

Application to Amend the Food Standards Code – New Food Produced Using Gene Technology

The Purple Tomato™
OECD Unique Identifier NPS-01201-8; Tomato event *Del/Ros1-N*

Submitting Company:
Norfolk Healthy Produce, Inc. (a wholly owned subsidiary of Norfolk Plant Sciences, Ltd.)

©2025 Norfolk Healthy Produce, all rights reserved

This document is protected under copyright law and is for use only by Food Standards Australia New Zealand in support of the actions requested by Norfolk Healthy Produce. Any other use of this material, without prior written approval of Norfolk Healthy Produce is strictly prohibited. By submitting this document, Norfolk Healthy Produce does not grant any party or entity any right or license to the information or intellectual property described herein.

Table of Contents

APPLICATION DETAILS	5
PURPOSE OF THE APPLICATION	5
JUSTIFICATION FOR THE APPLICATION	5
Costs and benefits of the application	6
Impact on international trade	6
INFORMATION TO SUPPORT THE APPLICATION	6
ASSESSMENT PROCEDURE	6
CONFIDENTIAL COMMERCIAL INFORMATION	6
OTHER CONFIDENTIAL INFORMATION	6
EXCLUSIVE CAPTURABLE COMMERCIAL BENEFIT (ECCB)	7
INTERNATIONAL AND OTHER NATIONAL STANDARDS	7
International standards	7
Other national standards or regulations	7
STATUTORY DECLARATION	8
CHECKLISTS	9
TECHNICAL INFORMATION ON THE FOOD PRODUCED USING GENE TECHNOLOGY	11
Nature and identity of the genetically modified food	11
History of use of the host and donor organisms	11

The nature of the genetic modification	12
CHARACTERISATION AND SAFETY ASSESSMENT OF NEW SUBSTANCES	17
New proteins	17
Other (non-protein) new substances	21
Compositional analyses of the food produced using gene technology	28
INFORMATION RELATED TO THE NUTRITIONAL IMPACT OF THE FOOD PRODUCED USING GENE TECHNOLOGY	33
REFERENCES	37
APPENDIX 1: COMPLETE SEQUENCE OF THE TRANSGENIC INSERT AND FLANKING GENOMIC REGIONS	45
APPENDIX 2: INTEGRITY OF T-DNA INSERT IN DEL/ROS1-N EVENT; ABSENCE OF VECTOR BACKBONE SEQUENCE	51
APPENDIX 3: ANALYSIS OF NEXT-GENERATION WHOLE GENOME SEQUENCING DATA FROM A COMMERCIAL LINE OF THE PURPLE TOMATO™	56
APPENDIX 4: FURTHER ANALYSIS OF INSERT STABILITY: T-DNA FLANKING SEQUENCES IN DEL/ROS1-N STOCK LINES DERIVED BY SELF-FERTILIZATION OVER >6 GENERATIONS	66
APPENDIX 5: SOUTHERN BLOT AND PCR ANALYSIS OF INSERT COPY NUMBER AND PRESENCE OF VECTOR BACKBONE	86
APPENDIX 6: BIOINFORMATICS ANALYSIS OF ORFS	109
APPENDIX 7: LC/MS ANALYSIS OF PROTEINS IN PURPLE TOMATO	131
APPENDIX 8: SEARCH OF DELILA AND ROSEA1 PROTEIN SEQUENCES AGAINST ALL TOXIN SEQUENCES IN UNIPROT	134
APPENDIX 9: <i>IN SILICO</i> ANALYSES AND PROTEASE DIGESTION STUDIES TO ASSESS THE LIKELY ALLERGENICITY OF DELILA AND ROSEA 1 PROTEINS	162
APPENDIX 10: FEEDING STUDY TO ASSESS EFFECT OF PURPLE TOMATOES IN DIET ON LABORATORY MICE	180

Application details

Applicant's name: Norfolk Healthy Produce, Inc., a wholly owned subsidiary of Norfolk Plant Sciences, Ltd.

Name of contact person: Eric Ward

Address: 2800 5th St., Suite 110, Davis, California 95618, USA

Telephone: [REDACTED]

Email: e.ward@norfolkhealthyproduce.com

Nature of the applicant's business: Norfolk Healthy Produce (NHP) develops and commercializes specialty produce varieties derived from novel biotechnology traits. Our first product is The Purple Tomato™, which is enriched throughout the fruit in the purple pigments called anthocyanins. NHP breeds tomato varieties containing this trait for both fresh market and processed food products. NHP works with growers and distributors on local, regional, and national levels to market products based on its technology.

Details of other individuals or organizations associated with the application: All Aussie Avocados, Pty., trading as All Aussie Farmers (Travis Murphy, Managing Director) will hold the DIR license (if and when granted) from the Gene Technology Regulator for The Purple Tomato™.

Purpose of the application

This application seeks to amend the Australia New Zealand Food Standards Code to allow for the inclusion of tomato event *Del/Ros1-N* (The Purple Tomato™) in Standard 1.5.2-Food Produced Using Gene Technology.

Justification for the application

The Purple Tomato™ is a new food produced using gene technology, and to be made available to consumers in Australia (and New Zealand), would require amendment of the Code.

To our knowledge, no other tomato varieties produced using gene technology have previously been applied for or approved by FSANZ.

The Purple Tomato™ is a source of the purple pigments called anthocyanins that are present in purple fruits and vegetables such as blueberries, blackberries, and purple potatoes. The presence of these compounds in the diet is associated with numerous positive health outcomes, and consumers are looking for ways to supplement their diets with purple foods. The Purple Tomato™ is a novel, savory source of these compounds, in a produce item that is widely consumed in Australia and New Zealand. In addition, The Purple Tomato™ is visually striking, delighting consumers when they see a tomato cut open that is dark purple throughout. The data presented in this application and conclusions that can be drawn from them, indicate no disadvantages to making this product available.

Costs and benefits of the application

The availability of The Purple Tomato™ to consumers will be a benefit from the standpoint of providing additional choice of a healthy produce item to include in their diet. Norfolk Healthy Produce nor its partners intend to make explicit health claims for The Purple Tomato™ beyond referencing studies in the scientific literature that describe the benefits of consuming anthocyanins. In addition, the ability to grow and market The Purple Tomato™ will provide Australian produce growers with an additional, high-value, novel item that consumers will seek out. We anticipate no additional regulatory cost (beyond review of this application), given that The Purple Tomato™ is easily and uniquely identifiable by its dark purple color throughout the fruit, resulting in inherent “identity preservation” that makes inadvertent consumption of the product impossible.

Impact on international trade

We anticipate no impact. Fruit of The Purple Tomato™ can only be legally sold and consumed in countries for which regulatory approval has been granted; to date, this is only the United States (with an application with Health Canada co-pending this current application). We are not aware of any export of fresh market tomatoes from Australia to either the US or Canada. Eventually, processed food products produced in Australia could be exported to other countries in which regulatory approval has been granted, but Norfolk Healthy Produce has no current plans for such.

Information to support the application

The detailed information required for an application for a new food produced using gene technology (Handbook section 3.5.1) appears in the sections below, identified by page in the Table of Contents (above).

Assessment procedure

Norfolk Healthy Produce considers it appropriate that FSANZ adopt the general procedure in assessing this application.

Confidential commercial information

This application does not contain confidential commercial information.

Other confidential information

This application does not contain other confidential information.

Release of Information:

Norfolk Healthy Produce is submitting the information in this application for review by Food Standards Australia New Zealand (FSANZ) for amendment to the Food Standard 1.5.2 - Food Produced Using Gene Technology. Norfolk Healthy Produce holds proprietary rights to the extent allowable by law to all such information and, by submitting this information, Norfolk Healthy Produce does not authorise its release to any third party except to the extent it is duly requested under the Freedom of Information Act 1982 (FOI Act) or in compliance with the responsibility of FSANZ to publish documents required under Sections 8, 8(A), 8(C) and 8(D) of the FOI Act; and this information is responsive to the specific aforementioned request. Accordingly, except as specifically stated above, Norfolk Healthy Produce does not authorise the release, publication or other distribution of this information (including website posting or otherwise), nor does Norfolk Healthy Produce authorise any third party to use, obtain, or rely upon this information, directly or indirectly, as part of any other application or for any other use, without Norfolk Healthy Produce's prior notice and written

consent. Submission of this information does not in any way waive Norfolk Healthy Produce's rights (including rights to exclusivity and compensation) to such information.

Exclusive capturable commercial benefit (ECCB)

Norfolk Healthy Produce intends to commercialize The Purple Tomato™, and products derived from it, in Australia in partnership with various growers, packers, shippers, distributors, and manufacturers. We initially plan to work with All Aussie Farmers to contract for the growing of and distribution of The Purple Tomato™. Norfolk Healthy Produce will realize a royalty stream from such partnerships.

International and other national standards

International standards

The Codex Alimentarius Commission Guideline for the Conduct of Food Safety Assessment of Foods Derived from Recombinant-DNA Plants (CAC/GL 45-2003) is applicable to the assessment of this application to amend the Australia New Zealand Food Standards Code to allow for the inclusion of event *Del/Ros1-N* in Standard 1.5.2 – Food Produced Using Gene Technology.

Other national standards or regulations

An application has been submitted to the Office of the Gene Technology Regulator for a license for a Dealing with Intentional Release (such license to be held by All Aussie Avocados, Pty., trading as All Aussie Farmers). In addition, review of this application is being carried out under the Shared Assessment Process with Health Canada, which agency is simultaneously reviewing The Purple Tomato™ for food safety in Canada. An application for environmental release has also been submitted to the Canadian Food Inspection Agency, to allow cultivation of The Purple Tomato™ in Canada without special permitting. In the US, US Department of Agriculture has determined that The Purple Tomato™ is not a regulated article¹, and the Food and Drug Administration has issued a letter confirming no questions regarding the safety of The Purple Tomato™ as a food². The Purple Tomato™ is now in its second season of commercial sales in the US, both as seeds and starter plants for home gardeners, and in partnership with local, regional, and national-scale growers, who provide fresh tomatoes through retail channels, including farmers markets and grocery chains.

¹ <https://www.aphis.usda.gov/sites/default/files/21-166-01rsr-review-response.pdf>

² <https://www.fda.gov/media/170056/download>

Statutory declaration

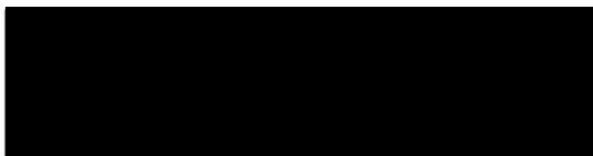
I, Eric R. Ward, Board Chair of Norfolk Healthy Produce, 2800 5th St., Suite 110, Davis, California, USA,

make the following declaration under the Statutory Declarations Act 1959:

1. the information provided in this application fully sets out the matters required
2. the information provided in this application is true to the best of my knowledge and belief
3. no information has been withheld that might prejudice this application, to the best of my knowledge and belief.

I understand that a person who intentionally makes a false statement in a statutory declaration is guilty of an offence under section 1 of the Statutory Declaration Act 1959, and I believe that the statements in this declaration are true in every particular.

Signed by me this 14th day of May, 2025:



Eric R. Ward

Durham County, North Carolina

I certify that the following person personally appeared before me this day, acknowledging to me that he signed the foregoing document: Eric R. Ward

Date: 05/14/25

(Official seal)



Official Signature of Notary



VANDANA DHINGRA, Notary Public

Notary's printed or typed name

My commission expires: 30 September 2029.

Checklist for General requirements

This Checklist will assist you in determining if you have met the mandatory format and information requirements as detailed in Guideline 3.1.1 General requirements. All applications **must** include this Checklist.

General requirements (3.1.1)		
Check	Page No.	Mandatory requirements
		A Form of application ☐ <i>Application in English</i> ☐ <i>Executive Summary (separated from main application electronically)</i> ☐ <i>Relevant sections of Part 3 clearly identified</i> ☐ <i>Pages sequentially numbered</i> ☐ <i>Electronic copy (searchable)</i> ☐ <i>All references provided</i> B Applicant details C Purpose of the application D Justification for the application ☐ <i>Regulatory impact information</i> ☐ <i>Impact on international trade</i> E Information to support the application ☐ <i>Data requirements</i> F Assessment procedure ☐ <i>General</i> ☐ <i>Major</i> ☐ <i>Minor</i> ☐ <i>High level health claim variation</i> G Confidential commercial information ☐ <i>CCI material separated from other application material</i> ☐ <i>Formal request including reasons</i> ☐ <i>Non-confidential summary provided</i> H Other confidential information ☐ <i>Confidential material separated from other application material</i> ☐ <i>Formal request including reasons</i> I Exclusive Capturable Commercial Benefit ☐ <i>Justification provided</i> J International and other national standards ☐ <i>International standards</i> ☐ <i>Other national standards</i> K Statutory Declaration L Checklist/s provided with application ☐ <i>3.1.1 Checklist</i> ☐ <i>All page number references from application included</i> ☐ <i>Any other relevant checklists for Chapters 3.2–3.7</i>

Checklist for applications for new foods

This Checklist is in addition to the Checklist for Guideline 3.1.1 and will assist you in determining if you have met the information requirements as specified in Guidelines 3.5.1 3.5.3.

Foods produced using gene technology (3.5.1)		
Check	Page No.	Mandatory requirements
		- A.1 Nature and identity - A.2 History of use of host and donor organisms - A.3 Nature of genetic modification - B.1 Characterisation and safety assessment - B.2 New proteins - B.3 Other (non-protein) new substances - B.4 Novel herbicide metabolites in GM herbicide-tolerant plants - B.5 Compositional analyses - C Nutritional impact of GM food - D Other information
Novel foods (3.5.2)		
Check	Page No.	Mandatory requirements
		A. Exclusive use - B.1 Type of novel food - B.2 Information on potential beneficial outcomes - B.3 Chemical and physical properties - B.4 Impurity profile - B.5 Manufacturing process - B.6 Specification for identity and purity - B.7 Analytical detection method C.1 Plant or animal extracts - C.1.1 Extraction and composition - C.1.2 Effects of food processing or preparation - C.1.3 Current use - C.1.4 Potential adverse effects C.2 Plant and animal extracts - C.2.1 Method or extraction and composition of extract - C.2.2 Use as a food in other countries - C.2.3 Toxicity studies - C.2.4 Safety assessments from other agencies C.3 Herbs (both non-culinary and culinary) including extracts - C.3.1.1 History of use

Technical information on the food produced using gene technology

Nature and identity of the genetically modified food

Norfolk Plant Sciences has introduced into tomato the *Delila* (*Del*) and *Rosea1* (*Ros1*) genes derived from the common garden snapdragon (*Antirrhinum majus*). Both genes are expressed under the tomato fruit specific promoter E8. The *Del* gene encodes a basic helix-loop-helix transcription factor and the *Ros1* gene encodes an R2R3MYB-related transcription factor (Schwinn *et al.*, 2006; Goodrich *et al.*, 1992; Butelli *et al.*, 2008). The *Del/Ros1*-N event also contains an NPTII selectable transformation marker. The *Del* and *Ros1* genes encode transcription activators of the endogenous tomato anthocyanin biosynthetic pathway, resulting in increased production of the two major anthocyanins in tomato fruit. The NPTII selectable transformation marker confers resistance to kanamycin, used during the transformation process. The event that is subject of this application is designated as *Del/Ros1*-N, and has been given OECD unique identifier NPS-01201-8. We anticipate that The Purple Tomato™ will be marketed under various trade names in Australia, including Purple Bliss. In the US, The Purple Tomato™ is marketed in partnership with Red Sun Farms under the Empress brand name.

