE SKRSEDMAEE CFDDLVCRFF FRYSWGNYVM NDSVHDLARW VSLDEYFRAD EDSPLHISKP IRHLSWCSER ITNVLEDNNT GGDAVNPLSS LRTLLFLGQS EFRSYHLLDR MFRMLSRIRV LDFSNCVIRN LPSSVGNLKH LRYLGLSNTR IQRLPESVTR LCLLQTLLLE GCELCRLPRS MSRLVKLRQL KANPDVIADI AKVGRLIELQ ELKAYNVDKK KGHGIAELSA MNQLHGDLSI RNLQNVEKTR ESRKARLDEK QKLKLLDLRW ADGRGAGECD RDRKVLKGLR PHPNLRELSI KYYGGTSSPS WMTDQYLPNM ETIRLRSCAR LTELPCLGQL HILRHLHIDG MSQVRQINLQ FYGTGEVSGF PLLELLNIRR MPSLEEWSEP RRNCCYFPRL HKLLIEDCPR LRNLPSLPPT LEELRISRTG LVDLPGFHGN GDVTTNVSLS SLHVSECREL RSLSEGLLQH NLVALKTAAF TDCDSLEFLP AEGFRTAISL ESLIMTNCPL PCSFLLPSSL EHLKLQPCLY PNNNEDSLST CFENLTSLSF LDIKDCPNLS SFPPGPLCQL SALQHLSLVN CQRLQSIGFQ ALTSLESLTI QNCPRLTMSH SLVEVNNSSD TGLAFNITRW MRRRTGDDGL MLRHRAQNDS FFGGLLQHLT FLQFLKICQC PQLVTFTGEE EEKWRNLTSL QILHIVDCPN LEVLPANLQS LCSLSTLYIV RCPRIHAFPP GGVSMSLAHL VIHECPQLCQ RCDPPGGDDW PLIANVPRIC LGRTHPCRCS TT
Length (aa)	1232
Full FASTA hits (E-value < 10^{-4})	0
Number of 80mers	1153
80mers hits (35% homology, E-value < 10^{-4})	0
Number of 8mers	1225
8mers with exact match	0

Table SR7-2. Search results using NPTII query sequence against the FARRP database on October 27, 2022

Database	AllergenOnline Database v21 (February 14, 2021)
Amino acid query	>NPTII MGIEQDGLHA GSPAAWVERL FGYDWAQQTI GCSDAAVFRL SAQGRPVLFV KTDLSGALNE LQDEAARLSW LATTGVPCAA VLDVVTEAGR DWLLLGEVPG QDLLSSHLAP AEKVSIMADA MRRLHTLDPA TCPFDHQAKH RIERARTRME AGLVDQDDLD EEHQGLAPAE LFARLKARMP DGEDLVVTHG DACLPNIMVE NGRFSGFIDC GRLGVADRYQ DIALATRDIA EELGGEWADR FLVLYGIAAP DSQRIAFYRL LDEFF
Length (aa)	265
Full FASTA hits (E-value < 10^{-4})	0
Number of 80mers	186
80mers hits (35% homology, E-value < 10^{-4})	0
Number of 8mers	258
8mers with exact match	0

7.3.2 Assessment of the toxicity potential of MamRGA2 and NPTII

To assess the potential for toxicity of MamRGA2 and NPTII, structural similarities shared between their predicted amino acid sequence and sequences in a local Geneious protein toxin BLAST database were evaluated as described in materials and methods section 7.2.2. BLAST (blastp) searches using the BLOSUM45 similarity scoring matrix and the amino acid sequence of NPTII as queries did not return any accessions of biological significance with an E-value acceptance criteria lower than 1 x 10^{-5} (Figure SR7-1).

The amino acid sequence of MamRGA2 matched three protein accessions that contain the keyword "toxin" in the UniProt Knowledgebase with an E-value lower than 1 x 10^{-5} (Figure SR7-2). Of these three proteins, two were from *Triticum aestivum* (wheat) and one was from *Hordeum vulgare* (barley). To assess the significance of these findings, an identical search into the same toxin database was made with the amino acid sequence of SvVNT1, another CC-NBS-LRR type resistance gene (*Rpi-vnt1*) product derived from *Solanum venturi* and that provides protection against foliar late blight in potato (*Solanum tuberosum*). As predicted, the blastp search using SvVNT1 as query returned the identical three sequences as the blastp search using MamRGA2 as query (Table SR7-3). SvVNT1 is expressed in the Innate® Hibernate (event Y9) and Innate® Acclimate (event X17) potatoes developed by J.R. Simplot Co. that have been approved for food, feed and cultivation in the USA and Canada since 2017. Both events were also approved for food that same year in Australia and New Zealand and for both food and feed in 2020 in the Philippines.

7.4 Conclusions

Using conservative search criteria, the amino acid sequences of MamRGA2 and NPTII showed no biological significant similarity to any protein known, or suspected, to be of allergenicity or toxicological concern.

Figure SR7-1. Search results using NPTII as query sequence against the Geneious toxin database. Top, hit table and bottom, query-centric alignment.

Figure SR7-2. Search results using MamRGA2 as query sequence against the Geneious toxin database. Top, hit table and bottom, query-centric alignment.

Table SR7-3. BLAST results using MamRGA2 and SvVNT1 as query sequence against the toxin database

Event	Query	Query match					
		Start-end	Identity	E-value	Accession	Description	Organism
QCAV-4	MamRGA2	303-695	24.7% in 393 aa overlap	$1.0x10^{-17}$	A0A3B6FZY2	NB-ARC domain-containing protein	*Triticum aestivum* (wheat)
		175-640	25.2% in 466 aa overlap	$4.8x10^{-16}$	A0A287K383	Uncharacterized protein	*Hordeum vulgare* (barley)
		1,239-1,813	24.9% in 575 aa overlap	$4.2x10^{-11}$	A0A3B6SRU1	rRNA N-glycosidase	*Triticum aestivum* (wheat)
Innate®	SvVNT1	175-630	28.7% in 456 aa overlap	$6.0 x10^{-39}$	A0A287K383	Uncharacterized protein	*Hordeum vulgare* (barley)
		306-674	28.5% in 369 aa overlap	$5.5x10^{-30}$	A0A3B6FZY2	NB-ARC domain-containing protein	*Triticum aestivum* (wheat)
		1,267-1,753	23.6% in 487 aa overlap	$1.4x10^{-11}$	A0A3B6SRU1	rRNA N-glycosidase	*Triticum aestivum* (wheat)
		139-260	36.8% in 122 aa overlap	$3.1x10^{-9}$	A0A3B6FZY2	NB-ARC domain-containing protein	*Triticum aestivum* (wheat)

7.5 Supporting documents

- SD7-1-AllergenOnline analysis of MamRGA2.pdf
- SD7-2-AllergenOnline analysis of NPTII.pdf

QUT2023-8: Characterisation of the MamRGA2 protein derived from an *E. coli* expression system and its equivalence to MamRGA2 expressed in event QCAV-4

STUDY IDENTIFICATION

QUT2023-8

AUTHORS

[redacted]
s 22
[redacted]

PERFORMING LABORATORY

[redacted]
-
[redacted]

[redacted] 2023

8.1 Introduction

Banana event QCAV-4 (QUT-QCAV4-6) was created by *Agrobacterium tumefaciens*-mediated transformation of banana (*Musa acuminata* subgroup Cavendish cv Grand Nain) embryogenic cells with plasmid pSAN3. This plasmid contains the *neomycin phosphotransferase* II (*npt*II) selectable marker gene from the Tn5 transposon of *Escherichia coli* strain K12 and the *MamRGA2* disease resistance (R) gene from the wild banana *Musa acuminata* ssp. *malaccensis* which confers resistance to the devastating banana fungal pathogen *Fusarium oxysporum* f. sp. *cubense* tropical race 4 (TR4) (Dale *et al*., 2017).

The purpose of this study was to fully characterise the microbial-expressed form of the MamRGA2 protein that was used as standard for the quantification of MamRGA2 *in planta* (study report QUT2023-4) and in the gastric enzymatic digestion and thermal stability assessment of MamRGA2 presented in study report QUT2023-9. MamRGA2 was expressed in an *E. coli* over-expression system and characterised by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), Western blot analysis and mass spectral analysis of tryptic peptide fragments.

8.2 Materials and Methods

8.2.1 Cloning and sequencing of the MamRGA2 expression vector

The 3,699 bp cDNA sequence of *MamRGA2* (accession #EU616673) was cloned into the multiple cloning site (MCS) of the pET28a(+) expression vector (Cat. #69864, Novagen, Merck) allowing fusion of a N-terminal His_6-tag (Figure SR8-1). The integrity of the resulting pET28a(+)-MamRGA2 construct was confirmed by cloning and sequencing using routine laboratory practices (Sambrook and Russell, 2001) and transfected into expression host *E. coli* strain Rosetta 2 (DE3) (Cat. #71397, Novagen, Merck).

8.2.2 Expression and purification of the MamRGA2 protein from a recombinant *E. coli*

For protein expression, 100 mL of recombinant *E. coli* was cultured in Luria-Bertani broth containing kanamycin (30 µg/mL) and chloramphenicol (34 µg/mL) at 37˚C with agitation (200 rpm) to an optical density (OD_{600nm} = 0.6) at which point isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce MamRGA2 expression. The culture was maintained at 37˚C with agitation (200 rpm) for 3 h, following which the cells were harvested by centrifugation (4,000 x*g*, 10 min, 4˚C). Cells were resuspended in BugBuster Master Mix (Cat. #71456, Merck) including protease inhibitors cOmplete™, Mini Protease Inhibitor Cocktail (Cat. #04693124001, Roche, Merck) and incubated at room temperature for 20 min. The homogenate was centrifuged at 16,000 x*g* for 20 min at 4˚C. At this stage the supernatant, which contains soluble proteins, is designated the soluble protein fraction (SPF). The pellet was washed three times with 1:10 dilution of BugBuster Master Mix by centrifugation (5,000 x*g*, 15 min, 4˚C). The washed pellet was then solubilised by resuspension with 1x denaturing equilibration buffer (50 mM sodium phosphate, 6 M guanidine-HCL, 300 mM NaCl) and centrifuged again (14,000 x*g*, 10 min) to clarify and remove any insoluble debris. At this stage the supernatant, which contains insoluble proteins, is designated the insoluble protein fraction (IPF).

8.2.3 Production of anti-MamRGA2 monoclonal antibodies

Since antibodies against MamRGA2 are not commercially available, a mouse anti-MamRGA2 monoclonal antibody designated 17F07 was custom made by Maine Biotechnology Services (MBS, BBI Solutions). The 17F07 antibody was produced against 3 peptides of MamRGA2 (QUT-P3 Ac-CSERITNVLEDNNT-amide,

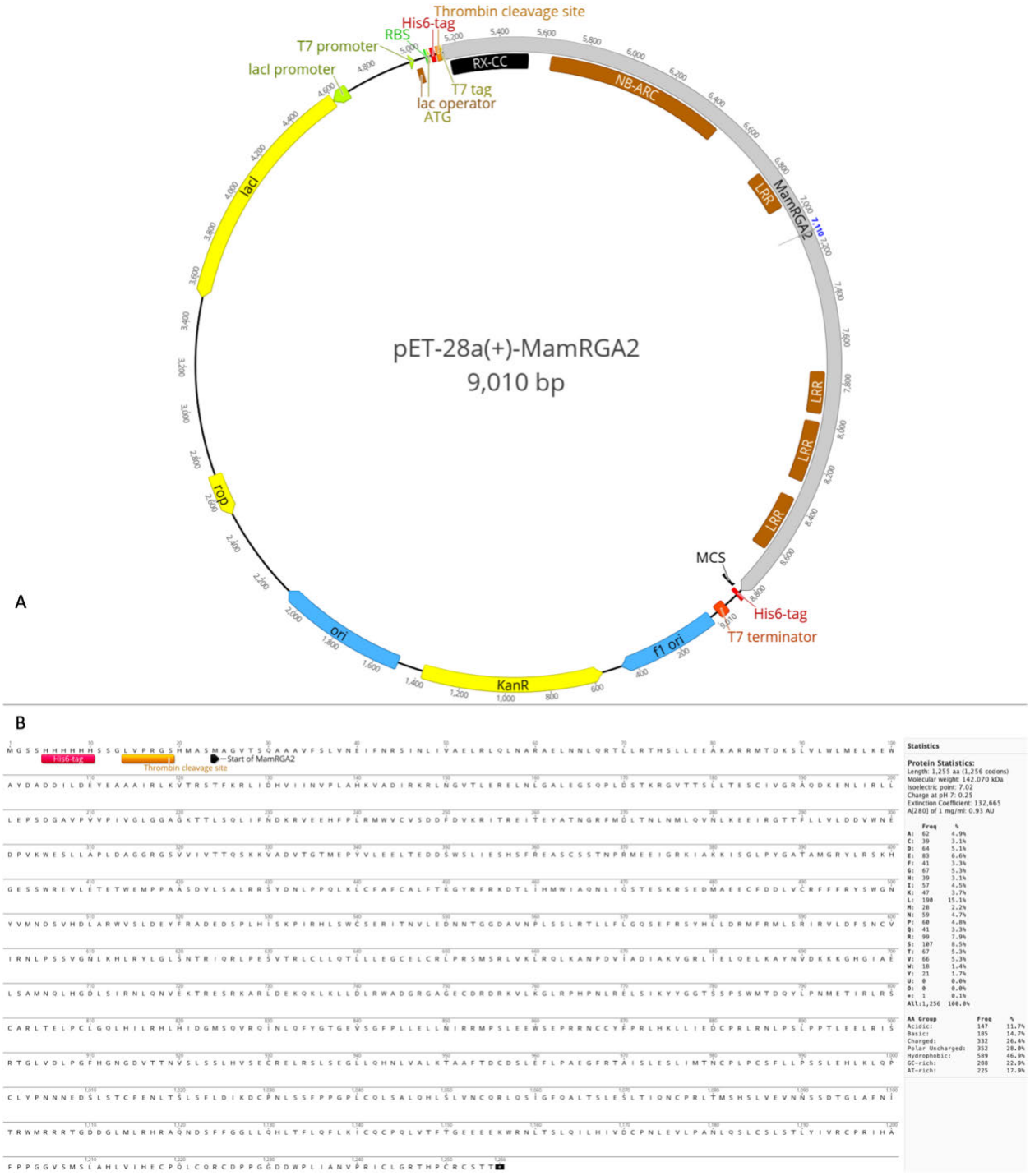


Figure SR8-1. Expression of the MamRGA2 protein using the pET-28a(+) expression system. A: schematic representation of the pET-28a(+)-MamRGA2 vector for expression of MamRGA2 in *E. coli* strain Rosetta 2 (DE3). B: Deduced amino acid (aa) sequence of the MamRGA2 protein. The N-terminal His_6-tag and thrombin cleavage site are indicated within the 23-amino acid leader sequence. The expressed MamRGA2 protein is 1,255 amino acids (aa) in length with a calculated molecular weight of ~142 kDa.

QUT-P4 Ac-CSEFRSYHLLDRMFR-amide and QUT-P5 Ac-CDPPGGDDWPLIAN-amide) and supplied at a concentration of 0.7 mg/mL. Details on the production of 17F07 can be found in SD8-1 (section 8.5).

8.2.4 SDS-PAGE and Western blot analysis

For SDS-PAGE, 4-20% Mini-PROTEAN TGX Precast Protein gels (Cat. #4561095, Bio-Rad) were used followed by staining with Coomassie brilliant blue R-250 (Cat. #0472-25G, Amresco). Gels were de-stained in a solution of 10% acetic acid and 50% methanol for 2-3 h followed by deionised water for 1-2 h and imaged on a Chemidoc MP imaging system (Bio-Rad). An estimate of the molecular weight was determined relative to the Colour Protein Standard markers, Broad Range (11-245 kDa or 10-250 kDa) (Cat. #P7712S, P7719S, New England Biolabs) included on all gels.

For Western immunoblot analysis, proteins in gels were transferred to a nitrocellulose membrane (Cat. #1704158, Bio-Rad), blocked in 3% bovine serum albumin (BSA, Cat. #P0834, Sigma)/Tris-buffered saline with Tween® 20 (TBST) for 1 h prior to incubation. For His_6-tag detection, an anti-His_6-tag monoclonal antibody-HRP (Cat. #MA1-135-HRP; ThermoFisher) was used at a 1:3,000 dilution in 3% BSA/TBST for 1 h. For specific MamRGA2 detection, custom-made mouse anti-MamRGA2 monoclonal antibody 17F07 (MBS, BBI Solutions) was used at 1:1,000 dilution in 3% BSA/TBST for 1 h followed by incubation with secondary goat-anti-mouse IgG-HRP (Cat. #62-6520, ThermoFisher) at 1:5,000 dilution in 3% BSA/TBST for 1 h. Chemiluminescence was visualised using the Clarity™ Western ECL Substrate (Cat. #1705061, Bio-Rad) and imaged on the Chemidoc MP imaging system (Bio-Rad) at various durations of exposure.

8.2.5 MamRGA2 protein purification by metal affinity chromatography

A mini-scale batch purification protocol was used to purify the *E. coli*-expressed MamRGA2 from the insoluble fraction of *E. coli* extracts using TALON® metal affinity resin (Cat. # 635501, Clontech) following recommendations from the manufacturer. In brief, under denaturing conditions, insoluble clarified protein fraction from 1 mL of IPTG-induced *E. coli* culture was added to the resin, allowed to bind for 10 min and centrifuged at 18,000 x*g* for 1 min to pellet the protein-resin complex. The supernatant containing unbound proteins was removed, and a sample taken for SDS-PAGE analysis. The resin was further washed twice with denaturing equilibration buffer (50 mM sodium phosphate, 6 M guanidine-HCl and 300 mM NaCl, 10% glycerol, pH 7) with samples collected from each wash for SDS-PAGE analysis (washes 1 and 2). The protein was then eluted twice from the resin using elution buffer (45 mM sodium phosphate, 5.4 M guanidine-HCL, 270 mM NaCl and 150 mM imidazole, pH 7) and samples collected from each eluate for SDS-PAGE analysis (elution 1 and 2).

8.2.6 Estimation of MamRGA2 protein concentration by densitometry

For densitometric analysis, BSA standards and semi-purified *E. coli*-expressed MamRGA2 IPF were prepared in 2x Laemmli sample buffer. A standard curve consisting of a serial doubling dilution of BSA (5 - 0.3 µg) and triplicate samples of MamRGA2 IPF (5 µL) were subjected to SDS-PAGE and stained with Coomassie. Gels were imaged on a Chemidoc MP imaging system (Bio-Rad) and densitometric analysis performed using the Image Lab Software, v6.1 (Bio-Rad). A linear regression was made from the intensity of the BSA bands and used to extrapolate the concentration of MamRGA2 within the semi-purified MamRGA2 extracts. A known amount of BSA (3 µg) and β-lactoglobulin (3 µg, Cat. #L3908, Sigma) were included as standards. The resulting BSA standards equation was y = 1.95E-08x - 0.949 (Figure SR8-2) from which the concentration of MamRGA2 was calculated at 0.4 µg/µL.

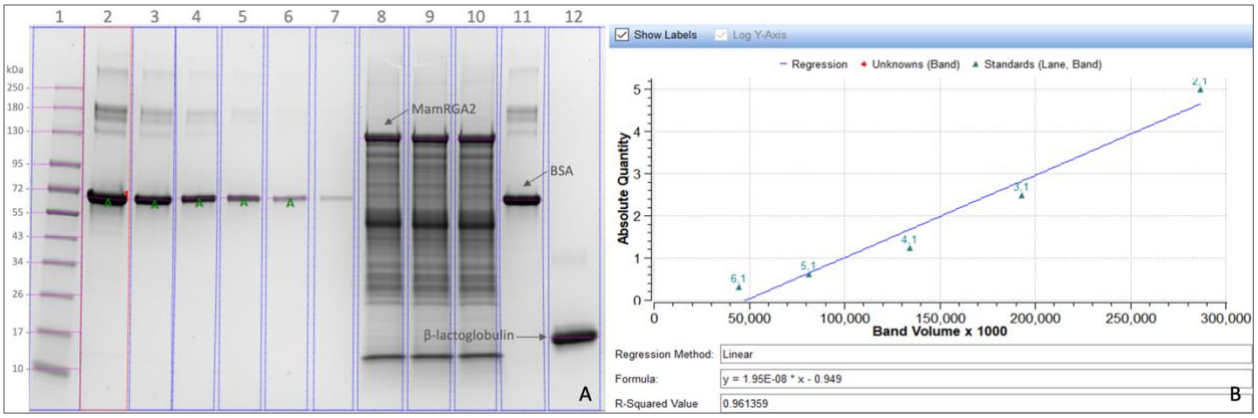


Figure SR8-2. MamRGA2 protein concentration estimation by densitometry. A: 4-20% Mini-PROTEAN TGX Precast Protein Gel, Coomassie brilliant blue R-250 stain (Cat. #0472-25G, Amresco). Lane 1, Colour Pre-stained Protein Standard, Broad Range (10-250 kDa) (Cat. #P7719S, New England Biolabs); lanes 2-7, BSA standard curve 5, 2.5, 1.25, 0.625, 0.313 µg and 0.156 µg, respectively; lanes 8-10, MamRGA2 IPF (5 µL) of unknown concentration; lane 11, BSA (3 µg) and lane 12, β-lactoglobulin (3 µg). B: Linear regression on Image Lab Software, v6.1 (Bio-Rad).

8.3 Results

8.3.1 MamRGA2 expression, purification and detection

The MamRGA2 protein, containing an N-terminal His_6-tag sequence, was expressed in *E. coli* Rosetta 2 (DE3) from its pET28a(+) expression vector by induction with IPTG. Production of the His_6-tag-MamRGA2 protein fusion could be seen 3 h post-IPTG induction using SDS-PAGE and Coomassie staining (Figure SR8-3A, lane 3). The expressed MamRGA2 protein was found exclusively in the IPF of *E. coli* extracts (Figure SR8-3A, lanes 5 and 6 or SR8-3C, lanes 6 and 7) and was detectable by Western blot analysis using an anti-His_6-tag antibody (Figure SR8-3B, lanes 3, 4 and 6). Attempts to purify MamRGA2 from the IPF using chromatography on TALON® metal affinity resin were largely unsuccessful (Figure SR8-3C, lanes 11 and 12). The use of Western blot analysis and the 17F07 mouse anti-MamRGA2 monoclonal antibody allowed the specific detection of the ~142 kDa *E. coli*-expressed MamRGA2 protein but only from the IPF of *E. coli* extracts (Figure SR8-4).