History of use of the host and donor organisms

The host organism is the cultivated tomato (*Solanum lycopersicum* L.), previously known as *Lycopersicon esculentum*. Tomato is native to Central America and western South America. The people native to these regions cultivated tomato and seeds were transported to Europe by Spanish explorers in the 1600s. Later, tomatoes were introduced in North America by European colonists (Rick, 1978). Hybrid varieties of tomato have been developed and are generated by manual crossing techniques. Tomato is not considered to be a weedy species (USDA, 1992) and it is not sexually compatible with related species within the family Solanaceae.

Tomato is a diploid species with 12 chromosome pairs and the genetics of tomato is well characterized (Tanksley, 1993; OECD, 2016). All commercial tomato varieties are self-compatible and inbreeding. Modern tomato varieties were selected for high fertility, which resulted in stigmas that are recessed in the anther cone and are therefore not able to receive pollen from outside the flower. As a result, cross pollination in the field is low, with self-pollination occurring at the rate of 99% or greater (Currence and Jenkins, 1942).

Tomatoes are widely consumed in Australia; indeed, the ANZ region, at 17kg/person annually, ranks second only to North America in per capita tomato consumption³. The fruit are the component of the plant that is eaten, both in fresh form and in various processed products, including sauces, pastes, and juices. No special handling or preparation of The Purple Tomato™ is required to consume it safely; it can be handled like any other tomato.

The *Del* and *Ros1* genes are derived from the snapdragon plant, whose flowers are edible and commonly used in summer salads and canapés. Both the *Del* and *Ros1* genes are expressed

³ https://www.tomatonews.com/en/world-consumption-study-of-tomato-products_2_2333.html

in the snapdragon flower (Goodrich et al. 1992.; Schwinn et al. 2016). Snapdragon flowers are widely acknowledged to be edible and non-toxic⁴.

The other expressed gene in the DNA introduced is the NPTII selectable marker, derived from *E. coli* Tn5. This protein has been previously assessed as safe by FSANZ. Additional genetic elements in the introduced DNA construct derive from *Agrobacterium tumefaciens* and Cauliflower Mosaic Virus. These additional elements have previously been assessed by FSANZ. A complete listing of the DNA sequence elements appears below.

The nature of the genetic modification

The transformation procedures and results of this work have been reported in detail (Butelli *et al.*, 2008). *Agrobacterium* transformation of tomato was performed essentially as described by McCormick (1991). Briefly, tomato leaf discs of the MicroTom variety were transformed using *Agrobacterium tumefaciens* strain LBA4404 containing the binary vector pDEL.ROS containing the *Del* and *Ros1* genes whose expression is under the control of the fruit-specific E8 promoter from tomato. In addition to the *Del* and *Ros1* genes, the *nptII* gene derived from the bacterial transposon Tn5 was included within the T-DNA of plasmid pDEL.ROS to provide a means of selection of the tomato transformants. No tumor-inducing genes are present on the plasmid. Plants regenerated from tissue culture were free of *Agrobacterium* contamination. Four primary transformants (T0 generation) carrying the pDEL.ROS vector were generated (lines C, N, Y and Z). These were analyzed by Southern blot to determine the T-DNA copy number using the *Ros1* cDNA as a probe. The event expressing the highest levels of anthocyanins, event N, was chosen for further breeding and analysis.

Individual plants in the T1 generation were used for copy number determination by qPCR. One individual plant in the T1 generation of the *Del/Ros1*-N event plant was crossed to MoneyMaker to generate a purple fruited F1 population. One of the F1 plants was selfed and then a dark purple F2 plant was selfed through 8 generations propagated by single seed descent to develop the stock line “*Del/Ros1*-N in MoneyMaker”. The same T1 *Del/Ros1*-N plant that was crossed to MoneyMaker was also selfed across 6 generations using single seed descent to generate the stock line “*Del/Ros1*-N in MicroTom”.

Plants from these stock lines were used for initial characterization of the T-DNA insertion and segregation analyses. Compositional analyses were carried out on fruit from the MoneyMaker F9 plants. Additional molecular characterization was carried out on the commercial breeding line derived from MoneyMaker F9. This commercial line resulted from a cross of *Del/Ros1*-N in MoneyMaker x Goldkrone (a small-fruited yellow tomato variety), followed by multiple generations of selfing and selection for fruit color and truss architecture.

⁴ <https://gardenerspath.com/plants/flowers/edible-snapdragons/>
<https://www.gardeningknowhow.com/ornamental/flowers/snapdragon/are-snapdragons-edible.htm>
<https://plants.ces.ncsu.edu/plants/antirrhinum-majus/>
<https://plantaddicts.com/are-snapdragons-poisonous/?srsltid=AfmBOopCM4TAuDDtpd1hKyAMMIA0l6nVWk8cP35eRCiJhYWRc5GLdvJ>
<https://floristsreview.com/a-florist-guide-to-edible-flowers/>
https://discover.texasrealfood.com/safe-bites/is-it-safe-to-eat-snapdragons#google_vignette
<https://hautebitesfood.com/2014/04/09/crystalized-edible-snapdragon-flowers/>

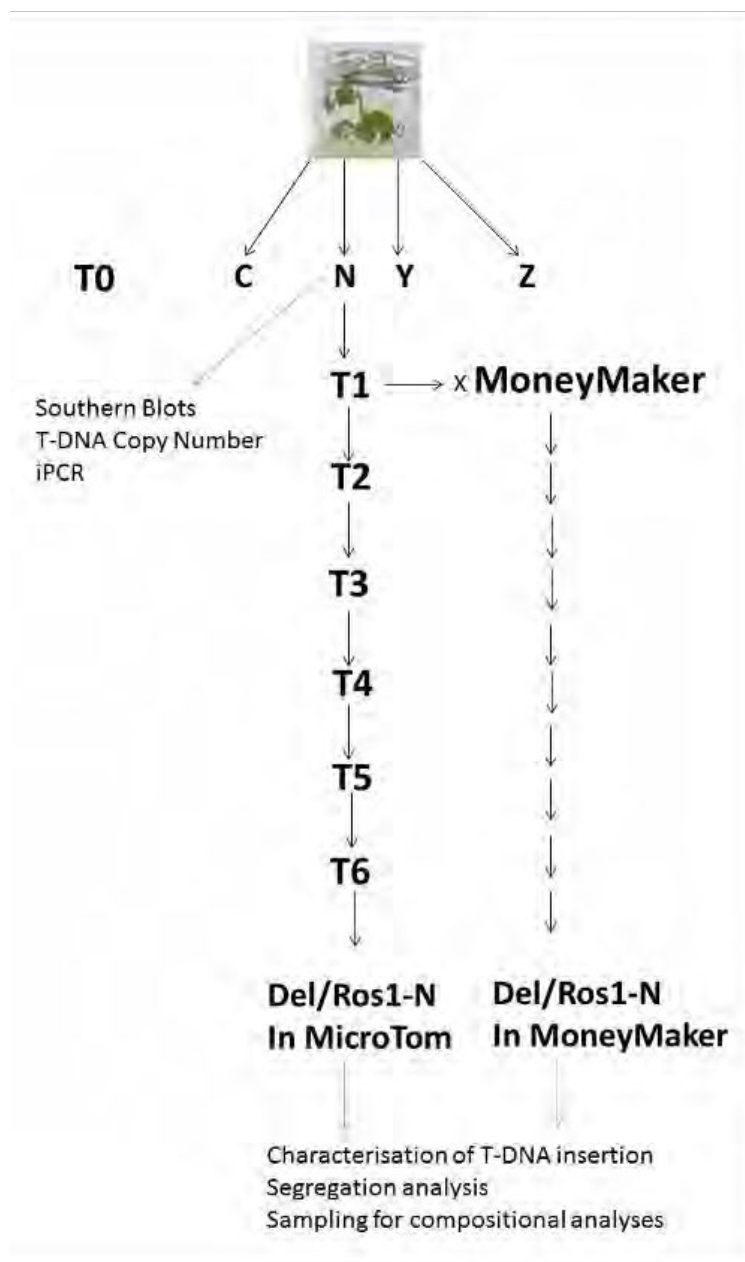


Figure 1. Pedigree diagram showing derivation of the T6 MicroTom and F9 MoneyMaker lines. In addition, the current commercial breeding line is the result of a cross of MoneyMaker F9 x Goldkrone, with subsequent selection of selfed progeny for purple fruit color and high fruit count per truss. The material used for commercial breeding and for molecular characterization by whole genome sequencing is >7 generations of selfing from the MoneyMaker x Goldkrone F1.

Table 1 lists the components of the transformation vector. The complete sequence of pDEL.ROS is deposited in GenBank as accession number MN580094.1.

Table 1. Description of the genetic elements in the 33,236 bp transformation construct pDEL.ROS (NCBI Accession Number MN580094.1).

Genetic Element	Description	Donor Organism	Function	Reference
LB	Left T-DNA Border	<i>Agrobacterium tumefaciens</i>	T-DNA boundary	Zambryski <i>et al.</i> , 1982
NOSp	Nopaline synthase promoter region	<i>Agrobacterium tumefaciens</i>	Transcriptional Promoter	Jones <i>et al.</i> , 1992
NPT II	Neomycin phosphotransferase gene	Bacterial transposon Tn5 from <i>Escherichia coli</i>	Selectable marker gene conferring resistance to kamamycin	van den Elzen <i>et al.</i> , 1985
Ocs 3	Octopine synthase termination region	<i>Agrobacterium tumefaciens</i>	Transcriptional terminator/ poly-adenylation signal	Jones <i>et al.</i> , 1992
E8p	E8 promoter	<i>Solanum lycopersicum</i>	Transcriptional promoter activated in the fruit upon ripening	Butelli <i>et al.</i> , 2008
DEL	<i>Delila</i> cDNA from snapdragon	<i>Antirrhinum majus</i>	Transcription factor gene coding sequence	Goodrich <i>et al.</i> , 1992
CMV	Cauliflower mosaic virus termination region	Cauliflower mosaic virus	Transcriptional terminator/ poly-adenylation signal	Hellens <i>et al.</i> , 2000
E8p	E8 promoter	<i>Solanum lycopersicum</i>	Transcriptional promoter activated in the fruit upon ripening	Butelli <i>et al.</i> , 2008
ROS	<i>Rosea1</i> cDNA	<i>Antirrhinum majus</i>	Transcription factor gene coding sequence	Schwinn <i>et al.</i> , 2006
CMV	Cauliflower mosaic virus termination region	Cauliflower mosaic virus	Transcriptional terminator/ poly-adenylation signal	Hellens <i>et al.</i> , 2000
RB	Right T-DNA Border	<i>Agrobacterium tumefaciens</i>	T-DNA boundary	Wang <i>et al.</i> , 1984
TETr	Tetracycline resistance gene	Bacterial plasmid pRK290	Bacterial selectable marker gene	Comai <i>et al.</i> , 1983
trfA	Replication function from pRK2	Bacterial plasmid pRK2 (NCBI Accession No. NC005970.1)	Plasmid origin of replication	Ditta <i>et al.</i> , 1980
rix	Cis-acting function for conjugal mobilization from pRK2	Bacterial plasmid pRK2	Conjugal mobilization of the plasmid	Ditta <i>et al.</i> , 1980
trfB	Replication function from pRK2	Bacterial plasmid pRK2	Plasmid replication	Ditta <i>et al.</i> , 1980
oriV	Origin of replication from pRK2	Bacterial plasmid pRK2	Plasmid origin of replication	Ditta <i>et al.</i> , 1980

A detailed map of the vector appears below as Figure 2.

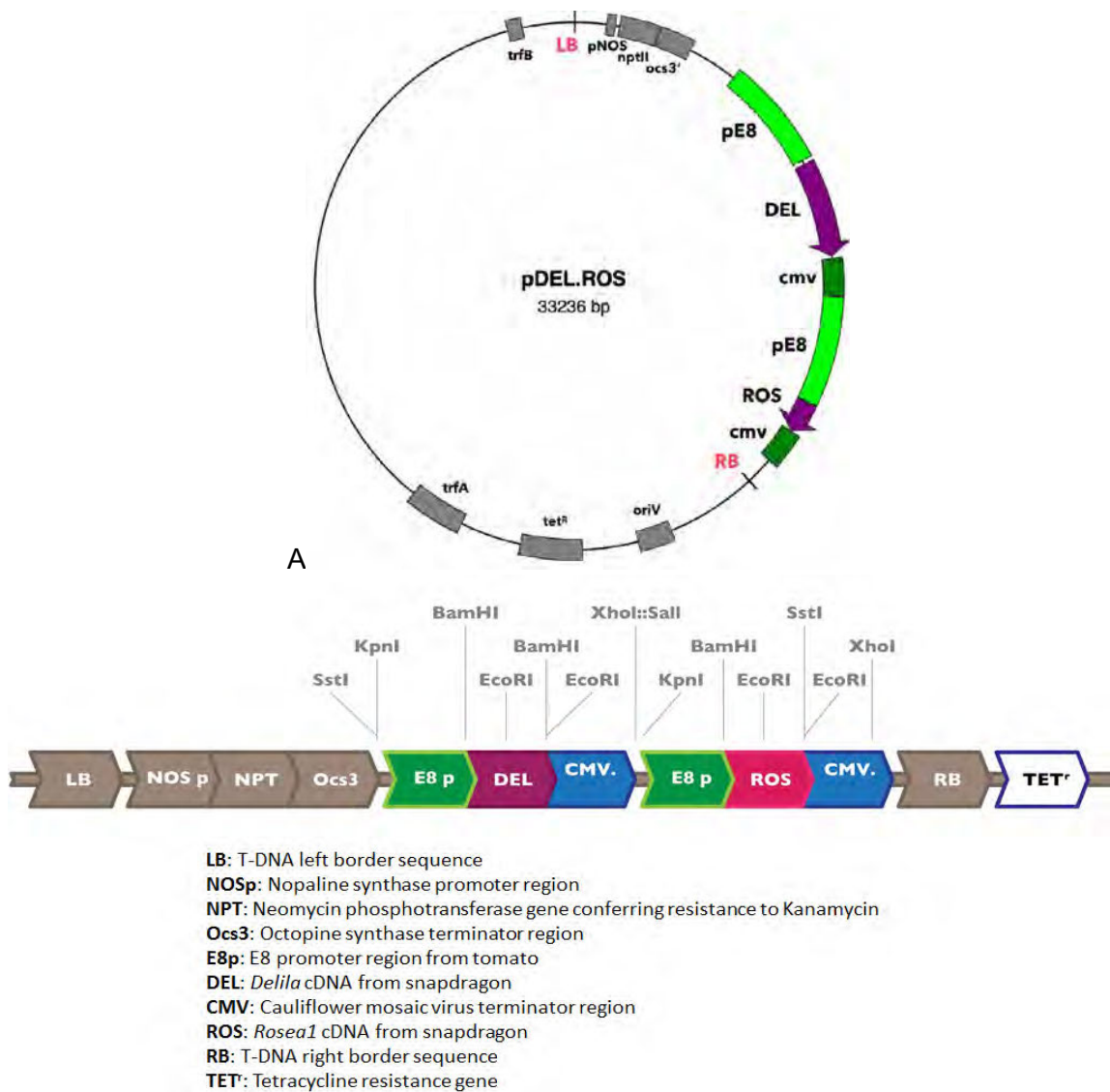


Figure 2. a) pDEL/ROS was constructed by inserting the *Del/Ros1* cassette shown at the top of the figure into bacterial plasmid pRK290. B) Molecular map of the pDEL.ROS T-DNA insert containing the *Del* and *Ros1* genes. Locations of some cleavage sites of restriction enzymes are shown.

The complete sequence of the pDEL.ROS insert present in The Purple Tomato™, with surrounding tomato genomic DNA, appears in Appendix 1.

Complete genome resequencing of a single *Del/Ros1*-N plant (MicroTom generation >6) confirmed that the T-DNA portion of the vector was inserted at a single locus with 100% match to the expected sequence with no mutations or mismatches. Short deletions occurred at each end of the inserted portion of the vector, not affecting the function of the inserted genes, and a 94 bp deletion of the tomato genome occurred at the site of insertion. This search also confirmed that no sequences matching the backbone of the vector are present in *Del/Ros1*-N lines (see Appendix 2).

In addition, a whole genome resequencing of the commercial breeding line MoneyMaker x Goldkrone F7+ was carried out to 40X depth, to confirm that: i) the commercial line carried a single T-DNA insert in the same position as the locus characterized by whole genome sequencing of the MicroTom T6 line, and ii) no vector backbone DNA was present. This information and the methodology for the analysis appears in appendix 3.

A schematic diagram of the T-DNA insertion in the genome of The Purple Tomato™ is depicted in Figure 3.

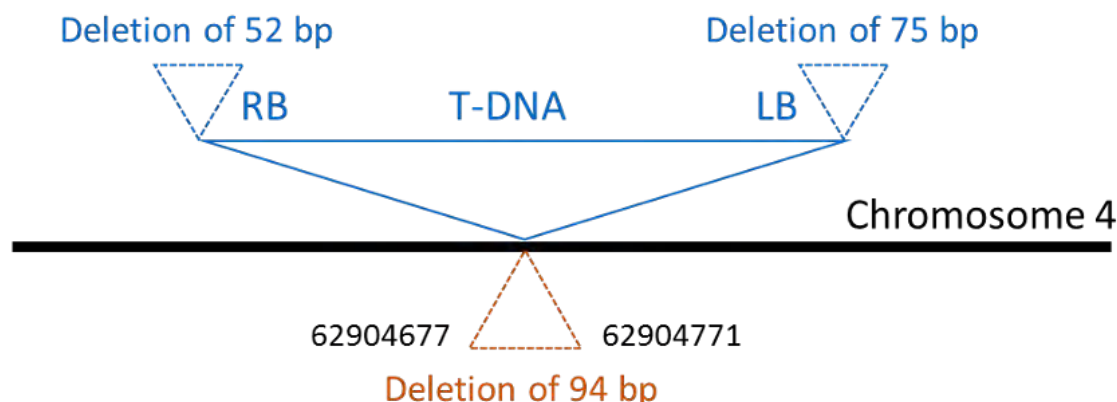


Figure 3. Schematic representation of the *Del/Ros1* T-DNA insert in the chromosome of the Purple Tomato. The complete sequence of the inserted DNA with flanking tomato genomic DNA appears in Appendix 1. Note that nucleotide coordinates indicating the position of the insert in the tomato genome have changed in the latest release of the reference tomato genome (see details in Appendix 3).

Stability of the single DNA insert was monitored and confirmed across multiple generations by molecular analysis (see Appendix 4). In addition, the purple trait has been introgressed from *Del/Ros1-N* into other tomato varieties such as “Ailsa Craig”, “Lucinda”, and VF36. The purple phenotype was inherited in a Mendelian fashion in all tomato genetic backgrounds, with this phenotypic segregation confirming the presence of the single, stable locus. For example, data for the F₂ of *Del/Ros1-N* crossed to VF36 are shown in Table 2.

Table 2: Genotypic and phenotypic segregation of *nptII*, kanamycin resistance and purple phenotype in F₂ generation of *Del/Ros1-N* (in MoneyMaker background F9) x VF36 cross. Analysis was carried out by germinating seeds on agar containing kanamycin. A control plate containing no kanamycin was used to confirm that germination was sufficiently high (>95%) to give a useful result. Chi-squared value = 0.267 with 1 degree of freedom. Two-tailed P value = 0.61. Therefore the observed results are not statistically significantly different from those expected for a 3:1 segregation ratio, consistent with a single genetic locus being present in MoneyMaker F9 *Del/Ros1-N* plants and progeny used for further development.

Presence of <i>nptII</i> gene		Kanamycin resistance		Color of tomato fruit	
Present	16	Resistant	16	Purple	16
Absent	4	Sensitive	4	Red	4

Purple Tomato lines now in commercial production in the US are >10 generations removed from the original transformation event, and no variation in penetrance of the trait has been observed within varieties among 1000s of plants grown.

Note that the parental T1 line that gave rise the final *Del/Ros1*-N event (present in MicroTom T6, MoneyMaker F9, and the commercial lines derived from MoneyMaker F9) contained two insertions of the T-DNA. These segregated from each other during subsequent selfing and outcrossing, and only the single locus (referred to as “locus B” in the initial analyses) remains in these advanced lines. The initial molecular evidence supporting the presence of a single locus and absence of vector backbone DNA in the MicroTom line is summarized in Appendix 5 for completeness, but is superseded and unequivocally corroborated by whole genome resequencing of a commercial line (described in Appendix 3).

The tomato genome and the junctions of T-DNA with the genome were analyzed for the interruption of any native tomato genes, or the creation of new open reading frames resulting from the T-DNA insertion. From these analyses we concluded that no endogenous tomato genes were interrupted. Only one putative new open reading frame was created. No potential matches to any known proteins, nor any matches to allergens, nor evidence of expression, were found for the putative new ORF. The results of these analyses appear in Appendix 6.

In conclusion, *Del/Ros1*-N tomato contains only a single, stable, well-characterized transgenic insert, identical to the intended sequence, which disrupts no existing tomato genes, nor creates any new expressed open reading frames.

Characterisation and safety assessment of new substances

Both *Del* and *Ros1* genes encode transcription factors that control anthocyanin biosynthesis in plants. They are closely related structurally to two endogenous tomato genes, Hoffman’s anthocyaninless and *ANT1*, which control anthocyanin production in tomato leaves (Mathews et al., 2003; Qiu et al., 2016). Transcription factors function in the nucleus and are not glycosylated since glycosylation would involve transfer to the endoplasmic reticulum where they could not function. No enzymatic function has ever been associated with either protein; rather, their sole function is to activate transcription of their target genes (Goodrich et al., 1992; Schwinn et al., 2006; Butelli et al., 2008). Figure 4 shows the upregulation of enzymes in the anthocyanin biosynthesis pathway resulting from the expression of *Del* and *Ros1* in ripening fruit.

New proteins

To confirm that no unexpected proteins were produced in the Purple Tomato, and to assess the amount of the DEL and ROS1 proteins in the fruit, total protein extracts were prepared (Faurobert et al., 2007) and analyzed using an Orbitrap LC-MS. Over 2,000 proteins were detected but neither DEL nor ROS1 transcription factor proteins were detected, nor any NPTII plant selectable marker protein. Consistent with the bioinformatic analysis indication that no potential fusion proteins were created as a result of the T-DNA insertion (see above, Appendix 3), no novel peptides were detected in the LC-MS analysis.

The limit of detection of the LC-MS analysis was calculated as less than 7 fmoles each of the DEL and ROS1 proteins per milliliter (mL) of extract. This corresponds to less than 0.2 ng ROS1 protein and less than 0.5 ng of DEL protein per mL, or the following values for Limit of Detection on a fresh weight basis: DELILA: <4ng/gfw; ROSEA1: <1.5ng/gfw; NPTII: <1.7ng/gfw. For details of the LC-MS analysis, see Appendix 7.

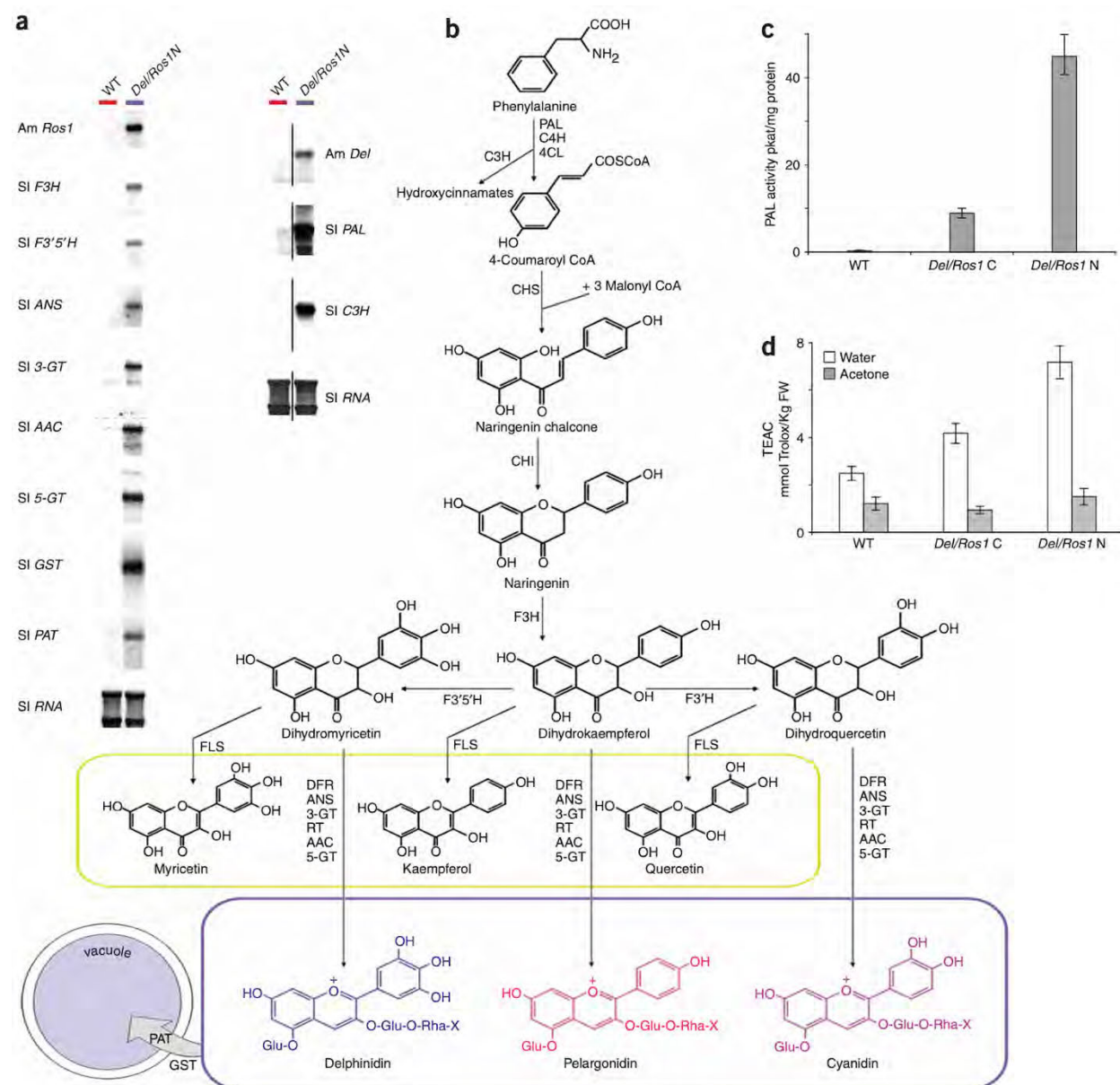


Figure 4. Expression of *Del* and *Ros1* causes the upregulation of genes required for anthocyanin biosynthesis and results in increased PAL activity and higher total antioxidant capacity. (A) Northern blots showing the differential expression of several anthocyanin biosynthetic genes identified by suppression subtractive hybridization. (B) Schematic representation of the anthocyanin biosynthetic pathway. Flavonoid classes relevant to this article are shown in boxes. Yellow box, flavonols; purple box, anthocyanins; PAL, phenylalanine ammonia lyase; 4CL, 4-coumarate:coenzyme A ligase; C4H, cinnamate 4-hydroxylase; C3H, 4-coumarate 3-hydroxylase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone-

3-hydroxylase; F3 ϕ H, flavonoid-3'-hydroxylase; F3'5'H, flavonoid-3'5'-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol reductase; ANS, anthocyanidin synthase; 3-GT, flavonoid 3-O-glucosyltransferase; RT, flavonoid 3-O-glucoside-rhamnosyltransferase AAC, anthocyanin acyltransferase; 5-GT, flavonoid-5-glucosyltransferase; GST, glutathione S-transferase; PAT, putative anthocyanin transporter. PAT and GST may be involved in transport of anthocyanins into the vacuole (purple circle). X, acyl group on the 3-glucoside moiety of the anthocyanin. (C) Analysis of specific PAL activity in ripe tomato fruit from wild-type and transgenic lines *Del/Ros1-C* and *Del/Ros1-N*. Histograms represent mean values $\pm$ s.d. of separate measurements (n = 3). Results were confirmed in two additional independent experiments. (D) Analysis of hydrophilic and lipophilic antioxidant activity in ripe tomato fruit from wild-type and transgenic lines *Del/Ros1-C* and *Del/Ros1-N*. Data represent mean values $\pm$ standard deviation and were derived from at least seven tomatoes per plant (taken from Butelli et al., 2008).

The protein expression analysis of The Purple Tomato was conducted on material from MoneyMaker generation F9 and later. We know that expression of the *Del* and *Ros1* transgenes is very weak, with protein detection being below the limit of mass spectrographic analysis. There is no evidence that the expression level would change in any tomato genetic background. The E8 promoter that drives both transgenes is the natural promoter of a hydroxylase enzyme involved in catabolizing α -tomatine during ripening. That protein, which has a catalytic activity and is not a structural protein, can only be detected by heterologous expression (Akiyama et al. 2021), indicating that the expression driven by the E8 promoter is quite weak. This is consistent with the undetectable level of expression seen for DEL and ROS1 proteins in The Purple Tomato™. We are developing additional varieties of tomatoes, introgressing the Purple Tomato trait from MoneyMaker F9 by traditional breeding⁵. There is no evidence to support that expression of these transgenes would be higher in other tomato genetic backgrounds.

Comparison of the DEL and ROS1 amino acid sequences to all toxin entries in the UniProt database yielded no significant similarities to any known toxin or anti-nutrient. The details of these analysis appear in Appendix 8.

The Rosea1 and Delila protein sequences exhibited no homology (>35%) to known allergens at the entire protein level or by looking for matches to 80 amino-acid peptides. In addition, no exact matches to 8-mers in known allergens were identified. The details of these results appear in Appendix 9.