8.3.2 Characterisation of the *E. coli*-expressed MamRGA2 protein from tryptic and chymotryptic peptides

Data Dependent Acquisition (DDA) LC-MS/MS analysis of trypsin-digested MamRGA2 protein identified 81 different tryptic peptides derived from the expected amino acid sequence of MamRGA2. This corresponds to 53% coverage of the supplied amino acid sequence of MamRGA2 (Figure SR8-5). To increase the coverage, (DDA) LC-MS/MS analysis of chymotrypsin-digested MamRGA2 protein identified 129 different chymotryptic peptides derived from the expected amino acid sequence of MamRGA2. This corresponds to 67% coverage of the supplied amino acid sequence of MamRGA2 (Figure SR8-6). When the result from both trypsin- and chymotrypsin-digested MamRGA2 are combined, a total sequence coverage of 82% was achieved (Figure SR8-7) which unambiguously identified the *E. coli*-expressed MamRGA2 as corresponding to the *Musa acuminata* ssp. *malaccensis* resistance gene analogue 2 protein (MamRGA2; accession #ACF21694) expressed in event QCAV-4.

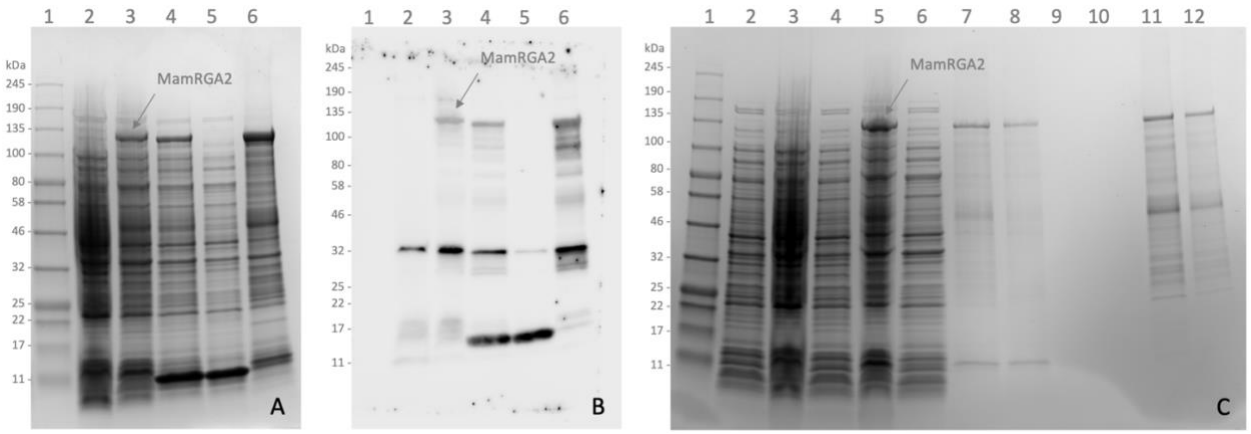


Figure SR8-3. MamRGA2 expression in *E. coli*. A and C: 4-20% Mini-PROTEAN TGX Precast Protein Gel, Coomassie brilliant blue R-250 stain (Cat. #0472-25G, Amresco) and B: Western immunoblot. Detection with anti-His_6-tag monoclonal antibody-HRP (Cat. #MA1-135-HRP, ThermoFisher), 1:3,000, chemiluminescent substrate development and auto optimal exposure time of 20 min. Lanes 1 on all panels, Colour Pre-stained Protein Standard, Broad Range (10-245 kDa) (Cat. #P7712S, New England Biolabs). A and B: 10 µL each for lane 2, *E. coli* total protein (TP) from uninduced culture; lane 3, TP 3 h post-IPTG induction; lane 4, TP 3 h post-IPTG induction followed by lysozyme/sonication extraction; lane 5 and 6, SPF and IPF 3 h post-IPTG induction. C: 10 µL each for lanes 2 and 3, TP from uninduced culture at 0 and 3 h; lanes 4 and 5, TP at 0 and 3 h post-IPTG induction; lanes 6 and 7, SPF and IPF 3 h post-IPTG induction; lanes 8-12, TALON® metal affinity purification with lane 8, unbound protein fraction, lane 9 and 10, washes 1 and 2 and lanes 11 and 12, elutions 1 and 2. SPF = soluble protein fraction and IPF = insoluble protein fraction.

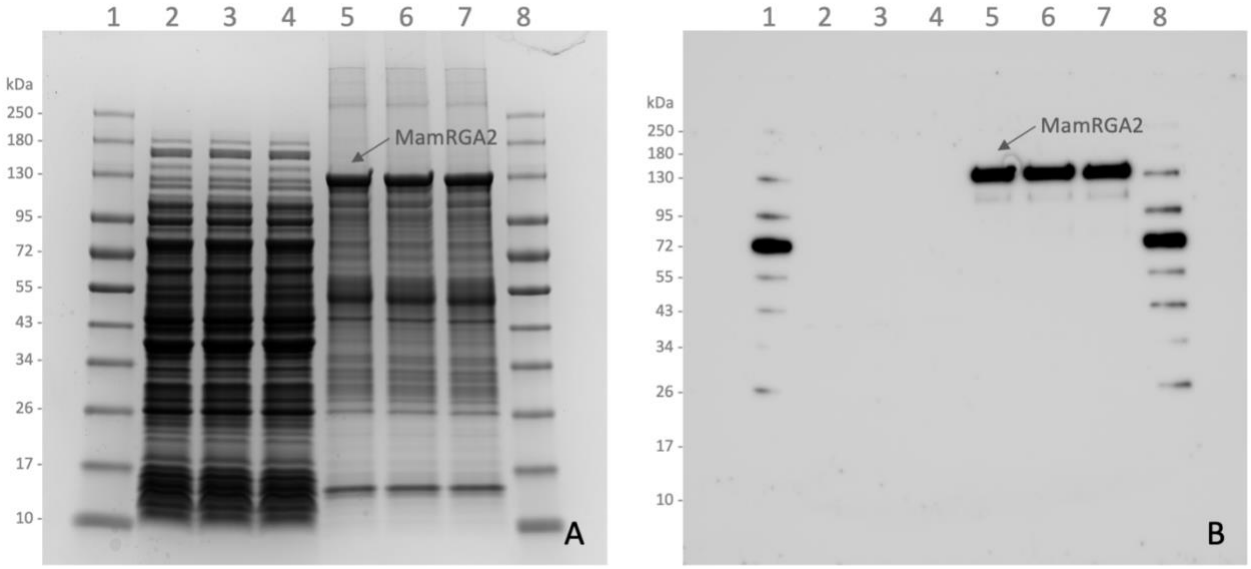


Figure SR8-4. Detection of *E. coli*-expressed MamRGA2 using an anti-MamRGA2 antibody. A: 4-20% Mini-PROTEAN TGX Precast Protein Gel, Coomassie brilliant blue R-250 stain (Cat. #0472-25G, Amresco) and B: Western immunoblot. Detection with primary mouse anti-MamRGA2 monoclonal antibody 17F07 (MBS, BBI Solutions), 1:1,000 followed by secondary goat-anti-mouse IgG-HRP (Cat. #62-6520, ThermoFisher), 1:5,000, chemiluminescent substrate development and auto optimal exposure time of 7 min. Lanes 1 and 8, Colour Pre-stained Protein Standard, Broad Range (10-250 kDa) (Cat. #P7719S, New England Biolabs); lanes 2-4, SPF of *E. coli* extract 3 h post-IPTG induction run in triplicate (16 µg SPF for A and 123 ng for B); lanes 5-7, IPF of *E. coli* extracts 3 h post-IPTG induction run in triplicate [11.5 µg IPF (1.3 µg of which is MamRGA2, ~11%) for A and 88 ng for B equivalent to 10 ng MamRGA2].

MGSSHHHHHH SSGLVPRGSH MASMAGVTSQ AAAVFSLVNE IFNRSINLIV AELRLQLNAR AELNNLQRTL LRTHSLLEEA KARRMTDKSL VLWLMELKEW
AYDADDILDE YEAAAIRLKV TRSTFKRLID HVIINVPLAH KVADIRKRLN GVTLERELNL GALEGSQPLD STKRGVTTSL LTESCIVGRA QDKENLIRLL
LEPSDGAVPV VPIVGLGGAG KTTLSQLIFN DKRVEEHFPL RMWVCVSDDF DVKRITREIT EYATNGRFMD LTNLNMLQVN LKEEIRGTTF LLVLDDVWNE
DPVKWESLLA PLDAGGRGSV VIVTTQSKKV ADVTGTMEPY VLEELTEDDS WSLIESHSFR EASCSSTNPR MEEIGRKIAK KISGLPYGAT AMGRYLRSKH
GESSWREVLE TETWEMPPAA SDVLSALRRS YDNLPPQLKL CFAFCALFTK GYRFRKDTLI HMWIAQNLIQ STESKRSEDM AEECFDDLVC RFFFRYSWGN
YVMNDSVHDL ARWVSLDEYF RADEDSPLHI SKPIRHLSWC SERITNVLED NNTGGDAVNP LSSLRTLLFL GQSEFRSYHL LDRMFRMLSR IRVLDFSNCV
IRNLPSSVGN LKHLRYLGLS NTRIQRLPES VTRLCLLQTL LLEGCELCRL PRSMSRLVKL RQLKANPDVI ADIAKVGRLI ELQELKAYNV DKKKGHGIAE
LSAMNQLHGD LSIRNLQNVE KTRESRKARL DEKQKLKLLD LRWADGRGAG ECDRDRKVLK GLRPHPNLRE LSIKYYGGTS SPSWMTDQYL PNMETIRLRS
CARLTELPCL GQLHILRHLH IDGMSQVRQI NLQFYGTGEV SGFPLLELLN IRRMPSLEEW SEPRRNCCYF PRLHKLLIED CPRLRNLPSL PPTLEELRIS
RTGLVDLPGF HGNGDVTTNV SLSSLHVSEC RELRSLSEGL LQHNLVALKT AAFTDCDSLE FLPAEGFRTA ISLESLIMTN CPLPCSFLLP SSLEHLKLQP
CLYPNNNEDS LSTCFENLTS LSFLDIKDCP NLSSFPPGPL CQLSALQHLS LVNCQRLQSI GFQALTSLES LTIQNCPRLT MSHSLVEVNN SSDTGLAFNI
TRWMRRRTGD DGLMLRHRAQ NDSFFGGLLQ HLTFLQFLKI CQCPQLVTFT GEEEEKWRNL TSLQILHIVD CPNLEVLPAN LQSLCSLSTL YIVRCPRIHA
FPPGGVSMSL AHLVIHECPQ LCQRCDPPGG DDWPLIANVP RICLGRTHPC RCSTT

Figure SR8-5. Amino acid sequence of MamRGA2 with tryptic peptides identified via LC-MS/MS highlighted in green. Note that the 81 tryptic peptides identified correspond to the 21 highlighted portions of the amino acid sequence of MamRGA2 due to redundancy in the sequence of the identified peptides.

MGSSHHHHHH SSGLVPRGSH MASMAGVTSQ AAAVFSLVNE IFNRSINLIV AELRLQLNAR AELNNLQRTL LRTHSLLEEA KARRMTDKSL VLWLMELKEW
AYDADDILDE YEAAAIRLKV TRSTFKRLID HVIINVPLAH KVADIRKRLN GVTLERELNL GALEGSQPLD STKRGVTTSL LTESCIVGRA QDKENLIRLL
LEPSDGAVPV VPIVGLGGAG KTTLSQLIFN DKRVEEHFPL RMWVCVSDDF DVKRITREIT EYATNGRFMD LTNLNMLQVN LKEEIRGTTF LLVLDDVWNE
DPVKWESLLA PLDAGGRGSV VIVTTQSKKV ADVTGTMEPY VLEELTEDDS WSLIESHSFR EASCSSTNPR MEEIGRKIAK KISGLPYGAT AMGRYLRSKH
GESSWREVLE TETWEMPPAA SDVLSALRRS YDNLPPQLKL CFAFCALFTK GYRFRKDTLI HMWIAQNLIQ STESKRSEDM AEECFDDLVC RFFFRYSWGN
YVMNDSVHDL ARWVSLDEYF RADEDSPLHI SKPIRHLSWC SERITNVLED NNTGGDAVNP LSSLRTLLFL GQSEFRSYHL LDRMFRMLSR IRVLDFSNCV
IRNLPSSVGN LKHLRYLGLS NTRIQRLPES VTRLCLLQTL LLEGCELCRL PRSMSRLVKL RQLKANPDVI ADIAKVGRLI ELQELKAYNV DKKKGHGIAE
LSAMNQLHGD LSIRNLQNVE KTRESRKARL DEKQKLKLLD LRWADGRGAG ECDRDRKVLK GLRPHPNLRE LSIKYYGGTS SPSWMTDQYL PNMETIRLRS
CARLTELPCL GQLHILRHLH IDGMSQVRQI NLQFYGTGEV SGFPLLELLN IRRMPSLEEW SEPRRNCCYF PRLHKLLIED CPRLRNLPSL PPTLEELRIS
RTGLVDLPGF HGNGDVTTNV SLSSLHVSEC RELRSLSEGL LQHNLVALKT AAFTDCDSLE FLPAEGFRTA ISLESLIMTN CPLPCSFLLP SSLEHLKLQP
CLYPNNNEDS LSTCFENLTS LSFLDIKDCP NLSSFPPGPL CQLSALQHLS LVNCQRLQSI GFQALTSLES LTIQNCPRLT MSHSLVEVNN SSDTGLAFNI
TRWMRRRTGD DGLMLRHRAQ NDSFFGGLLQ HLTFLQFLKI CQCPQLVTFT GEEEEKWRNL TSLQILHIVD CPNLEVLPAN LQSLCSLSTL YIVRCPRIHA
FPPGGVSMSL AHLVIHECPQ LCQRCDPPGG DDWPLIANVP RICLGRTHPC RCSTT

Figure SR8-6. Amino acid sequence of MamRGA2 with chymotryptic peptides identified via LC-MS/MS highlighted in green. Note that the 129 chymotryptic peptides identified correspond to the 29 highlighted portions of the amino acid sequence of MamRGA2 due to redundancy in the sequence of the identified peptides.

MGSSHHHHHH SSGLVPRGSH MASMAGVTSQ AAAVFSLVNE IFNRSINLIV AELRLQLNAR AELNNLQRTL LRTHSLLEEA KARRMTDKSL VLWLMELKEW
AYDADDILDE YEAAAIRLKV TRSTFKRLID HVIINVPLAH KVADIRKRLN GVTLERELNL GALEGSQPLD STKRGVTTSL LTESCIVGRA QDKENLIRLL
LEPSDGAVPV VPIVGLGGAG KTTLSQLIFN DKRVEEHFPL RMWVCVSDDF DVKRITREIT EYATNGRFMD LTNLNMLQVN LKEEIRGTTF LLVLDDVWNE
DPVKWESLLA PLDAGGRGSV VIVTTQSKKV ADVTGTMEPY VLEELTEDDS WSLIESHSFR EASCSSTNPR MEEIGRKIAK KISGLPYGAT AMGRYLRSKH
GESSWREVLE TETWEMPPAA SDVLSALRRS YDNLPPQLKL CFAFCALFTK GYRFRKDTLI HMWIAQNLIQ STESKRSEDM AEECFDDLVC RFFFRYSWGN
YVMNDSVHDL ARWVSLDEYF RADEDSPLHI SKPIRHLSWC SERITNVLED NNTGGDAVNP LSSLRTLLFL GQSEFRSYHL LDRMFRMLSR IRVLDFSNCV
IRNLPSSVGN LKHLRYLGLS NTRIQRLPES VTRLCLLQTL LLEGCELCRL PRSMSRLVKL RQLKANPDVI ADIAKVGRLI ELQELKAYNV DKKKGHGIAE
LSAMNQLHGD LSIRNLQNVE KTRESRKARL DEKQKLKLLD LRWADGRGAG ECDRDRKVLK GLRPHPNLRE LSIKYYGGTS SPSWMTDQYL PNMETIRLRS
CARLTELPCL GQLHILRHLH IDGMSQVRQI NLQFYGTGEV SGFPLLELLN IRRMPSLEEW SEPRRNCCYF PRLHKLLIED CPRLRNLPSL PPTLEELRIS
RTGLVDLPGF HGNGDVTTNV SLSSLHVSEC RELRSLSEGL LQHNLVALKT AAFTDCDSLE FLPAEGFRTA ISLESLIMTN CPLPCSFLLP SSLEHLKLQP
CLYPNNNEDS LSTCFENLTS LSFLDIKDCP NLSSFPPGPL CQLSALQHLS LVNCQRLQSI GFQALTSLES LTIQNCPRLT MSHSLVEVNN SSDTGLAFNI
TRWMRRRTGD DGLMLRHRAQ NDSFFGGLLQ HLTFLQFLKI CQCPQLVTFT GEEEEKWRNL TSLQILHIVD CPNLEVLPAN LQSLCSLSTL YIVRCPRIHA
FPPGGVSMSL AHLVIHECPQ LCQRCDPPGG DDWPLIANVP RICLGRTHPC RCSTT

Figure SR8-7. Amino acid sequence of MamRGA2 with sequences identified solely from chymotryptic peptides, solely from tryptic peptides, and by both chymotryptic and tryptic peptides highlighted in green, blue and yellow, respectively.

For more details on the mass spectrometric identification of the *E. coli*-expressed recombinant MamRGA2 protein from tryptic and chymotryptic peptides, including detail of the materials and methods used, refer to supporting documents SD8-2 to SD8-6 (section 8.5).

8.4 Conclusions

Based on a combination of physicochemical analyses, the *E. coli*-expressed MamRGA2 protein was found to be equivalent to the MamRGA2 expressed in event QCAV-4 and was a suitable surrogate for conducting relevant safety studies.

8.5 Supporting documents

- SD8-1-Mouse anti-MamRGA2 monoclonal antibody production.pdf
- SD8-2-APAF reports.pdf contains:
  - APAF report MS-R32978-2 - Mass spectrometric identification of an *Escherichia coli* expressed recombinant plant protein and associated files: SD8-3-APAF-P32978_rMamRGA_proteins.xlsx, SD8-4-APAF-P32978_rMamRGA_PSMs.xlsx
  - APAF report MS-R33006-2 - Mass spectrometric identification of an *Escherichia coli* expressed recombinant plant protein from chymotryptic peptides and associated files: : SD8-5-APAF-P33006_rMamRGA_proteins.xlsx and : SD8-6-APAF-P33006_rMamRGA_PSMs.xlsx

The "_proteins.xlsx" files list all proteins identified in the sequence database search from more than one peptide, together with details associated with each identification, with MamRGA2 highlighted in green. The proteins listed in this file include likely contaminant proteins derived from sample handling or carryover from previous mass spectrometry experiments. The "_PSMs.xlsx" files list peptide spectrum matches (PSMs) derived from rMamRGA2, and in-depth details associated with each match. Please note that single peptides can produce multiple PSMs. All MamRGA2 peptide sequences identified from the sequence database search are listed in column E.

QUT2023-9: *In vitro* assessment of the MamRGA2 protein lability to gastric enzymatic digestion and its thermal stability

STUDY IDENTIFICATION

QUT2023-9

AUTHORS

[redacted]
s 22

PERFORMING LABORATORY

[redacted]

[redacted] 2023

9.1 Introduction

Banana event QCAV-4 (QUT-QCAV4-6) was created by *Agrobacterium tumefaciens*-mediated transformation of banana (*Musa acuminata* subgroup Cavendish cv Grand Nain) embryogenic cells with plasmid pSAN3. This plasmid contains the *neomycin phosphotransferase* II (*npt*II) selectable marker gene from the Tn5 transposon of *Escherichia coli* strain K12 and the *MamRGA2* disease resistance (R) gene from the wild banana *Musa acuminata* ssp. *malaccensis* which confers resistance to the devastating banana fungal pathogen *Fusarium oxysporum* f. sp. *cubense* tropical race 4 (TR4) (Dale *et al*., 2017).

As part of the safety evaluation process for genetically engineered plants, the digestibility of newly introduced proteins is routinely examined. Proteins that can be easily digested are expected to behave like typical dietary proteins when consumed and are less likely to cause allergic or toxic reactions. The resistance of novel food proteins to pepsin digestion under acidic conditions is an important consideration for assessing their potential allergenicity and toxicity (Codex, 2003). Although the pepsin digestibility assay does not guarantee that a protein will always be digested in a consumer's stomach, it provides a standardised test to evaluate protein digestibility (Thomas *et al.,* 2004). The objective of this study was to evaluate the susceptibility of the MamRGA2 protein to (i) proteolytic degradation in simulated gastric fluid (SGF) containing pepsin and (ii) degradation induced by elevated temperatures.

9.2 Materials and Methods

9.2.1 Test proteins

Pepsin-susceptible Bovine Serum Albumin (BSA; Cat. #A7906, Sigma) and pepsin-resistant β-lactoglobulin (Cat. #L3908, Sigma) were dissolved in 1x denaturing equilibration buffer (50 mM sodium phosphate, 6 M guanidine-HCL, 300 mM NaCl) at a concentration of 2 mg/mL. The semi-purified *E. coli*-expressed MamRGA2 was prepared at 3.3 µg/µL total soluble protein (TSP, QUT2023-8) in 1x denaturing equilibration buffer. The equivalence of the *E. coli*-expressed MamRGA2 to the QCAV-4-expressed MamRGA2 is evaluated in study report QUT2023-8.