The only matches to Delila and Rosea1 of any significance are other proteins of the same functional classes. This result is not surprising, given their high degree of similarity to other bHLH (Del) and Myb-type (Ros1) transcription factors present in many other food crops, as described in Albert et al. 2021, Liu et al., 2015, Goodrich et al. 1992, Butelli et al. 2012, Cao et al. 2020 (a review of MYB transcription factors as regulators of secondary metabolism in plants), Zhao et al. 2019 (expression of bHLH and MYB genes in ripening blueberry fruit), Aharoni et al. 2001 (strawberry), Deluc et al 2008 (grape), and Wang et al. 2017 (apple).

A standardized digestion protocol was carried out in a simulated gastric environment to monitor the disappearance of the DEL and ROS1 proteins over time (Thomas et al., 2004). Both DEL and ROS1 proteins were rapidly degraded by pepsin (< 1 min and <10 min respectively). The details of the methods used and results appear in Appendix 9.

⁵ MicroTom was used only as the variety for the original transformation, and no additional lines have been derived from MicroTom, nor will it or any progeny be commercialized.

We cannot state categorically that the DEL and ROS1 proteins expressed in *E. coli*, used for proteolytic degradation studies, are identical in conformation to the native proteins expressed in plants. This is because the levels of those proteins in the *Del/Ros1-N* event are below the detection limit of our exceedingly sensitive MS analysis (see above, and Appendix 7) and are therefore impossible to characterize in greater detail given the vanishingly small amounts of these regulatory proteins expressed. Moreover, these exact proteins are present in the edible flowers of *Antirrhinum majus* (snapdragon), from which the genes were derived, and closely related proteins are present through the plant kingdom, in many edible plants (see references above, and BLAST search results in Appendix 8).

The NPTII protein has previously been assessed for safety by FSANZ. The alignment below shows the NPTII sequence in pDEL.ROS used to transform the Purple Tomato (query) compared to the canonical NPTII from *E. coli* transposon Tn5 (subject). There is a single, conservative amino acid substitution (valine for isoleucine at position 2, the result of a A>G transition, a change made to create an *NcoI* restriction site during vector construction) in the version present in the Purple Tomato™ compared to the canonical sequence.

Query: NPTII [Plant binary expression vector pDEL.ROS] Query ID: QGV13007.1
Length: 264
>aminoglycoside-3'-O-phosphotransferase [Escherichia coli]
Sequence ID: AAA73390.1 Length: 264
Range 1: 1 to 264

Score:535 bits(1377), Expect:0.0,
Method:Compositional matrix adjust.,
Identities:263/264(99%), Positives:264/264(100%), Gaps:0/264(0%)

Query	1	MVEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRPVLVFKTDLSGALNEL	60
		M+EQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRPVLVFKTDLSGALNEL	
Sbjct	1	MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRPVLVFKTDLSGALNEL	60
Query	61	QDEAARLSWLATTGVPCA AVLDDVVTEAGRDWLLLGEVPGQDLLSSHLAPA EKVSIMADAM	120
		QDEAARLSWLATTGVPCA AVLDDVVTEAGRDWLLLGEVPGQDLLSSHLAPA EKVSIMADAM	
Sbjct	61	QDEAARLSWLATTGVPCA AVLDDVVTEAGRDWLLLGEVPGQDLLSSHLAPA EKVSIMADAM	120
Query	121	RRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQGLAPAELFARLKARMPD	180
		RRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQGLAPAELFARLKARMPD	
Sbjct	121	RRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQGLAPAELFARLKARMPD	180
Query	181	GEDLVVTHGDA CLPNIMVENGRFSGFIDCGRLGVADRYQDIALATRDIAEELGGEWADRF	240
		GEDLVVTHGDA CLPNIMVENGRFSGFIDCGRLGVADRYQDIALATRDIAEELGGEWADRF	
Sbjct	181	GEDLVVTHGDA CLPNIMVENGRFSGFIDCGRLGVADRYQDIALATRDIAEELGGEWADRF	240
Query	241	LVLYGIAAPDSQRIAFYRLLDEFF	264
		LVLYGIAAPDSQRIAFYRLLDEFF	
Sbjct	241	LVLYGIAAPDSQRIAFYRLLDEFF	264

Note that the *nos* promoter is expressed at a relatively low level in transgenic plants, resulting in 30-fold lower transcript levels than those generated by the CaMV 35S promoter, and 110-fold lower NPTII activity than with 35S (Sanders et al. 1987; reference attached to response). *Nos* promoter-driven expression of NPTII protein in regenerating cells in tissue culture is, evidently,

sufficient to render the regenerants resistant to aminoglycoside antibiotics for the purpose of selecting transgenic shoots for further propagation, but expression levels are low relative to what would typically be considered constitutive expression (Sanders et al. 1987), consistent with the inability to detect NPTII in the sensitive mass spectrometric analysis conducted on The Purple Tomato™. Given that only the fruit of tomato plants are eaten, the presence of trace levels of NPTII protein in other plant parts would have no bearing on safety of the consumed fruit.

We conclude that the DEL and ROS1 proteins present in The Purple Tomato™ pose no food safety risk. This assessment is based on: i) the vanishingly tiny quantities of the proteins present, ii) the widespread presence of similar proteins in edible plants, iii) the lack of any similarity to any known toxins or allergens, and iv) the rapid degradation of the recombinant proteins in simulated gastric proteolysis. NPTII protein has previously been deemed safe for consumption by FSANZ.

Other (non-protein) new substances

As described above, the novel gene products in the Purple Tomato, the DEL and ROS1 proteins, are detectable by their action on the endogenous anthocyanin biosynthesis pathway present in tomato. In non-transgenic tomato, this anthocyanin biosynthesis pathway is active in the hypocotyls of germinating seedlings and in leaf tissue. The DEL and ROS1 proteins are expressed in the Purple Tomato under the control of the tomato E8 promoter, which is activated in tomato fruit during ripening. As a result, the presence of DEL and ROS1 in ripening fruit activates the endogenous anthocyanin biosynthesis pathway in this plant organ. The result is purple pigmentation throughout the ripe fruit (Figures 5 and 6) due to biosynthesis of the two major anthocyanin pigments present in tomato (Figure 7). These anthocyanins, delphinidin 3-O-(coumaroyl) rutinoside-5-O-glucoside, and petunidin 3-O-(coumaroyl) rutinoside-5-O-glucoside, are produced in leaves on non-transgenic tomato, to protect photosynthetic tissues from light stress, and are also found in the skin of tomato fruit resulting from crosses of *S. lycopersicum* (cultivated tomato) with the wild species *S. chilense* or *S. cheesmaniae*. Tomato fruits produced from these crosses are currently available to consumers as the commercial tomato varieties Indigo Rose and Sunblack, among others (Mes, 2005; Mes et al., 2008; Gruber, 2017).





B

Figure 5. (A) Comparison of the growth and development of a *Del/Ros1* tomato (left) in the MicroTom variety to a non-transgenic wild type MicroTom tomato (right). (B) Fruit derived from the wild type MicroTom (top) compared to fruit from the Purple Tomato (bottom) harvested at the mature green (left), breaker (second from left), breaker plus 2 days (third from left), and red ripe (right) stages.

A



Figure 6. (a) Whole and halved *Del/Ros1* (purple) and wild type red tomato fruit. (b) Whole and halved ripe red wild type (top) and *Del/Ros1* (bottom) tomato fruit.

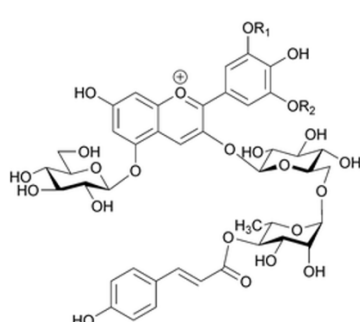


Figure 7. Structure of the major anthocyanins produced in tomato leaves of WT control plants and in the fruit of Purple Tomatoes.

Delphinidin 3-O-(coumaroyl)rutinoside-5-O-glucoside also known as Nasunin: R1 and R2 = H; Petunidin 3-O-(coumaroyl)rutinoside-5-O-glucoside: R1 = CH₃ and R2 = H.

Although Norfolk Healthy Produce does not intend to make explicit dietary health claims about of The Purple Tomato™, it is worth noting that dietary anthocyanins are associated with a reduced risk of chronic and degenerative diseases such as cardiovascular disease, obesity and certain cancers. Dietary anthocyanins have been reported to have protective effects against myocardial infarction and coronary heart disease (Cassidy *et al.*, 2013) as well as ameliorative effects in those already at risk of atherosclerosis (Zhu *et al.*, 2012). People consuming a high fruit and vegetable diet may eat as much as 60 mg of anthocyanins per day⁶, although beneficial effects have involved 5-fold higher anthocyanin consumption levels (>300 mg/day) in human intervention studies (Pojer *et al.*, 2013). Consequently, anthocyanin levels in the most commonly eaten fruits and vegetables are probably inadequate to confer optimal benefits. In addition, overall consumption of fruit and vegetables has declined over the last 10 years, with only 6% of the Australian population consuming the recommended daily intake of fruits and vegetables⁷. Similar trends hold true in other countries; for instance, in the US, only ~25% of the population consumes the recommended “daily 5” (Potter *et al.*, 2000).

Anthocyanin pigments occur in almost all higher plants, mostly in flowers and fruits, but they are also present in leaves, stems, and some roots, including several familiar fruits and vegetables (Table 3). Anthocyanins are also used as food additives in the European Union (EU food additive number E163) and in the U.S., the Food and Drug Administration (FDA) lists products rich in anthocyanins as approved for use as food colorants: Fruit juice - GMP - 73.250; Grape color extract – Non-beverage food only - 73.169; Grape skin extract (enocianina) - 73.170. The color of anthocyanins depends on their structure, and the pH of their environment. Anthocyanins are red under acidic conditions and turn blue under less acid conditions.

⁶ The consumption of anthocyanin-containing fruits and vegetables varies widely, depending on food availability and individual dietary preferences. As a result, average values estimated across an entire population vary widely.

⁷ <https://dietitiansaustralia.org.au/about-us/media-centre/were-not-getting-any-better-eating-fruit-veg-we-need-coordinated-approach-nutrition-now-media>

Table 3. Anthocyanin contents of different foods (in mg/100 g fresh weight of food)

Foodstuff	Manach	lpi.oregonstate	Researchgate
blackberry	100-400	100	82-180
blackcurrant	130-400	157	130-400
blueberry	25-500	163	83-530
cherry	35-450		350-450
chokeberry		485	506-1000
cranberry			46-200
eggplant (aubergine)	750		750
elderberry			200-1560
orange		5	
plum	2-25	6	2-25
radish		63	11-60
raspberry		47	10-60
red cabbage	25	210	25
red currant			12-19
red grape		48	30-750
red onions		10	up to 25
red wine	20-35	153	
rhubarb	200		up to 200
strawberry	15-75	27	13-36

Based upon extensive reviews of the published literature, no reports of adverse effects associated with the consumption of anthocyanins in food were identified (Pojer et al., 2013). Intervention studies supplementing daily anthocyanin consumption with 640 mg anthocyanins per day (Hassellund et al., 2012, 2013) reported no adverse effects, but rather positive effects in elevating HDL-cholesterol levels in healthy males. LD₅₀ values in excess of 2g/kg have been calculated, with no signs of ill effects at these levels (Pojer et al., 2013; Wallace and Giusti, 2015).

To investigate whether the levels of anthocyanins achieved in *Del/Ros1-N* purple tomatoes impacted growth, development or behavior in a dietary context in laboratory mice, 'wild type' C57B16 mice were fed with complete feed 'custom formula' for rodents, either without supplement, or supplemented with 10% powder from freeze-dried red (control) tomatoes or purple (*Del/Ros1-N*) tomatoes. It was concluded that supplementation of purple tomatoes at the equivalent of 6 mg/day (equivalent to 125 mg/day in humans; Reagan-Shaw et al., 2007) had no impact on food consumption or weight gain in mice. The details of the methods employed and results obtained appear in Appendix 10.

Transcription factors functionally and structurally equivalent to *Del* and *Ros1* control anthocyanin biosynthesis in all Angiosperm plants that produce these pigments including well known food crops such as pepper, potato, corn, apple, grape, blackberry, strawberry, blood orange (Borovsky et al., 2004; Jung et al., 2009; Ramsay and Glover, 2005; Paz-Ares et al., 1987; Espley et al., 2007; Kobayashi et al., 2002; Lin-Wang et al., 2010; Butelli et al., 2012). Because many foods are highly pigmented by anthocyanins, these pigments (and the transcription factor proteins that control their production) are ubiquitous in food and have a

previous history of safe use, including commercialized tomato varieties that produce anthocyanins in the fruit skin; Sun Black and Indigo Rose (Mes et al., 2008; Gruber, 2017). Considering this history of dietary exposure, there are no known allergies to anthocyanins (nor their biosynthetic enzymes, as described above).

The expression of *Del* and *Ros1* increases the transcript levels of almost all the genes encoding enzymes required for anthocyanin biosynthesis, genes encoding enzymes required for side-chain modification such as a putative anthocyanin acyltransferase, and two genes likely involved in the transport of anthocyanins into the vacuole (see Figure 4).

Anthocyanins were virtually undetectable in wild-type fruit but averaged 2.83 ± 0.46 mg of anthocyanin per g fresh weight in hemizygous MicroTom plants containing the *Del/Ros1-N* event. In other varieties of tomato the levels of anthocyanins were lower, largely due to the greater amount of water in larger tomato fruit. Thus, in the MoneyMaker genetic background, the levels were measured as 0.4 mg anthocyanin per g fresh weight (Butelli et al., 2008).

Varieties of tomatoes with purple skin have reported values in the literature. These are listed for comparison to the *Del/Ros1-N* event (in both MicroTom and Money Maker backgrounds) in Table 4.

Table 4. Anthocyanin levels in wild type (red), purple skinned, and *Del/Ros1-N* tomatoes.

Variety	Total anthocyanins (mg/gFW)	Source
wild type (red) tomatoes	undetectable	multiple, including Vu and Lee, 2019
Sun Black (purple skinned)	0.071	Blando et al. 2019
Indigo Rose (purple skinned)	0.1-0.3	https://horticulture.oregonstate.edu/oregon-vegetables/purple tomato faq
OSU P20	0.5-3	Mes et al., 2008
<i>Del/Ros1-N</i> MicroTom	2.83 ± 0.46	Butelli et al., 2008
<i>Del/Ros1-N</i> Money Maker	0.4	Butelli et al., 2008

Anthocyanins were detected using high-performance liquid chromatography (HPLC) in both peel and flesh of purple fruit, the major species being 3,5-diglucosides acylated with coumaric acid (Figure 8). There were no increases in flavonols in the flesh of purple tomatoes, although the main natural flavonol, rutin, was detected in the peel of purple fruit, while being barely detectable in the peel of wild-type comparator tomatoes. Additional analyses showed that the content of the other major phenylpropanoid (chlorogenic acid) was increased in the *Del/Ros1-N* tomato fruits compared to wild-type (about 2.5-fold). This change was likely the result of spill over of flavonoid intermediates from the increased flux to anthocyanins (Zhang et al., 2015; Tohge et al., 2015).