9.2.2 Bioinformatic searches for putative proteolytic enzymes cleavage sites

Searches were done on the PeptideCutter tool on the ExPASy Proteomics Site (Gasteiger *et al.,* 2005) using the amino acid sequence of MamRGA2 as query and default parameters for pepsin, trypsin and chymotrypsin under various conditions.

9.2.3 Conditions for protein digestibility assays in SGF containing pepsin

The methodology described by Thomas *et al.* (2004) was followed in this study. For each protein, a single tube containing 1.43 mL of SGF (0.084 M HCl, 35 mM NaCl, pH 1.4) and 50 µL of pepsin (10 µg/µL) was pre-heated to 37˚C prior to the addition of 120 µL of MamRGA2 protein solution (equivalent to ~27 µg of MamRGA2 based on densitometry) or 100 µL of control BSA and β-lactoglobulin protein solutions (equivalent to 200 µg control test protein). The tubes were mixed by brief vortexing and immediately placed in a 37˚C heating block. Samples (200 µL) were removed at 0.5, 2, 5, 10, 20, 30 and 60 min after initiation of the incubation. Each 200 µL sample was quenched by the addition of 70 µL of 200 mM $NaHCO_3$, pH 11, and 70 µL of 5x Laemmli buffer (40% glycerol, 5% β-mercaptoethanol, 10% SDS, 0.33 M Tris, 0.05% bromophenol blue, pH 6.8) (Laemmli, 1970). Quenched samples were heated to 95˚C for 10 min and analysed directly or stored at -20˚C. The zero-time point protein digestion samples were prepared by quenching the pepsin in the solution before adding the test protein. Control samples for pepsin auto-digestion (pepsin without test protein) and test protein stability (reaction buffer with test protein but without pepsin) were included. These control reactions were treated as described above except samples

were prepared only for the 0- and 60-min time points. Due to solubility constraints with MamRGA2, the dilution buffer required the presence of 6 M guanidine-HCL which precluded sample analysis by PAGE. To overcome this issue, prior to the addition of 5x Laemmli buffer, the protein was precipitated by adding 9-volumes of ice-cold ethanol (100%) to the quenched solution and placing samples at -20˚C for at least 1 h. Samples were then centrifuged at 15,000 x*g* for 10 min at 4˚C to pellet the protein. The pellet was further washed once with ice-cold ethanol and centrifuged again at 15,000 x*g* for 10 min at 4˚C. The pellet was then dried completely, resuspended in 270 µL dH_20 and 70 µL of 5x Laemmli buffer, and heated to 95˚C for 10 min to be analysed directly or stored at -20˚C.

9.2.4 Pepsin activity

Porcine gastric mucosa pepsin was purchased as a single lot with a supplied activity of 4036 U/mg (Cat. #P6887, Sigma). The digestibility assays of the control BSA and β-lactoglobulin proteins contained 0.5 mg of pepsin (50 µL of pepsin at 10 µg/µL) with an activity of 4036 U/mg and 200 µg of TSP (100 µL at 2 µg/µL TSP). This represents a pepsin to control test solution ratio of ~10 U/µg (2018 U/200 µg TSP) as recommended by Thomas *et al.* (2004). To be conservative, the pepsin to MamRGA2 solution ratio was purposely kept at half the recommended amount. The MamRGA2 digestibility assays contained 0.5 mg of pepsin (50 µL of pepsin at 10 µg/µL) with an activity of 4036 U/mg and 396 µg of TSP (120 µL of MamRGA2 protein solution at 3.3 µg/µL TSP). This represents a pepsin to MamRGA2 solution ratio of around 5 U/µg (2018 U/396 µg TSP).

9.2.5 SDS-PAGE and Western Blot analysis

Samples (10 µL) from each time point and control reaction were subjected to SDS-PAGE using 12-well, 4-20% Mini-PROTEAN TGX Precast Protein gels (Cat. #4561095, Bio-Rad). Colour Protein Standard markers, Broad Range (10-250 kDa) (Cat. #P7719S, New England Biolabs) or Broad Range (11-245 kDa) (Cat. #P7712S, New England Biolabs) were included on all gels. Gels were stained with Coomassie brilliant blue R-250 (Cat. #0472-25G, Amresco) and de-stained in a solution of 10% acetic acid and 50% methanol for 2-3 h followed by deionised water for 1-2 h. In addition, for the MamRGA2 digest samples, duplicate gels were used, and the proteins transferred to nitrocellulose membranes (Cat. #1704158, Bio-Rad) for Western blotting. One membrane was blocked in 3% BSA/Tris-buffered saline with Tween® 20 (TBST) for 1 h prior to incubation with His_6-tag monoclonal antibody-HRP (Cat. #MA1-135-HRP; ThermoFisher), 1:3,000 dilution in 3% BSA/TBST for 1 h. The second membrane was blocked in 3% BSA/TBST for 1 h prior to incubation with primary custom-made mouse anti-MamRGA2 monoclonal antibody 17F07 (MBS, BBI Solutions), 1:1,000 dilution in 3% BSA/TBST for 1 h followed by secondary goat-anti-mouse IgG-HRP (Cat. #62-6520, ThermoFisher), 1:5,000 dilution in 3% BSA/TBST for 1 h. Chemiluminescence was visualised using the Clarity™ Western ECL Substrate (Cat. #170-5060, Bio-Rad) and imaged on the Chemidoc MP imaging system (Bio-Rad) at various durations of exposure.

9.2.6 Thermal stability of MamRGA2

Samples of *E. coli*-expressed MamRGA2 protein were incubated at 60, 75 or 90°C for up to 60 min to determine the thermal stability of the protein. The production and characterisation of the MamRGA2 test protein is described in detail in study report QUT2023-8. Assay samples contained 24 µg of MamRGA2 (30 µL of a 0.8 ug/µL stock solution) diluted in 370 µL of denaturing buffer (or 60 ng/µL of MamRGA2). Original samples kept at 4°C and room temperature (RT) were included as controls. Aliquots (100 µL) were taken after 10, 30 and 60 min of incubation and combined with 100 µL 2x Laemmli buffer for SDS-PAGE analysis as described in 9.2.4. For Coomassie staining, 5 µL of this solution was loaded onto the gel which is equivalent to 150 ng of MamRGA2 per lane (30 ng/µL x 5 µL = 150 ng) while the equivalent of 5 ng per lane of MamRGA2 protein was loaded for Western blot analysis.

9.3 Results

9.3.1 MamRGA2 stability to proteolysis in simulated gastric fluid (SGF)

The susceptibility of MamRGA2 to digestion by proteolytic enzymes under acidic conditions was examined using two approaches. Firstly, a bioinformatic assessment was conducted to identify potential protease cleavage sites using the amino acid sequence of MamRGA2 as query in PeptideCutter. This analysis revealed MamRGA2 had multiple cleavage sites for pepsin (308 sites at pH 1.3 and 354 sites at pH >2), trypsin (145 sites), chymotrypsin (781 sites) (Table SR9-1). On this basis, MamRGA2 was considered likely to be as susceptible to digestion as most dietary proteins.

Table SR9-1. PeptideCutter search result using the amino acid sequence of MamRGA2 as query

Proteases	No. of PeptideCutter predicted cleavage sites
Pepsin (pH1.3)	308
Pepsin (pH>2)	354
Trypsin	145
Chymotrypsin	781

The safety of novel proteins inserted into genetically modified plants has routinely included an evaluation of their susceptibility to digestion. To examine this, the susceptibility of semi-purified *E. coli*-expressed MamRGA2 protein to digestion by pepsin in simulated gastric fluid (SGF) was performed. Detailed analysis assessing the suitability and equivalence of the *E. coli*-expressed MamRGA2 protein compared to the plant-expressed MamRGA2 protein is available in study report QUT2023-8. Protein extracts from *E. coli*-expressing MamRGA2 (~200 µg TSP of which MamRGA2 was estimated to be ~13.5 µg by densitometry) were incubated in the presence of SGF pH 1.4 containing pepsin at 37°C for 0, 0.5, 2, 5, 10, 20, 30 and 60 min (Figure SR9-1A-D). Control digestions with pepsin-insensitive β-lactoglobulin and pepsin-sensitive BSA were performed under the same conditions. Samples were removed at stated time points and separated by SDS-PAGE and either stained using Coomassie brilliant blue (Figure SR9-1A, C, E and F) or the proteins transferred to a nitrocellulose membrane for Western blot analysis (Figure SR9-1B and D). Since an N-terminal His_6-tag was added to MamRGA2 for purification purposes, a His_6-tag monoclonal antibody was also used for detection purposes in Western analysis (Figure SR9-1B) along with the monoclonal mouse anti-MamRGA2 monoclonal antibody 17F07 (Figure SR9-1D).

The pepsin-insensitive protein, β-lactoglobulin, was still present after 60 minutes of incubation in SGF containing pepsin (Figure SR9-1E), while the pepsin-sensitive protein, BSA, was completely digested within 30 seconds following the addition of pepsin to the assay (Figure SR9-1F). MamRGA2 in SGF was substantially degraded 30 seconds after the addition of pepsin (Figure SR9-1A-D) and was undetectable after 2 minutes of digestion (Figure SR9-1D). Faint, low molecular weight degradation products were visible (by Coomassie) in samples subjected to up to 60 minutes of digestion (Figure SR9-1A and C), but these were not detected by Western blot (Figure SR9-1B and D).

9.3.2 Thermal stability of MamRGA2

Although most Cavendish bananas in Australia are consumed as a fresh fruit, some are fried to prepare fritters or banana chips while green (starchy, unripe) bananas have been used to make banana flour for use in baking. As an NLR protein, MamRGA2 is thought to play a role in the activation of the plant defences against Fusarium wilt TR4. As such, it was not possible to design a protein activity assay to assess the functionality of MamRGA2 following exposure to thermal treatments. Therefore, the effect of heat on the