We note that Chlorogenic acid (CGA) is the major soluble phenolic in Solanaceous species such as potato, tomato and eggplant. It also accumulates to substantial levels in apples, pears, plums and coffee. CGA can limit low-density lipid (LDL) oxidation, the major determinant of the initial events in atherosclerosis. Furthermore, it removes particularly toxic reactive species by scavenging alkylperoxyl radicals and may prevent carcinogenesis by reducing the DNA damage they cause. (Niggeweg et al., 2004). CGA is consumed most highly in coffee and it has been estimated that a heavy coffee drinker may consume as much as 1 g CGA per day. Against the high content in many common foods including potatoes, chicory, and coffee, an increase from an average of 1.84 mg per 100g to 4.6 mg per 100g (one large tomato) would remain well within normal limits of exposure, as indicated from the table of results for CGA (synonymous with 5-caffeoylquinic acid) from the Phenol Explorer database: <http://phenol-explorer.eu/contents/polyphenol/475> (an extract of that table appears in Appendix 12). For instance, apple cider contains 49.33mg/100g CGA, and blueberries contain between 86.99-131.18mg/100g, while wild type tomatoes contain between 1.84-3.74mg/100g. So even at 2.5X elevated levels over wild type, the amount of CGA in Purple Tomatoes is many-fold lower than the amount found in commonly consumed foods.

The major anthocyanins present in purple tomato fruit are: Delphinidin 3-(coumaroyl)-rutinoside-5-glucoside, Delphinidin 3-(caffeoyl)-rutinoside-5-glucoside, Delphinidin 3-(feruloyl)-rutinoside-5-glucoside, Petunidin 3-(coumaroyl)-rutinoside-5-glucoside and Petunidin 3-(feruloyl)-rutinoside-5-glucoside as reported by Butelli et al., (2008) and Tohge et al., (2015) (Figure 8). The feruloylated species of anthocyanin are minor relative to the coumaroylated versions (<10%).

Chemically the anthocyanins in Purple Tomato are identical to those in eggplant and purple potatoes (see Azuma et al., 2008; Lachman et al., 2009; Condurache et al. 2021) and the anthocyanins that accumulate are exactly the same, chemically, as those that accumulate in leaves of regular tomatoes, in the skin of some purple-skinned tomato varieties already available commercially, and in eggplant and in purple-fleshed potatoes.

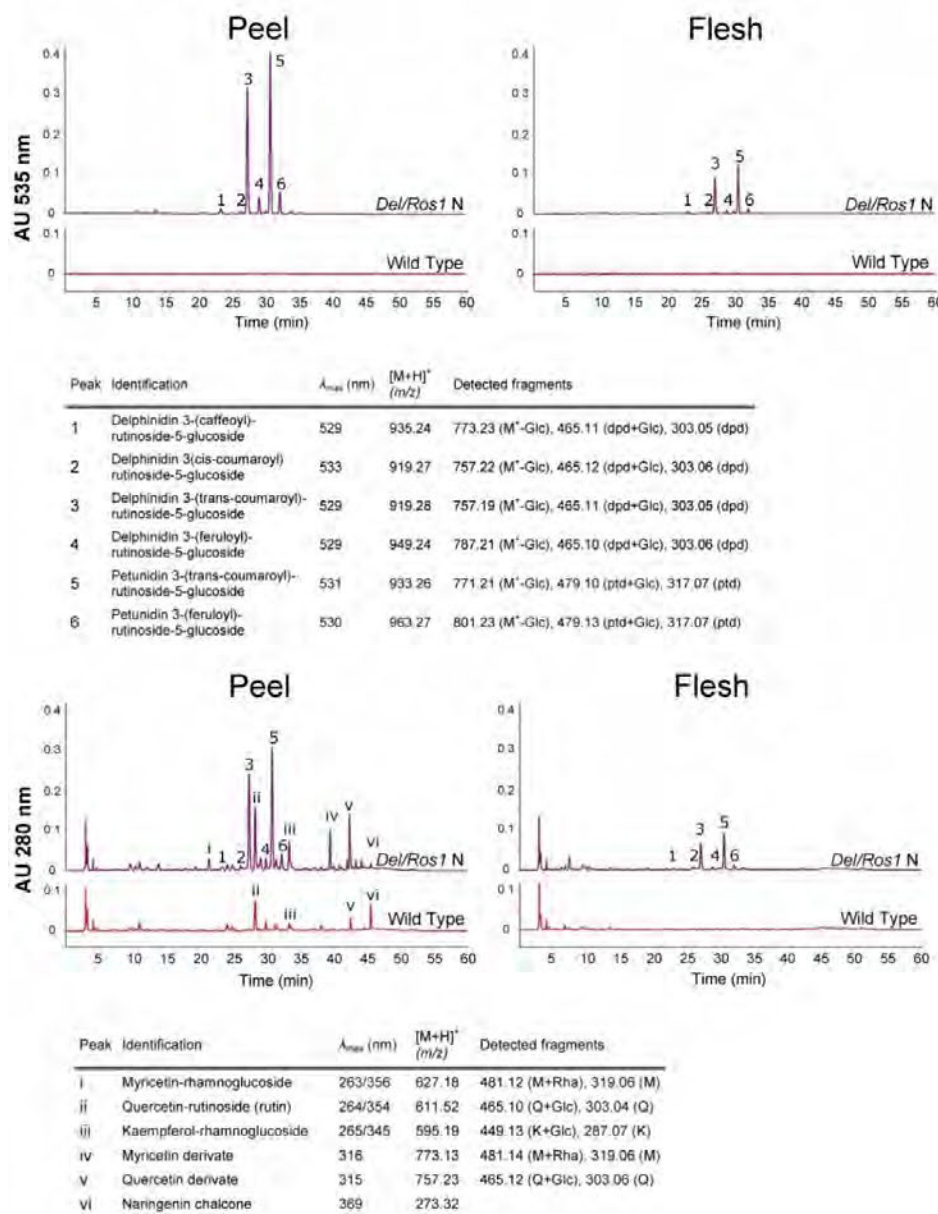


Figure 8. Comparative analysis of phenylpropanoid content and composition in purple vs. wild type tomato fruit. HPLC chromatogram of methanol extracts of peel or flesh of ripe fruit from *Del/Ros1-N* (purple line) and wild-type (red line) tomato, recorded at 535 nm. Peaks marked with numbers represent anthocyanins. Classification and identification of methanol soluble compounds was performed based on PDA absorbance and ESI-Q-TOF mass spectrometry. The purified compounds were analyzed by HPLC and ESI-MS/MS. Spectral characteristics, molecular ions and fragments obtained are tabulated. Identification was confirmed by hydrolysis and HPLC analysis of the respective acyl and sugar moieties.

Reviews of the anthocyanins present in the comparator foods show that a wide variation in types and amounts of anthocyanin species has been recorded within the same crop. For example, in blueberry, a variety of anthocyanins are present, including exactly those described in the Purple Tomato, but varying depending on the cultivar and environment (Yang et al.,

2022). To the extent that the functions of anthocyanins are known, there is no evidence that any of these compounds have different activity.

Compositional analyses of the food produced using gene technology

To assess compositional differences between purple tomatoes and isogenic non-transformed tomatoes, *Del/Ros1-N* tomatoes in the MoneyMaker genetic background (generated by backcrossing the *Del/Ros1-N* T1 transformant in the MicroTom background into the variety MoneyMaker and then selfing through 8 generations of single-seed descent) were compared to non-transformed ('wild type') control MoneyMaker tomatoes. Representative composite samples of both types of tomatoes were analyzed for their content of nutrients and non-nutrients.

Tomatoes were analyzed for:

Proximates – Protein, fat, total carbohydrates, fiber, ash

Minerals – Magnesium and potassium

Carotenoids – β -carotene and lycopene

Vitamins – C, K and folate (B9)

Nutrient composition was measured by Eurofins Food Testing UK Ltd (GLP accredited). Values are shown in Table 5 and given as mean $\pm$ SEM (standard error of means) from 5 replicate samples for red tomatoes and 5 replicate samples for purple *Del/Ros1-N* tomatoes. The values of all major nutritional components in the *Del/Ros1-N* Purple Tomatoes were similar to those in the red MoneyMaker tomatoes, and also similar to values listed by USDA and McCance and Widdowson Food Compositional database 2019 for raw, standard tomatoes. OECD values were not easy to relate to Eurofins values since they are reported on the basis of dry weight rather than fresh weight of tomatoes.

Although MicroTom was the laboratory variety used for the original transformation, no further lines from the MicroTom F6 have been developed, and MicroTom will not be sold. (It's strictly a laboratory strain and not considered a tasty variety.) We will be continuously breeding new tomato varieties using the transgenic locus that gives rise to the Purple Tomato phenotype. We have no evidence that the nutritional composition of any variety bearing the *Del/Ros1-N* trait will vary significantly from the wild type genetic background into which it is introgressed, with the exception of the expected elevated levels of the anthocyanin pigments.

Tomato varieties already vary widely with respect to various micronutrients (e.g. yellow varieties can have 10X higher flavonol levels than red varieties; Wang et al. 2018). Moreover, the environmental conditions in which tomatoes are grown can impact the macronutrient levels, and this within variety variation can be greater than between variety variation. As a result we conclude that any variety bred to incorporate the Purple Tomato trait will be subject to the same range of compositional variation as its wild type counterpart.

Table 5⁸		Nutrient composition of Tomatoes – standard, raw					
Analyte	units	RED	PURPLE	USDA (avg)	USDA (min)	USDA (max)	McCance and Widdowson 2019
moisture	g/100g	94.72 ± 0.12	95.1 ± 0.12	94.52	92.7	95.73	94.6
Crude protein	g/100g	0.64 ± 0.05	0.70 ± 0.09	0.88	0.59	1.06	0.5
Ash	g/100g	0.32 ± 0.02	0.48 ± 0.02	0.50	0.37	0.60	n/a
CHO	g/100g	3.06 ± 0.29	3.26 ± 0.27	3.89	n/a	n/a	3
Fructose	g/100g	1.46 ± 0.05	1.22 ± 0.09	1.37	1.1	2.32	1.6
Galactose	g/100g	<0.1 ± 0	0.1 ± 0	0.00	0.00	0.00	0.00
Glucose	g/100g	1.26 ± 0.02	1.04 ± 0.08	1.25	0.49	2.67	1.4
Lactose	g/100g	<0.1 ± 0	0.1 ± 0	0.00	0.00	0.00	0.00
Maltose	g/100g	<0.1 ± 0	0.1 ± 0	0.00	0.00	0.00	0.00
Sucrose	g/100g	<0.1 ± 0	0.1 ± 0	0.00	0.00	0.02	0.00
Total sugar	g/100g	2.72 ± 0.06	2.24 ± 0.17	2.63	1.59	5.01	3.00
Total fibre AOAC	g/100g	1.1 ± 0.2	0.7 ± 0.1	1.2	0.7	2	1
Energy kcal	kcal/100g	18.4 ± 0.83	16.8 ± 0.77	18	n/a	n/a	4
Energy kJ	kJ/100g	81.8 ± 1.78	71 ± 2.95	74	n/a	n/a	61
Total fat	g/100g	<0.3 ± 0.04	<0.3 ± 0	0.2	0.07	0.80	0.10
Salt ⁹	g/100g	<0.025 ± 0	<0.025 ± 0	n/a	n/a	n/a	n/a
Mono unsat FAs	g/100g	<0.1 ± 0	<0.1 ± 0	0.031	n/a	n/a	0.03
Poly unsat FAs	g/100g	<0.1 ± 0	<0.1 ± 0	0.083	n/a	n/a	0.05
Sat Fas	g/100g	<0.1 ± 0	<0.1 ± 0	0.028	n/a	n/a	0.03
Trans FAs	g/100g	<0.1 ± 0	<0.1 ± 0	n/a	n/a	n/a	0
Mg	g/100g	0.01 ± 0.00	0.01 ± 0.00	0.011	0.007	0.015	0.008
K	g/100g	0.16 ± 0.00	0.23 ± 0.01	0.237	0.144	0.385	0.0223
Na	g/100g	<0.01 ± 0	0.01 ± 0.00	0.005	0.001	0.024	0.002
beta carotene	ug/100g	451.6 ± 42.4	661.8 ± 64.8	449	184	572	349
Folate B9	ug/100g	8.92 ± 0.28	14.22 ± 0.31	13.7	7.8	19.8	23
Ascorbate C	mg/100g	6.86 ± 0.18	8.1 ± 0.86	15	1	36	22
Phylloquinone K1	ug/100g	2.69 ± 0.14	1.99 ± 0.1	7.9	2.2	60	6
Lycopene	mg / kg	n/a	20.9	25.73	11.36	34.19	n/a
Average and Min/Max values available online:							

⁸ Any values below the level of detection would be marked by Eurofins as below the Limit of Quantification” (LOQ), e.g. “< 0.0X mg/100g (LOQ)”. There were no values marked in this way for the compositional analysis, so the values given were detectable, but very low. In the case where a sample was <X and other samples had a measurable number, the <X values were recorded as zero for statistical purposes and analysis of variance was applied to the samples.

⁹ Salt (sodium chloride) was calculated on the basis of 2.5 times the content of sodium according to protocol UD815 UKAS accredited, for sodium chloride.

USDA Food Composition Databases Show Foods—Tomatoes, Red, Ripe, Raw, Year Round https://ndb.nal.usda.gov/ndb/ accessed 09-18-2018
McCance and Widdowson: The Composition of Foods Integrated Dataset 2019 available online: https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid
From 22 samples of standard raw commercial tomatoes accessed 08-25-2019

There were minor differences in most proximates (glucose, fructose, total CHO, total fat, total protein and fiber), vitamin C, the minerals Mg, K and Na between the Del/Ros1-N and the wild-type control. There were >25% differences in total folate, lycopene, β -carotene and α -tomatine between the Del/Ros1-N and the wild-type control. However, all contents were within the ranges reported in the scientific literature for commercial tomato varieties. Although some of these values are different with statistical significance, none are outside the reported ranges for tomatoes. For instance, folate can range between 4.1 and 35.3 ug/100g (Iniesta et al. 2009).

Carotenoids were not detected in all samples by Eurofins, so independent assays were undertaken using the acetone extraction method of Zhang and Hamauzu (2004), with the addition of an apo-8'-carotenal internal standard to quantify losses occurring during extraction and processing of the samples. All sample manipulations were carried out in a room with gold lighting to limit light-induced losses of carotenoids. The carotenoids were separated using the HPLC method of Hart and Scott (1995). β -carotene was used as the reference standard for quantification purposes, and a relative response factor correction of 1.26 applied to correct the values for lycopene. The main carotenoids described in tomatoes are lycopene (major) and β -carotene (minor).

The concentrations of carotenoids in the tomatoes appear in Table 6, below.

Table 6: Concentration of carotenoids in the tomato fruits.

	μg per 100g fresh weight			
	E-lycopene	Z-lycopene	Total lycopene	β-carotene
WT control	5720 $\pm$ 140	273 $\pm$ 74	5990 $\pm$ 200	755 $\pm$ 45
<i>Del/Ros1-N</i>	7020 $\pm$ 150	358 $\pm$ 10	7380 $\pm$ 160	866 $\pm$ 33

We believe that the data in Table 6 are more accurate than those in Table 5. Some data in nutritional composition databases used methods that may not have prevented losses during extraction, compared to the method used to generate the data in Table 6. Nonetheless, the β -carotene content of Purple Tomatoes is well below that found in many fruits and vegetables, such as mango, apricot, cantaloupe, carrots, red peppers, pumpkin, sweet potatoes, broccoli, and leafy greens (Holden et al., 1999). Traditionally bred tomato varieties containing 10X more β -carotene than recorded have been released¹⁰. There are many β -carotene-rich tomatoes available on the market including traditional red fruited cultivars, cv. Dhanshree at 1300 $\mu\text{g}/100\text{ g}$, cv. Pusa Ruby (90 $\mu\text{g}/100\text{ g}$) and orange fruited cultivars have 3800 $\mu\text{g}/100\text{ g}$, much higher values than the Del/Ros1-N tomatoes (Salunke et al. 2012). Orange-fruited varieties have not been listed in these databases, but are widely available on the market.

¹⁰ <https://www.ars.usda.gov/news-events/news/research-news/1998/usda-releases-new-tomatoes-with-increased-beta-carotene>

Similarly, the amount of lycopene found in purple tomatoes is well within the range seen in many foods, including foods based on wild type tomatoes (Holden et al., 1999). Moreover, no evidence of toxicity from ingestion of lycopene has been found (McClain and Bausch, 2003). We conclude that the amounts of lycopene and β -carotene observed in Purple Tomatoes are within normal limits of other foods and pose no additional risk upon ingestion compared to wild type tomatoes, or many other fruits and vegetables.

Eurofins also failed to detect Tomatine in either red or purple tomatoes. Consequently triplicate 50 mg portions of freeze-dried tomato and a taxifolin internal standard were dissolved into 1 cm³ of 70% methanol and vortex mixed for 1 hour. After diluting 1 part with 4 parts water, the samples were analyzed using HPLC-MS in positive ion mode. This method was used in favor of the convoluted Friedman method which involves precipitation of tomatine. The results appear in Table 7.

Table 7: α -Tomatine content in tomato fruits.

Sample Name	mg/kg α -tomatine in dry sample (mean)
Del/Ros1-N	88.0 $\pm$ 5.84
WT control	36.8 $\pm$ 1.71
OECD reported range ^a	0.0-389 mg/kg dry weight
OECD typical level ^a	16.9 mg/kg dry weight

^a Data taken from OECD ENV/JM/MONO(2008)26; Consensus Document on Compositional Considerations for New Varieties of Tomato

The contents of α -tomatine were significantly higher in the *Del/Ros1-N* tomatoes compared to the wild-type MoneyMaker control, but the observed levels were well within the broad range of α -tomatine concentrations for ripe tomatoes reported in the scientific literature. Further, the higher concentrations of α -tomatine in the *Del/Ros1-N* plants were consistent with the reported slower ripening of *Del/Ros1-N* fruits compared to the equivalent wild-type line (Zhang et al., 2013).

In the literature the range of tomatine content of ripe red tomato fruits is reported up to 202 mg/kg dry matter, with values approaching twenty times that, when fruit are in their immature green state (Dall'Asta et al., 2009; quoted as 36 – 202 mg/kg dry matter in ripe tomatoes). Friedman (2004) quoted 4 mg/kg dry weight for standard tomatoes to 42 mg/kg dry weight for cherry tomatoes, and a range of 4.0 (standard) – 157 mg/kg (sungold cherry). In a Czech language paper, Kalac reported 21 – 81 mg/kg in large maturing fruits, but levels below detection limit in ripe fruit, with no significant difference between field crops and crops grown under plastic. These papers report values relative to either dry weight or fresh weight. An approximate conversion is mg/kg dry weight / 14.49 = mg/kg fresh weight. Our analysis was carried out on freeze dried samples, so we simply converted from dry weight to fresh weight using an experimentally measured factor of 1g dry weight = 16.9g fresh weight. The water content of ripe tomatoes is typically between 90 - >95% (see, e.g. Zhang et al. 2023). Using our conversion factor, the dry weight value of 88 $\pm$ 5.84 mg/kg we report would

correspond to 5.17 ± 0.34 mg/kg fresh weight (compare to the cited “between undetectable and 23 mg/kg” for red ripe fruit in the OECD document)¹¹.

The MoneyMaker cultivar was used for compositional analysis because it was the closest in appearance to varieties into which we are introgressing the trait for commercialization. We are unable to provide compositional analysis for each cultivar that will be commercialized. The *Del/Ros1-N* locus is the target for introgression into multiple varieties that are trialed in Norfolk’s breeding greenhouse, and in collaboration with other breeders. These varieties are in continuous development through traditional breeding. The best of these varieties, as judged by agronomic (plant architecture, disease resistance, etc.) and gustatory (sweetness, acidity, etc.) and appearance (uniformity of color, absence of fruit cracking, etc.) characteristics is promoted each season to further development and seed multiplication. Thus, the variety currently being grown for commercial sale will change frequently—annually or at a minimum every couple of years.

Tomato composition varies within reasonable limits, as noted in the references to the compositional analysis section of the dossier. Indeed, the range of variation in composition is a result less of the genetics of the variety than the growing conditions (Collins et al. 2022) and, even more so, the ripeness of the fruit (Zhu et al. 2022). Any changes to fruit composition among varieties commercialised by Norfolk will be a result of the traditional breeding in which Norfolk is engaged, the growing conditions for the fruit, and their ripeness level, not a result of the now highly stable purple trait supplied by the *Del/Ros1-N* locus.

The aim of our compositional assessments was to determine whether or not the values in the new variety lie within the range of values for varieties already on the market. Standard errors of the mean are given for values measured for WT and *Del/Ros1-N* to provide a measure of the degree of variation in our compositional analyses, not to determine statistical significance of the values between the lines, because the differences are within standard ranges for tomatoes. For almost all the nutrients, our values are within the ranges quoted by these data bases, with the exceptions noted and explained above.

It was concluded that the nutritional composition of *Del/Ros1-N* tomatoes falls within the published expected ranges for tomatoes. Additional details of the analyses appear in Appendix 12.

¹¹ Dry matter content varies widely among tomato cultivars and growing conditions. As cited in Kurina et al. (2021), the dry matter content of cultivated tomatoes can range from 3.72% to 8.88% of fresh weight. This corresponds to a range of conversion factors from 11.36 to 26.9. Databases typically do not define the precise type/variety of tomato being reported; consequently, the values given for average conversion ratios are averaged across different types of tomato.

Information related to the nutritional impact of the food produced using gene technology

There is a wealth of published information regarding mechanisms of action of anthocyanins in humans, including their bioactivity, uptake, absorption, bioavailability, whole body distribution, and tissue localization (Lila, 2004, Pojer et al., 2013). Studies on the absorption and clearance kinetics of anthocyanins showed that maximum plasma concentrations are between 1.4 and 592 nmol/L and occur at 0.5-4 h post-consumption (doses; 68-1300 mg).

Absorption, distribution, metabolism, and elimination data from human-intervention studies have suggested maximal serum concentrations of anthocyanins and anthocyanidin phase II conjugates are reached by 1.5 h post ingestion at amounts of 100 nmol/L after doses of 500 mg anthocyanins (Kay, 2006; Garcia-Alonso et al., 2009). These findings suggest that anthocyanins are considerably less bioavailable than other flavonoid subclasses. Recent bioavailability studies indicate that anthocyanins have a minimum relative bioavailability of $12.3 \pm 1.3\%$, and their microbial metabolites reach a 42-fold higher peak serum concentration that occurs much later than that of the parent anthocyanin. The major metabolites found in plasma are phenolic acid derivatives produced by the gut intestinal microbiota prior to absorption (Czank et al., 2013).

Average urinary excretion is between 0.03 and 4.00 % of the ingested dose, with an elimination half-life of between 1.5-3.0 h (Kay, 2006, Pojer et al, 2013). Metabolites in the diet are subject to three types of metabolism: a) as a result of enzymes produced in the gastrointestinal (GI) tract, b) microbial metabolism in the GI tract and c) metabolism post-absorption. Metabolism of anthocyanins post-absorption is limited in humans and microbial metabolism, predominantly in the colon, is the most important type of metabolism. During metabolism of anthocyanins post-absorption, conjugation reactions including glucuronidation, methylation, and sulfation can occur but such conjugates are relatively minor constituents in plasma (Pojer et al., 2013). Anthocyanins are absorbed and transported in human serum primarily as microbial metabolites, with as much as 68-80 % of anthocyanins found as metabolized derivatives in human urine (Kay, 2006).

The average daily human consumption of anthocyanins (dietary exposure) was estimated as part of the EU-funded ATHENA research program. Average daily consumption of anthocyanins in a sample of 450 Italian volunteers (119 vegetarians and 331 omnivores) was 27.5 mg/day (SD 45.84) with a median value of 13.6 mg/day (Figure 9; Rizzi *et al.*, 2016). These data derive from subjects largely adherent to a Mediterranean diet rich in fruit and vegetables. In contrast, Wu *et al.* (2006) estimated anthocyanin intake in US adults to be only 12.5 mg/day. It is likely that with the switch from traditional diets to 'western-style' diets, the average consumption of anthocyanins has declined.

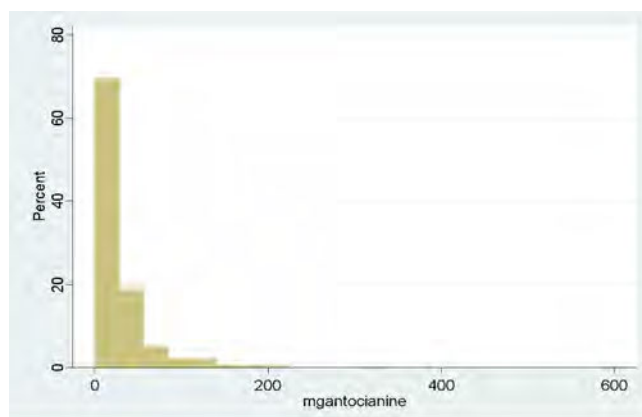


Figure 9. Estimation of the daily consumption of anthocyanins by Italians n=450 (Rizzi et al., 2016)

In assessing dietary exposure to anthocyanins in Purple Tomatoes, it was assumed that Purple Tomatoes would be commercially available in a manner similar to unmodified tomatoes. Consequently, our exposure scenario includes complete replacement of tomatoes in the diet by Purple Tomatoes. The safety of the anthocyanins in the Purple Tomatoes has been determined by comparison of current daily intake of anthocyanins (assumed to be completely from diet) with the resulting probable exposure to anthocyanins upon replacement of all tomatoes with Purple Tomatoes.

Intake of tomatoes in the diet was estimated using The National Health and Nutrition Examination Survey (NHANES, performed by the US Centers for Disease Control¹²) data for 2-day consumption of all foods that might contain tomatoes and recipe-species amounts of tomatoes in each food. Using data from the most recent set for which data are available, collected in the 2009-10 cycle of NHANES, it was estimated that total tomato consumption averages 35 grams per day and rises to 75 grams per day for the 90th percentile. Approximately 56% of the US population report having eaten a tomato-containing food. If these were all Purple Tomatoes, with anthocyanins present at 0.3%, daily exposure to the anthocyanins would average 100 mg and rise to 225 mg for the 90th percentile. Note that in Australia, average consumption of tomatoes and tomato products was 22g/person/day in 2023-24¹³, lower than the value reported in the US.

Background dietary exposure to anthocyanins was estimated using the levels in the foods shown in Table 5 combined with the 2-day consumption data for those foods (taken from the 2009-16 combined data sets). The high end of the reported ranges of anthocyanins in each food was assumed. Approximately 20% of the population reported eating at least one of these foods. The daily exposure to anthocyanins (background exposure) therefore averages 130 mg rising to 340 mg for the 90th percentile.

Overall consumption of anthocyanins, if all tomatoes consumed were Purple Tomatoes, would be less than the sum of these two means, as not all consumers of tomatoes are consumers of

¹² <https://www.cdc.gov/nchs/nhanes/index.html>

¹³ <https://www.abs.gov.au/statistics/health/health-conditions-and-risks/apparent-consumption-selected-foodstuffs-australia/latest-release>

other anthocyanin-containing foods. The simple addition of the mean of 100mg from Purple Tomatoes to the mean of 130mg from other sources, *i.e.* 230 mg/day in total, would be conservative relative to known safe values.

Table 5. Anthocyanin levels used to compute background dietary exposure Food.

Species	Anthocyanin Level (mg/kg)
Cherry	4000
Blackberry	3250
Blueberry	5000
Raspberry	600
Strawberry	350
Radish	600
Red Wine	1000
Eggplant	7500

The Joint WHO/FAO Expert Committee on Food Additives considered exposure to anthocyanins from black carrot extract at its 87th meeting in 2019 as part of a safety assessment of its use as a food coloring agent. The meeting concluded that mean daily exposure ranged from approximately 0.1 to 1.3 mg/kg-body weight (bw)/d, the latter for the toddler age grouping (Europe). This would be equivalent to approximately 6 to 20 mg/d for a 60 kg adult and a 15 kg toddler. At the 95th percentile, these figures range up to 35 mg/d for a toddler. The committee also noted that overall background exposure could be as high at 4 mg/kg-bw for a toddler (60 mg/d). These estimates are lower than those estimated above because of the use of longer dietary surveys for the food consumption data, which are known to result in lower chronic exposure estimates.

The estimates of exposure to anthocyanins from Purple Tomatoes are of the order of those from consumption of colored fruits. For example, consumption of an 8 oz. (237 ml) glass of purple tomato juice¹⁴ is approximately equivalent to the consumption of one serving of purple eggplant (200 grams, Manach, 2004) or two servings of blueberries (200 grams, Manach, 2004).

Although consumption of purple tomatoes would increase exposure to anthocyanins, our conservative estimates suggest that consumption would no more than double for people currently consuming anthocyanin-colored foods. People who do not currently consume purple foods would be exposed at the mean to approximately the same amount of anthocyanins as current consumers of colored food.

From these analyses it was concluded that the levels of anthocyanins in Purple Tomatoes are equivalent to levels found in many fruits and vegetables. We also conclude that any additional dietary exposure from purple tomatoes, even at the 90th percentile level of 75g/day (amounting to ~225mg of anthocyanins) is well within the range of anthocyanin values for foods already routinely consumed by significant portions of the human population.

¹⁴ The anthocyanin content in juice depends on the tomato variety from which the juice has been made. Larger tomatoes contain more water and so the anthocyanins are less concentrated. Based on repeated measurements of MoneyMaker purple tomatoes we measured 17.2 ± 1.2 mg cyanidin 3-glucoside equivalents per 100mL juice (which corresponds to 40.76 mg C-3-G equivalents in 237 mL [1 cup, or 8 fl. oz.] of juice).