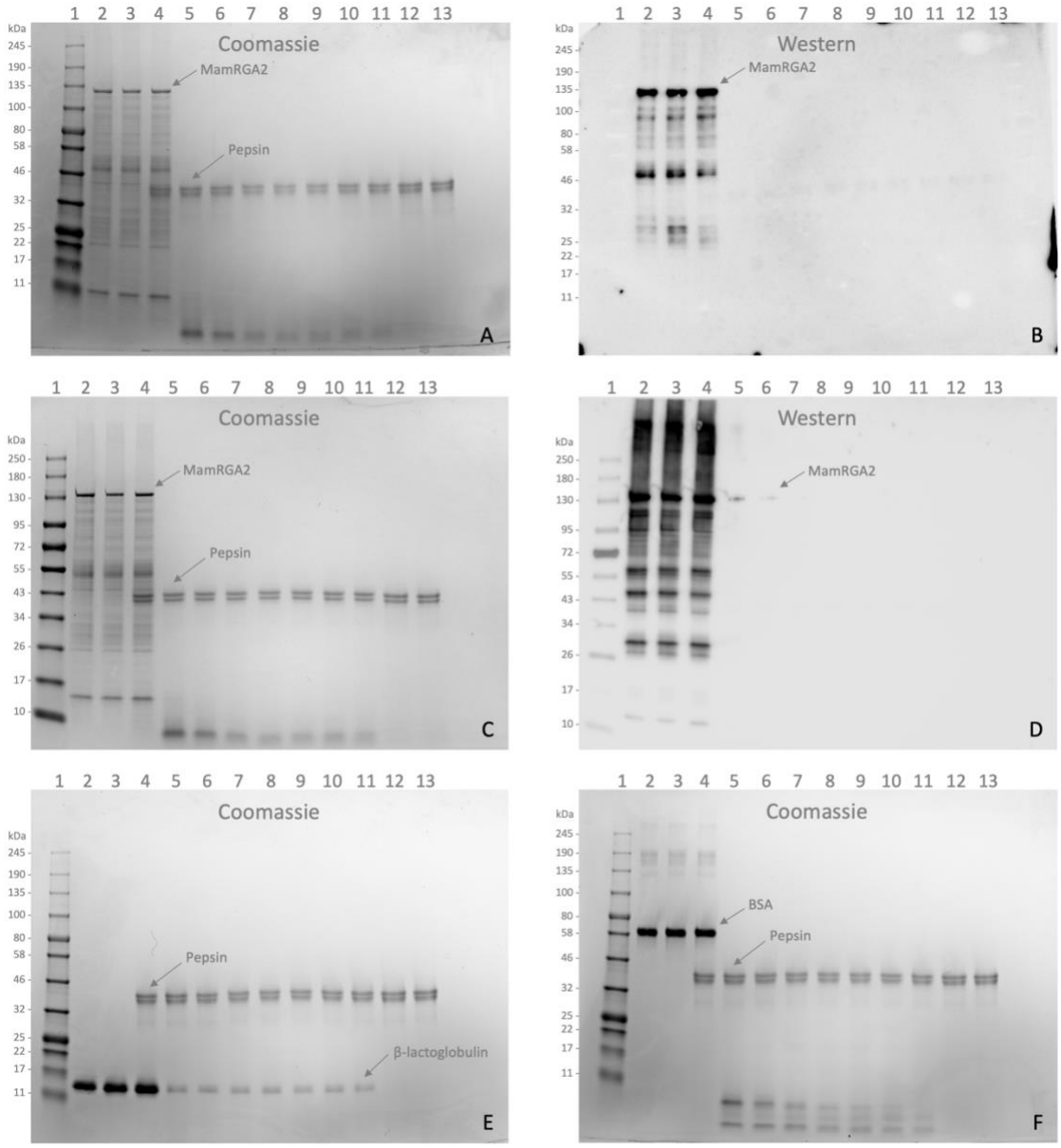


Figure SR9-1. Lability of *E. coli*-expressed MamRGA2 protein in SGF pH 1.4 containing pepsin. Digests of *E. coli*-expressed MamRGA2 (A-D), pepsin-insensitive β-lactoglobulin (Cat. #L3908, Sigma) control (E) and pepsin-sensitive bovine serum albumin (BSA, Cat. #P0834, Sigma) control (F), all under denaturing conditions. Panels A, C, E and F, 4-20% Mini-PROTEAN TGX Precast Protein Gel, Coomassie brilliant blue R-250 stain (Cat. #0472-25G, Amresco) and panels B and D, Western immunoblots. Panel B, detection with His_6-tag monoclonal antibody-HRP (Cat. #MA1-135-HRP, ThermoFisher), 1:3,000, chemiluminescent substrate development and exposure time of 20 min. Panel D, detection with primary mouse anti-MamRGA2 monoclonal antibody 17F07 (MBS, BBI Solutions), 1:1,000 followed by secondary goat-anti-mouse IgG-HRP (Cat. #62-6520, ThermoFisher), 1:5,000, chemiluminescent substrate development and exposure time of 98 s. Lane 1, molecular weight standards either Colour Pre-stained Protein Standard, Broad Range (11-245 kDa) (Panels A, B, E and F) or (10-250 kDa) (Panels C and D) (Cat. #P7712S and P7719S, New England Biolabs); lane 2-3, *E. coli*-expressed MamRGA2 or control protein incubated in SGF pH 1.4 for 0 and 60 min; lanes 4-11, *E. coli*-expressed MamRGA2 or control protein incubated in SGF pH 1.4 containing pepsin for 0, 0.5, 2, 5, 10, 20, 30 and 60 min; and lanes 12-13, SGF containing pepsin only for 0 and 60 min.

structural integrity and immuno-detectability of the MamRGA2 protein was evaluated by SDS-PAGE and Western blot analysis, respectively. Aliquots of *E. coli*-expressed MamRGA2 (~150 ng) were heated to 60, 75 and 90°C for 15, 30 and 60 minutes, while control samples were kept either at 4°C or 22°C for 60 min before being subjected to SDS-PAGE (Figure SR9-2A). For Western blot analysis, samples were diluted 30x to load the equivalent to 5 ng of MamRGA2 per lane (Figure SR9-2B). The intensity of the ~142 kDa *E. coli*-expressed MamRGA2 protein kept at 4°C was equivalent to that of the sample kept at room temperature (22°C). Further, no visible differences in band intensity or degradation of MamRGA2 were observed irrespective of the temperature used or the length of incubation (Figure SR9-2). These results suggest that the MamRGA2 protein is not significantly degraded at temperatures up to 90°C.

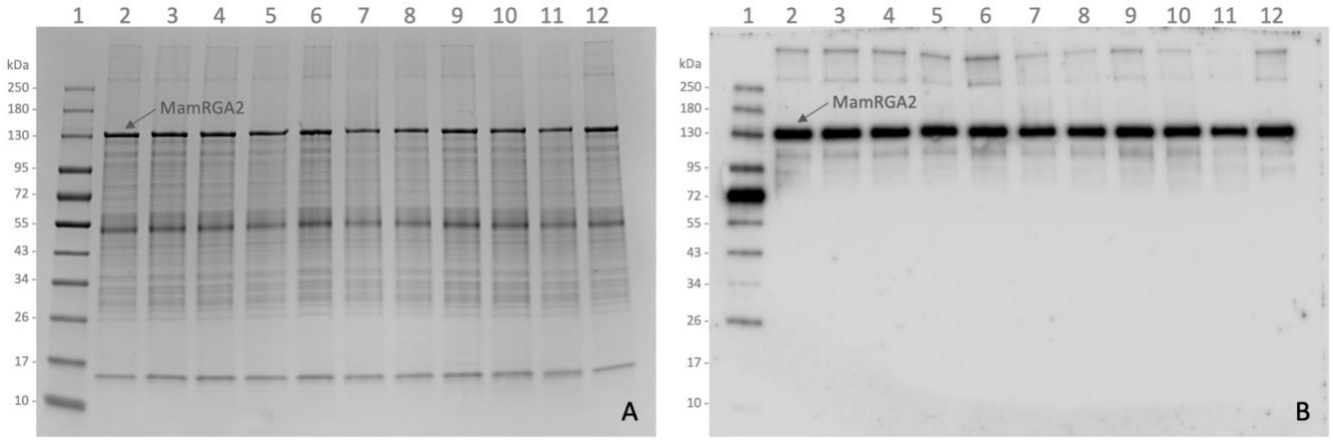


Figure SR9-2. Thermal stability of MamRGA2. A: ~150 ng of heat-treated *E. coli*-expressed MamRGA2 per lane of a 4-20% Mini-PROTEAN TGX Precast Protein Gel, Coomassie brilliant blue R-250 stain (Cat. #0472-25G, Amresco) and B: Western immunoblot of 5 ng of heat-treated *E. coli*-expressed MamRGA2 per lane. Detection with primary mouse anti-MamRGA2 monoclonal antibody 17F07 (MBS, BBI Solutions), 1:1,000 followed by secondary goat-anti-mouse IgG-HRP (Cat. #62-6520, ThermoFisher), 1:5,000, chemiluminescent substrate development and auto optimal exposure time of 30 min. Lanes 1, Colour Pre-stained Protein Standard, Broad Range (10-250 kDa) (Cat. #P7719S, New England Biolabs); lanes 2-12, *E. coli*-expressed MamRGA2 treatment as follows: lanes 2, 4°C for 60 min; lanes 3-5, 60°C for 10, 30 and 60 min, lanes 6-8, 75°C for 10, 30 and 60 min, lanes 9-11, 90°C for 10, 30 and 60 min and lane 12, 22°C for 60 min.

9.4 Conclusions

The results of this *in vitro* digestion analysis support the conclusion that the MamRGA2 protein, like most conventional dietary proteins, will be readily digested in a typical mammalian gastric environment. Further, the integrity of the protein is not likely to be affected by moderate temperatures used during cooking. Therefore, no increased risk of allergenicity or toxicity would be anticipated from dietary exposure to this protein (Codex, 2003).

9.5 Supporting documents

- SD9-1-PeptideCutter search results.xlsx
- SD9-2-PeptideCutter search results.pdf

QUT2023-10: Nutrient composition of fruit and peel tissues harvested from banana event QCAV-4 and non-GM Grand Nain control GN212-12

STUDY IDENTIFICATION

QUT2023-10

AUTHORS

[redacted]
s 22
[redacted]
[redacted]

PERFORMING LABORATORY

[redacted]
[redacted]
[redacted]

-

[redacted]
[redacted]

[redacted]

[redacted] 2023

10.1 Introduction

Banana event QCAV-4 (QUT-QCAV4-6) was created by *Agrobacterium tumefaciens*-mediated transformation of banana (*Musa acuminata* subgroup Cavendish cv Grand Nain) embryogenic cells with plasmid pSAN3. This plasmid contains the *neomycin phosphotransferase* II (*npt*II) selectable marker gene from the Tn5 transposon of *Escherichia coli* strain K12 and the *MamRGA2* disease resistance (R) gene from the wild banana *Musa acuminata* ssp. *malaccensis* which confers resistance to the devastating banana fungal pathogen *Fusarium oxysporum* f. sp. *cubense* tropical race 4 (TR4) (Dale *et al*., 2017).

The trait introduced into Cavendish banana to generate event QCAV-4 was not intended or expected to alter the composition of its edible parts. In this study, a comparative compositional analysis is made between event QCAV-4 and its non-GM counterpart (GN212-12) for evaluating the safety and potential risks associated with the genetic modification. The analysis helps to establish substantial equivalence by identifying any unintended changes in the composition of event QCAV-4 that may have an impact on the safety and nutritional value of the food derived from it.

10.2 Materials and Methods

10.2.1 Sample collection and processing

Fruit for this analysis was grown and harvested from a confined field trial conducted in the Northern Territory under OGTR licence DIR146 (2018-present). Details on the field trial site, including location, experimental design, management and growth conditions, are contained in study report QUT2023-3.

Fruit was harvested from the third top hand of full-green bunches, boxed and shipped by QANTAS freight to QUT, Gardens Point Campus, Brisbane and subsequently ripened by exposure to ethylene (Ripe Gas, BOC) to a ripening stage 66 (BAN-C-1, 2001). Pulp and peel samples (~300 g) were sliced and stored at -80°C until all the samples were collected.

Two batches of fruit were sent on dry ice to the s 22 for analysis. The first batch containing ripe fruit samples from 10 QCAV-4 and 10 non-GM GN212-12 control plants harvested from the 5th generation (ratoon 4) of plants was sent to NMI on October 12th, 2021. The second batch containing ripe fruit samples from 10 QCAV-4 and 10 non-GM control plants from the 6th generation (ratoon 5) as well as peel samples from ripened fruit from six QCAV-4 and two non-GM control plants from the 7th generation (ratoon 6) of plants was sent to NMI on June 21st, 2022.