THIS PAGE INTENTIONALLY LEFT BLANK

References

- Aharoni, A., Ric De Vox, C.H., Wein, M., Sun, Z., Greco, R., Kroon, A., Mol, J.N.M., and O Connell, A.P. (2001). The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *Plant J.* **28**: 319-332.
<https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-3113X.2001.01154.x?sid=nlm%3Apubmed>
- Akiyama, R., Nakayasu, M., Umemoto, N., Kato, J., Kobayashi, M., Lee, H.J., Sugimoto, Y., Iijima, Y., Saito, K., Muranaka, T., and Mizutani, M. (2021). Tomato E8 encodes a C-27 hydroxylase in metabolic detoxification of α -tomatine during fruit ripening. *Plant Cell Physiol.* **62**: 775-783.
- Albert, NW, Butelli, E, Moss, SMA, Piazza, P, Waite, CN, Schwinn, KE, Davies, KM, Martin, C. (2021). Discrete bHLH transcription factors play functionally overlapping roles in pigmentation patterning in flowers of *Antirrhinum majus*. *New Phytol.* **231**: 849-863.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC8248400/>
- Azuma, K., Ohyama, A., Ippoushi, K., Ichiyanagi T., Takeuchi, A., Saito, T. and Fukuoka, H. (2008). Structures and antioxidant activity of anthocyanins in many accessions of eggplant and its related species. *J. Agric. Food Chem.* **56**: 10154–10159.
<https://pubs.acs.org/doi/10.1021/jf801322m>
- Blando F, Berland H, Maiorano G, Durante M, Mazzucato A, Picarella ME, Nicoletti I, Gerardi C, Mita G, and Andersen ØM. (2019). Nutraceutical Characterization of Anthocyanin-Rich Fruits Produced by "Sun Black" Tomato Line. *Front Nutr.* **6**: 133.
<https://www.frontiersin.org/journals/nutrition/articles/10.3389/fnut.2019.00133/full>
- Borovsky, Y., Oren-Shamir, M., Ovadia, R., De Jong, W. and Paran, I., (2004). The A locus that controls anthocyanin accumulation in pepper encodes a MYB transcription factor homologous to Anthocyanin2 of *Petunia*. *Theoretical and Applied Genetics*, **109**: 23-29.
<https://link.springer.com/article/10.1007/s00122-004-1625-9>
- Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., Schijlen, E.G.W.M., Hall, R.D., Bovy, A.G., Luo, J., and Martin, C. (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nature Biotechnology* **26**: 1301 – 1308.
<https://www.nature.com/articles/nbt.1506>
- Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G. and Martin, C., (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *The Plant Cell* **24**: 1242-1255.
<https://academic.oup.com/plcell/article/24/3/1242/6097241>
- Cao, Y., Li, K., Zhao, X., and Wang, L. (2020). MYB transcription factors as regulators of secondary metabolism in plants. *Biology* **9**, 61.
<http://dx.doi.org/10.3390/biology9030061>
[reference added to describe ubiquity of MYB transcription factors in plants]

Cassidy A, Mukamal KJ, Liu L, Franz M, Eliassen AH, Rimm EB. (2013) High anthocyanin intake is associated with a reduced risk of myocardial infarction in young and middle-aged women. *Circulation* 127:188–96.

https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.112.122408?url_ver=Z39.88-2003&rft_id=ori:rid:crossref.org&rft_dat=cr_pub%20%20pubmed

Collins, E.J., Bowyer, C, Tsouza, A., and Chopra, M. (2022). Tomatoes: an extensive review of the associated health impacts of tomatoes and factors that can affect their cultivation. *Biology* 11: 239.

<https://doi.org/10.3390/biology11020239>

Comai, L., Schilling-Cordaro, C., Mergia, A. and Houck, C.M., (1983). A new technique for genetic engineering of *Agrobacterium* Ti plasmid. *Plasmid*, 10:21-30.

<https://www.sciencedirect.com/science/article/abs/pii/0147619X83900549?via%3Dihub>

Condurache (Lazar), N-N., Croitoru, C., Enachi, E., Bahrim, G-E., Stanciuc, N., and Rapeanu, G. (2021). Eggplant peels as a valuable source of anthocyanins: extraction, thermal stability and biological activities.

<https://doi.org/10.3390/plants10030577>

Currence, T.M. and J.M. Jenkins (1942). Natural crossing in tomatoes as related to distance and direction. *Proc. Amer. Soc. Horticul. Sci.* 42:273-276.

[We are unable to obtain an original copy of this reference. We note that the key finding of this study, on frequency of outcrossing, is summarized in the below-referenced OECD Consensus Document on the Biology of Tomato (2016), at paragraph 55.]

Czank, C., Cassidy, A., Zhang, Q., Morrison, DJ., Preston, T., Kroon, PA., Botting, NP. and Kay, CD. (2013) Human metabolism and elimination of the anthocyanin, cyanidin-3-glucoside: a ¹³C-tracer study. *Am. J. Clin. Nutr.* 97: 995-1003.

[https://ajcn.nutrition.org/article/S0002-9165\(23\)05506-5/fulltext](https://ajcn.nutrition.org/article/S0002-9165(23)05506-5/fulltext)

Dall'Asta, C., Falavigna, C., Galaverna, G., Sforza, S., Dossena, A. and Marchelli, R., (2009). A multiresidual method for the simultaneous determination of the main glycoalkaloids and flavonoids in fresh and processed tomato (*Solanum lycopersicum* L.) by LC-DAD-MS/MS. *Journal of Separation Science*, 32: 3664-3671.

<https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/full/10.1002/jssc.200900365>

Deikman, J., Kline, R., and Fischer, R.L. (1992). Organization of ripening and ethylene regulatory regions in fruit-specific promoter from tomato (*Lycopersicon esculentum*). *Plant Physiol.* 100:2013-2017.

<https://academic.oup.com/plphys/article/100/4/2013-2017/6087326>

Deluc, L., Bogs, J., Walker, A.R., Ferrier, T., Decendit, A., Merillon, J-M., Robinson, S.P., and Barrieu, F. (2008). The transcription factor VvMYB5B contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. *Plant Physiol.* 147: 2041-2053.

<https://pmc.ncbi.nlm.nih.gov/articles/PMC2492604/>

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D.R. (1980). Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* 77:7347-735.

<https://www.pnas.org/doi/abs/10.1073/pnas.77.12.7347>

Espley, R.V., Hellens, R.P., Putterill, J., Stevenson, D.E., Kuty-Amma, S. and Allan, A.C., (2007). Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *The Plant Journal*, **49**: 414-427.

<https://onlinelibrary.wiley.com/doi/10.1111/j.1365-313X.2006.02964.x>

Faurobert, M., et al. (2007). Major proteome variations associated with cherry tomato pericarp development and ripening. *Plant Physiol.* **143**:1327-46.

<https://academic.oup.com/plphys/article/143/3/1327/6106672>

Friedman, M., (2004). Analysis of biologically active compounds in potatoes (*Solanum tuberosum*), tomatoes (*Lycopersicon esculentum*), and jimson weed (*Datura stramonium*) seeds. *Journal of Chromatography A*, **1054**: 143-155.

<https://www.sciencedirect.com/science/article/abs/pii/S002196730400648X?via%3Dihub>

Garcia-Alonso M, Minihane A-M, Rimbach G, Rivas-Gonzalo JC, de Pascual-Teresa S. (2009) Red wine anthocyanins are rapidly absorbed in humans and affect monocyte chemoattractant protein 1 levels and antioxidant capacity of plasma. *J. Nutr. Biochem.* **20**: 521–9.

<https://www.sciencedirect.com/science/article/abs/pii/S0955286308001198?via%3Dihub>

Goodrich, J., Carpenter, R., and Coen, E.S.. (1992). A common gene regulates pigmentation pattern in diverse plant species. *Cell* **68**:955–964.

[https://www.cell.com/cell/abstract/0092-8674\(92\)90038-](https://www.cell.com/cell/abstract/0092-8674(92)90038-E?returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2F009286749290038E%3Fshowall%3Dtrue)

[E?returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2F009286749290038E%3Fshowall%3Dtrue](https://www.cell.com/cell/abstract/0092-8674(92)90038-E?returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2F009286749290038E%3Fshowall%3Dtrue)

Gruber, K. (2017). Agrobiodiversity: The living library. *Nature*, **544**:S8-S10.

<https://www.nature.com/articles/544S8a>

Hart, D.J. and Scott, K.J., (1995). Development and evaluation of an HPLC method for the analysis of carotenoids in foods, and the measurement of the carotenoid content of vegetables and fruits commonly consumed in the UK. *Food Chemistry*, **54**: 101-111.

<https://www.sciencedirect.com/science/article/abs/pii/030881469592669B>

Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S. and Mullineaux, P.M. (2000). pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Molecular Biology*, **42**:819-832.

<https://link.springer.com/article/10.1023/A:1006496308160>

Holden, J.M., Eldridge, A.L., Beecher, G.R., Buzzard, I.M., Bhagwat, S., Davis, C.S., Douglass, L.W., Gebhardt, S., Haytowitz, D., and Schakel, S. (1999). Carotenoid contents of U.S. foods: an update of the database. *J. Food Comp. and Analysis* **12**: 169-196.

Iniesta, M.D., Pérez-Conesa, D., García-Alonso, J., Ros, G., and Periado, M.J (2009). Folate content in tomato (*Lycopersicon esculentum*). Influence of cultivar, ripeness, year of harvest, and pasteurization and storage temperature. *J. Agric. Food Chem.* **57**: 4739-4745.

<https://pubs.acs.org/doi/10.1021/jf900363r>

Jones, J.D.G., Shlumukov, L., Carland, F., English, J., Scofield, S.R., Bishop, G.J. and Harrison, K. (1992). Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. *Transgenic Res.* **1**:285–297.

<https://link.springer.com/article/10.1007/BF02525170>

Jung, C.S., Griffiths, H.M., De Jong, D.M., Cheng, S., Bodis, M., Kim, T.S. and De Jong, W.S., (2009). The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. *Theoretical and Applied Genetics*, **120**: 45-57.

<https://link.springer.com/article/10.1007/s00122-009-1158-3>

Kay CD (2006) Aspects of anthocyanin absorption, metabolism, & pharmacokinetics in humans. *Nutr. Res. Rev.* **19**: 137–46.

<https://www.cambridge.org/core/journals/nutrition-research-reviews/article/aspects-of-anthocyanin-absorption-metabolism-and-pharmacokinetics-in-humans/BD2C2539E17B6596E514C865DAD6DFC8#>

Kobayashi, S., Ishimaru, M., Hiraoka, K. and Honda, C., (2002). Myb-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis. *Planta* **215**: 924-933.

<https://link.springer.com/article/10.1007/s00425-002-0830-5>

Kurina, A.B., Solovieva, A.E., Khrapalova, I.A., and Artemyeva, A.M. (2021). Biochemical composition of tomato fruits of various colors. *Vavilov J. Genetics Breeding* 25:514-527. DOI 10.18699/VJ21.058

Lachman, J., Hamouz, K.K., Orsak, M., Pivec, V., Hejtmankova, A., Dvoron, P. and Cepl, J. (2009). Cultivar differences of total anthocyanins and anthocyanidins in red and purple-fleshed potatoes and their relation to antioxidant activity. *Food Chemistry* **114**:836–843.

DOI:[10.1016/j.foodchem.2008.10.029](https://doi.org/10.1016/j.foodchem.2008.10.029)

Lila, M.A. (2004) Anthocyanins and human health: an in vitro investigative approach. *Journal of Biomedicine and Biotechnology* 5: 306–313.

<https://onlinelibrary.wiley.com/doi/10.1155/S111072430440401X>

Lin-Wang, K., Bolitho, K., Grafton, K., Kortstee, A., Karunairetnam, S., McGhie, T.K., Espley, R.V., Hellens, R.P. and Allan, A.C., (2010). An R2R3 MYB transcription factor associated with regulation of the anthocyanin biosynthetic pathway in Rosaceae. *BMC Plant Biology* **10**: 50.

<https://bmcplantbiol.biomedcentral.com/articles/10.1186/1471-2229-10-50>

Liu J, Osbourn A, Ma P. (2015) MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. *Mol Plant* **8**: 689-708.

[https://www.cell.com/molecular-plant/fulltext/S1674-2052\(15\)00180-X?returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS167420521500180X%3Fshowall%3Dtrue](https://www.cell.com/molecular-plant/fulltext/S1674-2052(15)00180-X?returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS167420521500180X%3Fshowall%3Dtrue)

Manach, C, Scalbert, A, Morand, C, Rémésy, C. and Jiménez, J. (2004) Polyphenols: food sources and bioavailability. *Am J Clin Nutr* 79:727–47.

<https://www.sciencedirect.com/science/article/pii/S0002916522039144?via%3Dihub>

- McClain, R.M., and Bausch, J. 2003. Summary of safety studies conducted with synthetic lycopene. *Regulatory Tox. and Pharmacology* **37**: 274-285.
- McCormick, S. (1991). Transformation of tomato with *Agrobacterium tumefaciens*. *Plant Tissue Culture Manual* **B6**: 1-9.
- Mes, P. (2005). Breeding Tomatoes for Improved Antioxidant Activity. Ph.D. Thesis, Oregon State University, Corvallis, OR.
https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/v118rg71s
- Mes, P.J., Boches, P., Myers, J.R. and Durst, R. (2008). Characterization of tomatoes expressing anthocyanin in the fruit. *J. Am. Soc. Hort. Sci.*, **133**:262-269.
<https://journals.ashs.org/jashs/view/journals/jashs/133/2/article-p262.xml>
- Niggeweg, R., Michael, A. & Martin, C. (2004). Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nat Biotechnol.* **22**: 746–754 (2004). <https://doi.org/10.1038/nbt966>
- OECD (2016). Consensus document on the biology of tomato.
<https://doi.org/10.1787/9789264279728-6-en>
[https://one.oecd.org/document/ENV/JM/MONO\(2016\)53/en/pdf](https://one.oecd.org/document/ENV/JM/MONO(2016)53/en/pdf)
- Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A. and Saedler, H., 1987. The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *The EMBO Journal*, **6**: 3553-3558.
<https://www.embopress.org/doi/abs/10.1002/j.1460-2075.1987.tb02684.x>
- Pojer E., Mattivi F., Johnson D., Stockley C.S. (2013). The case for anthocyanin consumption to promote human health: a review. *Compr Rev Food Sci Food Saf.* **12**:483–508.
<https://ift.onlinelibrary.wiley.com/doi/10.1111/1541-4337.12024>
- Potter JD, Finnegan JR, Guinard J-X, et al. (2000) 5-A-Day for Better Health. Program evaluation Report. Bethesda, MD: National Institutes of Health, National Cancer Institute. November 2000; NIH Publication No. 01-4904.
[Unable to access the original. Similar findings of the relatively low frequency of consumption of “five a day” in the Canadian population are found at:
<https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310009612>
and
<https://www150.statcan.gc.ca/n1/pub/82-003-x/2020004/article/00001-eng.htm>
- Ramsay, N.A. and Glover, B.J., (2005). MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science*, **10**: 63-70.
<https://www.sciencedirect.com/science/article/abs/pii/S1360138504002973>
- Reagan-Shaw, S., Nihal, M. & Ahmad, N. (2007) Dose translation from animal to human studies revisited. *FASEB J.*, **22**: 659–661.
<https://faseb.onlinelibrary.wiley.com/doi/10.1096/fj.07-9574LSF>
- Rick, C.M. (1978). The Tomato. *Scientific Amer.* **239**:76-87.
<https://www.scientificamerican.com/article/the-tomato-1978-08/>

Rizzi, F., Conti, C., Dogliotti, E., Terranegra, A., Salvi, E., Braga, D., Ricca F., Lupoli, S., Mingione, A., Pivari, F., Brasacchio, c., Barcella, M., Chittani, M., D'Avila, F., Turiel, M., Lazzaroni, M., Soldati, L., Cusi, D. and Barlassina, C. (2016) Interaction between polyphenols intake and PON1 gene variants on markers of cardiovascular disease: a nutrigenetic observational study. *J. Transl. Med.* **14**:186
<https://dx.doi.org/10.1186/s12967-016-0941-6>

Salunke, D.S., Jadhav, A.S., Pawar, B.D., Kale, A.A. and Chimote, V.P., 2012. Diversity analysis of tomato genotypes using RAPD markers and high-performance liquid chromatography in relation to β -carotene content. *Vegetos*, **25**: 95-101.
Sanders P.R., Winter J.A., Barnason, A.R., Rogers, S.G., and Fraley, R.T. (1987). Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. *Nucleic Acids Res.* **15**: 1543-1558.