10.2.2 Compositional analysis

All samples were analysed by for the content of (i) proximates (moisture, total fat, total protein, ash, carbohydrates and energy), (ii) three minerals (magnesium, manganese, potassium), and (iii) two vitamins (ascorbic acid (vitamin C) and pyridoxine (vitamin B6)) which represent the highest contributors to the percent daily values from a 2,000 calorie reference diet for adults and children aged four or over (USDA, 2019; https://nutritiondata.self.com/). For each analyte tested, the mean (± standard deviation (SD)) was calculated and the associated range of the data provided. In addition, a combined literature range (CLR) of values for each analyte adapted from the Australian Food Composition Database (FSANZ, 2022) and the FoodData Central database of the U.S. Department of Agriculture (USDA, 2019) is reported.

10.2.3 Statistical analysis

Statistical analysis (independent samples t-Test) was done on SPSS® Statistics Version 27 (IBM®) and significant differences with the non-GM control reported at $p<0.05$.

10.3 Results

In the absence of a banana-specific OECD consensus document on compositional considerations for new banana varieties, samples were analysed for the content of (i) proximates (moisture, total fat, total protein, ash, carbohydrates and energy), (ii) three minerals (magnesium, manganese, potassium), and (iii) two vitamins (ascorbic acid (vitamin C) and pyridoxine (vitamin B6)) which represent the highest contributors to the percent daily values from a 2,000 calorie reference diet for adults and children aged four or over (USDA, 2019; https://nutritiondata.self.com/). Further analysis of individual amino acid and fatty acid composition was not deemed a necessary consideration for the safety assessment of event QCAV-4 based upon the following considerations. The recommended daily intake (RDI) of protein for Australian males and females is 64-81 g/day and 46-57 g/day, respectively (NRVANZ, 2006). With a protein concentration of approximately 1 g/100 g (fresh weight, FW) and assuming 100% bioavailability, an average male would need to consume more than 3.2 kg (~25 peeled fruit) of banana per day to ingest 50% of their RDI. Regarding fatty acid intake, the Nutrient Reference Values for Australia and New Zealand (NRVANV, 2006) state that for children, adolescents and adults, the estimated average requirement (EAR), RDI or adequate intake (AI) for total fat has not been set since it is the type of fats consumed that relate to essentiality and to many of the physiological and health outcomes. However, AIs have been set for the most common dietary polyunsaturated fatty acid, linoleic acid (LA), with the female and male AIs established at around 8 and 13 g/day, respectively (NRVANZ, 2006). With a total fat concentration of around 0.2 g/100 g (FW), assuming 100% of banana fat was LA and 100% bioavailability, an average male would need to consume more than 6.5 kg (~50 peeled fruit) of banana per day to ingest the recommended AI for LA. An average Australian male requires a minimum of 23 kg (64 g x 365 days) of protein and 4.7 kg (13 g x 365 days) of linoleic acid per year as extrapolated from the protein RDI and linoleic acid AI for these nutrients (NRVANV, 2006). To meet these requirements for protein and LA, an average Australian male would need to consume a minimum of 2.3 tonnes and 470 kg of banana per year, respectively. Since Australians consume on average only 16 kg of Cavendish banana per year (Hort Innovation, 2022), this would only represent ~0.7% of both their protein (160 g) and fat (32 g assuming 100% LA) yearly requirement. In summary, bananas cannot be considered a significant contributor to the protein and fat intake of the typical Australian diet. Therefore, any changes (even significant) in the amino acid or fatty acid composition of edible parts of event QCAV-4 are inconsequential to the biosafety assessment of this event.

The nutritional composition of both banana fruit and peel tissue from event QCAV-4 was compared to fruit and peel tissue derived from its non-GM Grand Nain (GN212-12) counterpart (control). The data obtained from the 5th generation of QCAV-4 fruit showed statistical differences with the non-GM counterparts for most analytes except for fat, manganese and pyridoxine (Table SR10-1). Despite these differences, the mean data from all QCAV-4 fruit fell within the range of the non-GM control fruit grown at the same location. Further, for both QCAV-4 and non-GM control fruit, the levels of all analytes fell within the CLR except for manganese, ascorbic acid and pyridoxine which were lower than the CLR. When fruit from the 6th generation was analysed, a statistical difference between the QCAV-4 and the non-GM dataset was only observed for the average fruit manganese concentration (Table SR10-2). Like the analyses from the 5th generation of fruit, all mean QCAV-4 fruit data fell within the reported range of the non-GM control fruit grown at the same location. Further, for both QCAV-4 and the non-GM control fruit, the levels of all analytes fell within the CLR except for ascorbic acid which was found to be lower than the

Table SR10-1. Generation 5 banana fruit proximates, minerals and vitamins

Analyte	Genotype	n	Mean ± SD	*P* value[1]	Range (min - max)	CLR[2] (min - max)
Moisture (g/100 g)	QCAV-4	10	79.1 ± 1.0	0.0002***	77.6 - **80.4**	71.3 - 80.6
	non-GM	10	76.9 ± 1.1		75.5 - 78.6	
Fat (g/100 g)	QCAV-4	10	0.19 ± 0.00	0.0618	0.19 - 0.19	0.00 - 0.72
	non-GM	10	0.23 ± 0.05		0.19 - 0.30	
Protein (g/100 g)	QCAV-4	10	0.97 ± 0.09	0.0069**	**0.80** - 1.10	0.62 - 1.40
	non-GM	10	1.08 ± 0.06		1.00 - 1.20	
Ash (g/100 g)	QCAV-4	10	0.79 ± 0.11	0.0470*	0.70 - 1.00	0.43 - 1.00
	non-GM	10	0.89 ± 0.10		0.70 - 1.00	
Carbohydrates (g/100 g)	QCAV-4	10	19.1 ± 1.0	0.0018**	**18.0** - 21.0	17.3 - 27.5
	non-GM	10	20.9 ± 1.2		19.0 - 22.0	
Energy (kJ/100 g)	QCAV-4	10	340 ± 17	0.0003***	**320** - 370	287 - 426
	non-GM	10	377 ± 19		350 - 400	
Magnesium (mg/kg)	QCAV-4	10	261 ± 14	0.0004***	**230** - 280	180 - 380
	non-GM	10	294 ± 20		270 - 340	
Manganese (mg/kg)	QCAV-4	10	0.52 ± 0.08	0.1173	0.39 - 0.64	0.93 - 8.29
	non-GM	10	0.65 ± 0.23		0.39 - 0.98	
Potassium (mg/kg)	QCAV-4	10	3,852 ± 178	0.0436*	3,550 - **4,120**	3,000 - 4,260
	non-GM	10	3,652 ± 231		3,360 - 4,030	
Ascorbic Acid (mg/100 g)	QCAV-4	10	1.96 ± 0.18	0.0470*	1.70 - 2.20	4.0 - 15.1
	non-GM	10	1.72 ± 0.31		1.30 - 2.30	
Pyridoxine (mg/100 g)	QCAV-4	10	0.12 ± 0.04	0.6278	0.10 - 0.20	0.19 - 0.42
	non-GM	10	0.13 ± 0.05		0.10 - 0.20	

[1]Independent samples t-Test, significant differences with control asserted at 95%*, 99%**and 99.9%***

[2]Combined literature range (CLR) from the Australian Food Composition Database (F000262: Banana, cavendish, peeled, raw) (FSANZ, 2022) and the FoodData Central database of the U.S. Department of Agriculture, Agricultural Research Service for "Bananas, raw (FDC ID: 173944)", "Bananas, overripe, raw (FDC ID: 1105073)" and "Bananas, ripe and slightly ripe, raw (FDC ID: 1105314)" (USDA, 2019).

n = biological replicates; SD = standard deviation.

Mean values for QCAV-4 and/or non-GM control analytes wholly outside the CLR ranges are shaded in orange.

Value in bold extend outside the range of the non-GM control dataset.

Table SR10-2. Generation 6 banana fruit proximates, minerals and vitamins

Analyte	Genotype	n	Mean ± SD	*P* value[1]	Range (min - max)	CLR[2] (min - max)
Moisture (g/100 g)	QCAV-4	10	79.2 ± 1.2	0.0913	78.1 - **81.3**	71.3 - 80.6
	non-GM	10	78.1 ± 1.7		75.5 - 80.8	
Fat (g/100 g)	QCAV-4	10	0.19 ± 0.00	1.0000	0.19 - 0.19	0.00 - 0.72
	non-GM	10	0.19 ± 0.00		0.19 - 0.19	
Protein (g/100 g)	QCAV-4	10	1.07 ± 0.29	0.1108	**0.30** - 1.40	0.62 - 1.40
	non-GM	10	1.24 ± 0.13		1.10 - 1.40	
Ash (g/100 g)	QCAV-4	10	1.05 ± 0.41	0.7934	**0.60** - **1.90**	0.43 - 1.00
	non-GM	10	1.10 ± 0.43		0.70 - 1.80	
Carbohydrates (g/100 g)	QCAV-4	10	18.8 ± 1.5	0.3785	**14.0** - 21.0	17.3 - 27.5
	non-GM	10	19.5 ± 1.9		17.0 - 22.0	
Energy (kJ/100 g)	QCAV-4	10	336 ± 25	0.2281	**250** - 360	287 - 426
	non-GM	10	352 ± 32		310 - 400	
Magnesium (mg/kg)	QCAV-4	10	295 ± 20	0.2467	**260** - 330	180 - 380
	non-GM	10	308 ± 28		270 - 370	
Manganese (mg/kg)	QCAV-4	10	0.63 ± 0.14	0.0475*	**0.42** - 0.92	0.93 - 8.29
	non-GM	10	0.88 ± 0.34		0.55 - 1.70	
Potassium (mg/kg)	QCAV-4	10	3,788 ± 204	0.7682	**3,450** - 4,110	3,000 - 4,260
	non-GM	10	3,760 ± 214		3,510 - 4,120	
Ascorbic Acid (mg/100 g)	QCAV-4	10	2.28 ± 0.43	0.1011	1.40 - **3.00**	4.0 - 15.1
	non-GM	10	1.97 ± 0.37		1.30 - 2.60	
Pyridoxine (mg/100 g)	QCAV-4	10	0.33 ± 0.03	0.9436	**0.27** - 0.39	0.19 - 0.42
	non-GM	10	0.34 ± 0.03		0.30 - 0.40	

[1]Independent samples t-Test, significant differences with control asserted at 95%*, 99%**and 99.9%***
[2]Combined literature range (CLR) from the Australian Food Composition Database (F000262: Banana, cavendish, peeled, raw) (FSANZ, 2022) and the FoodData Central database of the U.S. Department of Agriculture, Agricultural Research Service for "Bananas, raw (FDC ID: 173944)", "Bananas, overripe, raw (FDC ID: 1105073)" and "Bananas, ripe and slightly ripe, raw (FDC ID: 1105314)" (USDA, 2019).
n = biological replicates; SD = standard deviation.
Mean values for QCAV-4 and/or non-GM control analytes wholly outside the CLR ranges are shaded in orange.
Value in bold extend outside the range of the non-GM control dataset.

CLR. The average fruit manganese concentration from QCAV-4 was also lower than the CLR (Table SR10-2).

The concentration of the selected analytes was also measured in peel tissue collected from fruit obtained from six QCAV-4 plants and two non-GM control plants growing in generation 7 (Table SR10-3). A statistical difference between the QCAV-4 and the non-GM dataset was only observed for protein which was found to be significantly higher in the peel of QCAV-4 fruit. Further, the mean values (value ± SD) for protein and magnesium in QCAV-4 peel were found to be outside the range of the non-GM mean values. However, at less than 1% protein content, banana peel is not a large contributor to the protein intake of the human diet making these differences insignificant from a nutritional perspective. Due to the lack of reliable data for banana peel, a CLR could not be accurately determined making the interpretation of our data in the wider context particularly difficult.

Table SR10-3. Generation 7 banana peel proximates, minerals and vitamins

Analyte	Genotype	n	Mean ± SD	*P* value[1]	Range (min - max)	CLR[2]
Moisture (g/100 g)	QCAV-4	6	90.6 ± 1.3	0.7670	89.5 - **92.8**	89.8
	non-GM	2	90.3 ± 1.8		89.0 - 91.5	
Fat (g/100 g)	QCAV-4	6	0.47 ± 0.12	0.8641	**0.30** - **0.60**	0.58
	non-GM	2	0.45 ± 0.07		0.40 - 0.50	
Protein (g/100 g)	QCAV-4	6	0.73 ± 0.10	0.0125*	0.60 - **0.90**	0.83
	non-GM	2	0.45 ± 0.07		0.40 - 0.50	
Ash (g/100 g)	QCAV-4	6	2.25 ± 0.79	0.5388	1.70 - 3.80	1.31
	non-GM	2	2.75 ± 1.48		1.70 - 3.80	
Carbohydrates (g/100 g)	QCAV-4	6	5.8 ± 1.5	0.9120	5.0 - 8.0	8.7
	non-GM	2	6.0 ± 2.8		4.0 - 8.0	
Energy (kJ/100 g)	QCAV-4	6	128 ± 26	0.9009	90 - 160	NA
	non-GM	2	125 ± 49		90 - 160	
Magnesium (mg/kg)	QCAV-4	6	125 ± 12	0.0781	**110** - 140	140
	non-GM	2	145 ± 7		140 - 150	
Manganese (mg/kg)	QCAV-4	6	2.55 ± 1.22	0.6446	**1.10** - **4.00**	2.25
	non-GM	2	2.10 ± 0.57		1.70 - 2.50	
Potassium (mg/kg)	QCAV-4	6	8,147 ± 241	0.0721	7,840 - **8,520**	6,479
	non-GM	2	7,645 ± 431		7,340 - 7,950	
Ascorbic Acid (mg/100 g)	QCAV-4	6	0.90 ± 0.00	0.6036	0.90 - 0.90	NA
	non-GM	2	0.90 ± 0.00		0.90 - 0.90	
Pyridoxine (mg/100 g)	QCAV-4	6	0.10 ± 0.02	0.1144	0.08 - **0.12**	NA
	non-GM	2	0.08 ± 0.02		0.06 - 0.09	

[1]Independent samples t-Test, significant differences with control asserted at 95%*, 99%**and 99.9%***
[2]Combined literature range (CLR) difficult to establish due to the lack of quality references for banana peel, however peel data from Cavendish banana cv Grand Nain published by Emaga *et al.* (2007) was converted as a guide.
n = biological replicates; SD = standard deviation.
Mean values for QCAV-4 analytes wholly (value ± SD) outside the range of the non-GM mean values are shaded in green.
Value in bold extend outside the range of the non-GM control dataset.

10.4 Conclusions

In summary, while there were statistical differences in the levels of some of the analytes between QCAV-4 and the non-GM controls, the mean values for proximates, vitamins and minerals from fruit and peel were mostly within the compositional variation reported in the literature. Further, no consistent pattern indicated that expression of the *MamRGA2* and *npt*II transgenes impacted the nutritional composition of QCAV-4. Taken together and considering that bananas are not a major contributor to diets in Australia and that QCAV-4 is not intended to completely replace the current Cavendish banana production, it is unlikely that any differences observed would have a nutritional impact on consumers of QCAV-4. Therefore, QCAV-4 should be considered nutritionally equivalent to conventional Grand Nain banana for the levels of all proximates, vitamins and minerals reported.

10.5 Supporting documents

- SD10-1-[redacted]-Composition analysis summary.xlsm
- SD10-2-[redacted]-Composition analysis summary.pdf
- SD10-3-[redacted]-Reports and methods.pdf
- SD10-4-[redacted]-Methods.pdf

s 22

References

BAN-C-1 (2001) BANANA RIPENING GUIDE, COLOR INDEX NUMBERS FOR BANANA RIPENING. https://www.ams.usda.gov/sites/default/files/media/Bananas_Visual_Aid%5B1%5D.pdf

Beck, E., Ludwig, G., Auerswald, E. A., Reiss, B. and Schaller, H. (1982) Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. Gene 19, 327-336. https://doi.org/10.1016/0378-1119(82)90023-3.

Codex Alimentarius Commission (2003) Guideline for the conduct of food safety assessment of foods derived from the recombinant-DNA plants. Risk Analysis *CAC/GL 45*, 1-18.

Dale, J., James, A., Paul, J-Y., Khanna, H., Smith, M., Peraza-Echeverria, S., Garcia-Bastidas, F., Kema, G., Waterhouse, P., Mengersen, K. and Harding, R. (2017) Transgenic Cavendish bananas with resistance to Fusarium wilt tropical race 4. Nature Communications 8, 1496. https://doi.org/10.1038/s41467-017-01670-6.

Delaney, B., Astwood, J. D., Cunny, H., Conn, R. E., Herouet-Guicheney, C., MacIntosh, S., Meyer, L. S., Privalle, L., Gao, Y., Mattsson, J. and Levine, M. (2008) Evaluation of protein safety in the context of agricultural biotechnology. Food and Chemical Toxicology 46, S71-S97. https://doi.org/10.1016/j.fct.2008.01.045.

Emaga, T. H., Andrianaivo, R. H., Wathelet, B., Tchango, J. T. and Paquot, M. (2007) Effects of the stage of maturation and varieties on the chemical composition of banana and plantain peels. Food Chemistry 103, 590-600. https://doi.org/10.1016/j.foodchem.2006.09.006.

FSANZ (2022) Australian Food Composition Database. https://www.foodstandards.gov.au/science/monitoringnutrients/afcd/Pages/fooddetails.aspx?PFKID=F000262.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005) Protein identification and analysis tools on the ExPASy server. In: Walker, J. M. (eds) The proteomics protocols handbook, pp. 571-607. Totowa, NJ: Humana Press. https://doi.org/10.1385/1-59259-890-0:571. https://web.expasy.org/peptide_cutter/.

Hernandez-Garcia, C. M. and Finer, J. J. (2014) Identification and validation of promoters and *cis*-acting regulatory elements. Plant Science 217-218, 109-119. https://doi.org/10.1016/j.plantsci.2013.12.007.

Holligan, E., Cook, S., Poggio, M. and Rattray, D. (2017) Economic assessment of best management practices for banana growing, report to the Department of Environment and Heritage Protection through funding from the Reef Water Quality Science Program, RP140B Technical Report. Department of Agriculture and Fisheries (DAF) and the Department of Natural Resources and Mines (DNRM), Queensland.

Hort Innovation (2022) Australian Horticulture Statistics Handbook 2020/21. Fruit. https://www.horticulture.com.au/contentassets/a68c8934a8bf40b4becdc487bacdb60f/hort-innovation-ahsh-20-21-fruit.pdf

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227(5259):680-685. https://doi.org/10.1038/227680a0.

Latham, J. R., Wilson, A. K. and Steinbrecher, R. A. (2006) The mutational consequences of plant transformation. Journal of Biomedicine and Biotechnology 25376. https://doi.org/10.1155/JBB/2006/25376.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P. and Rombauts, S. (2002) PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Research 30, 325-327. https://doi.org/10.1093/nar/30.1.325. https://bioinformatics.psb.ugent.be/webtools/plantcare/html/.

Martínez, L., Caballero-Mellado, J., Orozco, J. and Martínez-Romero, E. (2003) Diazotrophic bacteria associated with banana (*Musa* spp.). Plant and Soil 257, 35-47. https://doi.org/10.1023/A:1026283311770.

NRVANZ (Nutrient Reference Values for Australia and New Zealand) including recommended dietary intakes (2006) NHMRC, Commonwealth of Australia. https://www.eatforhealth.gov.au/nutrient-reference-values/nutrients/protein. https://www.eatforhealth.gov.au/nutrient-reference-values/nutrients/fats-total-fat-fatty-acids.

OGTR (2023) The Biology of *Musa* L. (banana). Version 3. OGTR website. https://www.ogtr.gov.au/resources/publications/biology-musa-l-banana

OGTR (2017) Risk Assessment Reference: Marker Genes in GM Plants. https://www.ogtr.gov.au/sites/default/files/files/202106/risk_assessment_reference_marker_genes_in_gm_plants.pdf

Porto, M. S., Pinheiro, M. P. N., Batista, V. G. L., dos Santos, R. C., de Albuquerque Melo Filho, P. and de Lima, L. M. (2014) Plant promoters: an approach of structure and function. Molecular Biotechnology 56, 38-49. https://doi.org/10.1007/s12033-013-9713-1.

Salamov A. A. and Solovyev V. V. (1997) Recognition of 3'-processing sites of human mRNA precursors. CABIOS 13, 1, 23-28. https://doi.org/10.1093/bioinformatics/13.1.23.

Sambrook, J. and Russell, D. W. (2001) *Molecular Cloning: A laboratory manual*. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

Solovyev, V. V., Shahmuradov, I. A., and Salamov, A. A. (2010) Identification of promoter regions and regulatory sites. In: Ladunga, I. (eds) Computational biology of transcription factor binding. Methods in Molecular Biology, vol 674, pp. 57-83. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-60761-854-6_5.

Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98, 503-517. https://doi.org/10.1016/S0022-2836(75)80083-0.

Thomas, K., Aalbers, M., Bannon, G. A., Bartels, M., Dearman, R. J., Esdaile, D. J., Fu, T. J., Glatt, C. M., Hadfield, N., Hatzos, C., Hefle, S. L., Heylings, J. R., Goodman, R. E., Henry, B., Herouet, C., Holsapple, M., Ladics, G.S., Landry, T. D., MacIntosh, S. C., Rice, E. A., Privalle, L. S., Steiner, H. Y., Teshima, R., van Ree, R., Woolhiser, M. and Zawodny, J. (2004) A multi-laboratory evaluation of a common *in vitro* pepsin digestion assay protocol used in assessing the safety of novel proteins. Regulatory Toxicology and Pharmacology 39, 87-98. https://doi.org/10.1016/j.yrtph.2003.11.003.

USDA (2019) FoodData Central. https://fdc.nal.usda.gov/fdc-app.html#/?query=banana.