Schwinn, K., Venail, J., Shang, Y., Mackay, S., Alm, V., Butelli, E., Oyama, R., Bailey, P., Davies, K., and Martin, C. (2006). A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell* **18**:831–851.
<https://academic.oup.com/plcell/article/18/4/831/6114873>

Tanksley, S.D. (1993). Linkage map of tomato *Lycopersicon esculentum*. In: Genetic Maps; S. O'Brein ed., Cold Springs Harbor Laboratory Press, Cold Springs Harbor, NY. pp 6.39-6.60. [Cannot access original reference. The reference was simply to state that the tomato genome has 12 chromosomes. We substitute instead the publication of the tomato genome sequence, and have added the PDF for the reference: The Tomato Genome Consortium, 2012
<https://www.nature.com/articles/nature11119>]

Tohge, T., Zhang, Y., Peterrek, S., Matros, A., Rallapalli, G., Tandron, Y.A., Butelli, E., Kallam, K., Hertkorn, N., Mock, H.P., Martin, C. and Fernie, A.R. (2015). Ectopic expression of snapdragon transcription factors facilitates the identification of genes encoding enzymes of anthocyanin decoration in tomato. *Plant J.* **83**:686–704.
<https://onlinelibrary.wiley.com/doi/10.1111/tpj.12920>

USDA (1992). Interpretive Ruling on Calgene, Inc., Petition for Determination of Regulatory Status of FLAVR SAVR™ Tomato. Federal Register 57, no. 202, pp. 47608-47616.
<https://www.federalregister.gov/citation/57-FR-20377>

van den Elzen, P., Lee, K.Y., Townsend, J. and Bedbrook, J. (1985). Simple binary vectors for DNA transfer to plant cells. *Plant Molecular Biology*, **5**:149-154.
<https://link.springer.com/article/10.1007/BF00015678>

Vu, A.T., Lee, J.M. (2019). Genetic variations underlying anthocyanin accumulation in tomato fruits. *Euphytica* **215**:196.
<https://link.springer.com/article/10.1007/s10681-019-2519-x>

Wallace, T.C., and Giusti, M.M. (2105). Anthocyanins. *Adv. Nutr.* **6**:620-622.
<https://www.sciencedirect.com/science/article/pii/S2161831323001072?via%3Dihub>

Wang, K., Herrera-Estrella L., Van Montagu, M., Zambryski, P. (1984). Right 25 bp terminus sequence of the nopaline T-DNA is essential for and determines direction of DNA transfer from *Agrobacterium* to the plant genome. *Cell* **38**:455-462.

<https://www.sciencedirect.com/science/article/abs/pii/S0092867484905002>

Wang, N., Xu, H., Jiang, S., Zhang, Z., Lu, N., Qiu, H., Qu, C., Wang, Y., Wu, S., and Chen, X. (2017). MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (*Malus sieversii* f. *niedzwetzkyana*). *Plant J.* **90**: 276-292.

<https://onlinelibrary.wiley.com/doi/10.1111/tpj.13487>

Wang, S., Chu, Z., Jia, R., Dan, F., Shen, X., Li, Y., and Ding, X. (2018) SIMYB12 regulates flavonol synthesis in three different cherry tomato varieties. *Sci. Reports* **8**:1582

DOI:10.1038/s41598-018-19214-3

Wu X., GR. Beecher, JM. Holden, DB. Haytowitz, SE. Gebhardt and RL. Prior (2006) Concentrations of Anthocyanins in Common Foods in the United States and Estimation of Normal Consumption. *J. Agric. Food Chem.* **54**:4069-4075.

<https://pubs.acs.org/doi/10.1021/jf060300l>

Yang, W., Guo, Y., Liu, M., Chen, X., Xiao, X., Wang, S., Gong, P., Ma, Y., and Chen, F. (2022). Structure and function of blueberry anthocyanins: a review of recent advances. *J. Functional Foods* **88**: 104864.

<https://doi.org/10.1016/j.jff.2021.104864>

Zambryski, P., Depicker, A., Kruger, K., Goodman, H.M. (1982). Tumor induction by *Agrobacterium tumefaciens*: analysis of the boundaries of T-DNA. *J. Mol. Appl. Genet.* **1**:361-370.

<https://pubmed.ncbi.nlm.nih.gov/7108407/>

[Original reference is unavailable, but this is the original source for identification of the T-DNA borders, now widely corroborated in subsequent publications, and used extensively in vectors for *Agrobacterium* transformation of plants.]

Zhang, Y., Butelli, E., De Stefano, R., Schoonbeek, H., Magusin, A., Pagliarini, C., Wellner, N., Hill, L., Orzaez, D., Granell, A., Jones, J.D.G. and Martin, C. (2013). Anthocyanins Double the Shelf Life of Tomatoes by Delaying Overripening and Reducing Susceptibility to Gray Mold. *Current Biology* **23**:1094–1100.

[https://www.cell.com/current-biology/fulltext/S0960-9822\(13\)00513-7?returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0960982213005137%3Fshowall%3Dtrue](https://www.cell.com/current-biology/fulltext/S0960-9822(13)00513-7?returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0960982213005137%3Fshowall%3Dtrue)

Zhang, Y. Butelli, E. Alseekh, S. Tohge, T. Rallapalli, G. Luo, J. Kwar, P.G. Hill, L. Santino, A. Fernie, A.R. & Martin, C. (2015) Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat. Commun.* **26**: 8635.

<https://www.nature.com/articles/ncomms9635>

Zhang, D. and Hamauzu, Y., (2004). Phenolics, ascorbic acid, carotenoids and antioxidant activity of broccoli and their changes during conventional and microwave cooking. *Food Chemistry*, **88**: 503-509.

<https://www.sciencedirect.com/science/article/abs/pii/S0308814604001517>

Zhang, J., Liu, S., Zhu, X., Chang, Y., Wang, C., Ma, N., Wang, J, Zhang, X., Lyu, J., and Xie, J. (2023). A Comprehensive Evaluation of Tomato Fruit Quality and

Identification of Volatile Compounds. *Plants* **12**: 2947. <https://doi.org/10.3390/plants12162947>

Zhao, M., Li, J., Chang, P., Li, L., and Zhang, L. (2019). Identification and characterization of MYB-bHLH-WD40 regulatory complex members controlling anthocyanidin biosynthesis in blueberry fruits development. *Genes* 10: 296.
<http://dx.doi.org/10.3390/genes10070496>

Zhu, F., Wen, W., Cheng, Y., and Fermie, A.R. (2022) The metabolic changes that effect fruit quality during tomato fruit ripening. *Mol. Horticulture* 2: 2
<https://doi.org/10.1186/s43897-022-00024-1>

Zhu Y., Ling W., Guo H., Song F., Ye Q., Zou T., Li D., Zhang Y., Li G., Xiao Y., Liu F., Li Z., Shi Z., and Yang Y. (2012). Anti-inflammatory effect of purified dietary anthocyanin in adults with hypercholesterolemia: a randomized controlled trial. *Nutr Metab Cardiovasc Dis.* **23**:843–849.
<https://www.sciencedirect.com/science/article/abs/pii/S0939475312001573>

Appendix 1: Complete sequence of the transgenic insert and flanking genomic regions

Annotation of the *Del/Ros1* T-DNA insertion. Abbreviations are as listed in Table 1. Figure 3 shows an overall representation and orientation of the *Del/Ros1* T-DNA insert.

Nucleotide position		Genetic Element
Start	End	
1	332	Tomato chromosomal DNA ("chr4") position 62904346...62904677 (Build SL 3.0). 94 bp of genomic sequence after position 62904677 have been deleted
333	864	RB region of the T-DNA insert 52 bp at the end of the RB region have been deleted (including the RB)
865	1502	CMV (Cauliflower mosaic virus termination region)
1615	2277	ROS (<i>Rosea1</i> cDNA from snapdragon)
2284	4470	E8p (E8 promoter region from tomato)
4500	5227	CMV (Cauliflower mosaic virus termination region)
5287	7221	DEL (<i>Delila</i> cDNA from snapdragon)
7305	9493	E8p (E8 promoter region from tomato)
10535	11242	Ocs 3 (Octopine synthase termination region)
11269	12063	NPT II (Neomycin phosphotransferase gene conferring resistance to Kanamycin)
12150	12329	NOSp (Nopaline synthase promoter region)
12330	12915	LB region of the T-DNA insert 75 bp at the end of the LB region have been deleted (including the LB)
12916	13098	Tomato chromosomal DNA ("chr4") position 62904771...62904953 (Build SL 3.0)

Nucleotide sequence of the *Del/Ros1* T-DNA locus including tomato genomic flanking DNA. Tomato genomic DNA sequence is shown in lower case letters while the sequence of the T-DNA insert is shown in upper case letters. The nucleotide sequences of the different genetic elements are color coded to match the corresponding color of the elements listed above.

```

1  atacgtaaaagggtcaaataattattatttaaatttttagaaaaatagattaagaaaataaatgaataagta      70
71  ggtaagtcgatgacaaaataaataaggaagaagaggaaataaataggaagaagggttgtaaccgtaaaa      140
141  ggggtcaaataattattatttaaattttttctcacttacgttaaaggtaagcaattttcatgaatctttaatca      210
211  acatttttatgtttaaatgacgcatgagaaatgaggttgagaattaaaaatagaacaagattaaatcggag      280
281  catcaagacaaaatgtatgagcaaattaaagaggtgatgtatgtattgggccCGAATAGTTTGAAATTAG      350
351  AAAGCTCGCAATTGAGGTCTACAGGCCAAATTCGCTCTTAGCCGTACAATATTACTCACCGGTGCGATGC      420
421  CCCCCATCGTAGGTGAAGGTGGAAATTAATGATCCATCTTGAGACCACAGGCCACACAGCTACCAGTT      490
491  TCCTCAAGGGTCCACCAAAAACGTAAGCGCTTACGTACATGGTTCGATAAGAAAAGGCAATTTGTAGATGT      560

```

561 TCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCAATTAATGCAGCTGGCACGACAGGTTTCCC 630
631 GACTGGAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCCAGGCTT 700
701 TACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAG 770
771 CTATGACCATGATTACGCCAAGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTGGTACGTACCGGG 840
841 CCCCCCTCGAGATATCGCATGCGATCCCCGTCACCGGTGTGAGGGAAGTAGTTTTGATCTTGAAAGATC 910
911 TTTTATCTTTAGAGTTAAGAACTCTTTCGTATTTTGGTGAGGTTTTATCCTCTTGAGTTTTGGTCATAGA 980
981 CCTATTCATGGCTCTGATACCAATTTTTAAGCGGGGGCTTATGCGGATTATTTCTTAAATTGATAAGGGG 1050
1051 TTATTAGGGGGTATAGGGTATAAATACAAGCATTCCCTTAGCGTATAGTATAAGTATAGTAGCGTACCTC 1120
1121 TATCAAATTTCCATCTTCTTACCTTGACAGGGCCTGCAACCTTATCCTTCCTTGTCTTCCTCCTTCCTT 1190
1191 CCGTCCACTTCATCATATTTAAACCAAACCTACGGGGGAGTCAACGTAACCAACCTGCCTTAGCATCTT 1260
1261 TTCCCTAACGGCCTCCTGCCTAAGCGGTACTTCTAGCTTCGAACGGCGTCTGGGCTCCAGGTTTAGTCGT 1330
1331 CTCGTGTCTGGTTTATATTCACGACAAAGATCTATAGGGACTTTAGGAGATCTGGATTTTAGTACTGGAT 1400
1401 TTTGGTTTTAGGAATTAGAAATTTTATTGATAGAAGTATTTTACAAATACAAATACATACTAAGGGTTTC 1470
1471 TTATATGCTCAACACATGAGCGAAACCTATAAGAACCCTAATTTCCCTTATCGGGAACTACTCACACA 1540
1541 TTATTTATGGAGAAAATAGAGAGAGATAGATTTGTAGAGAGAGACTGGTGATTTACGCGTACCGAATTCC 1610
1611 CGGGTTAAATTTCCAATTTGTTGGGCCTCCTCGAATAGGTTTCTTAATCCCCATCCTCCGTTGTTTCTAG 1680
1681 CAACTTACTCCACCCTGAATGCAGTCTTCAACTTCATCTTGTGGCGACGCAACATCATGTGAAAATTGC 1750
1751 GTTTGCTTCTCACAATCTGGAATCTCATCAGTTGTTAACCGGACATTTGAAAATTCATCGGTTTTTCCGA 1820
1821 CTTCTCTCGGCCAAGTAACGTGCAATCCGGTGAAGGTCCGAGCTCGGGGTCTTACGATATTAGTCAGCTT 1890
1891 AATGGTTTTTGTGTTTCATAACATTTTTCCGGCATCGTTCTCCATCCTCGCCTAAATTTCTCCCCACATGA 1960
1961 GTATTCCAAAAGTTCTTCACGTCATTAGCTGTCCTTCCAGGAATTTCTACCAGCAATCAGCGACCATTGT 2030
2031 TACCCAACAGCTTATGAAGCCTCACAATTAGGTCCACTTCATCTCTCGAAAACCGACCTCTTTTGATATT 2100
2101 TGGCCTCAGATAATTCAACCACCTCAGCTGCAACTCTTCTACACCGGTTCAACCCTGCTCTGTGTGGA 2170
2171 ACTTGATGCCATTTCCCTTACCATACTCTTCTATACATTGCCTCAAGAGAGTGTCTTCTTCTTTGGTCC 2240
2241 AAGTACCTTTTCTCACTCCACGACAATTCTTTTCCATGGATCCGCACGTGAATGATTAGAATAATTTCT 2310
2311 AAAAATCCCAATATGAGGATGCCATATTTATAATAGAATAAAAATAAAATGTGAACAAAGAAAGAGATAAA 2380
2381 GTAGTTCACTTTTTGAAATCTAAGAGAGAAATGGGAACAAGAAAGAGACAAAAGTAGTTTCAAACAAACT 2450
2451 TCTCTTCTAAGTTTAGTCCCTTTTAAATATGAAACCAATACGCTCGATTAAAGAATAGAAAAATATCA 2520
2521 AATTTTCAATATAATTTATACTAATCGTTTGAATTTTCTACTGATATAGTGACGTTTCATCATAAC 2590
2591 AACCAAAACGTTGTTGCTTCACAACAATAATATAGTAGTAGTTAATTTATTATTTAGTAATAAGTGGTCC 2660
2661 TAAAAATTAGATAAATATTACTATGATAATATAAAAAATATTTGAGTCAGTCCTAAAAAATTTATTTAGTAT 2730
2731 TCATACATGAATCAAACGATTAGTTAAGTGTCAACAATTGGACAAGTGGCATGGAGGTTGTAAAAGAAT 2800
2801 GACATAAGCCAACCTGCTATTTTTATCCAAAAAAGAAGACAACCTTGACAACCTACATTTCTTTTATTTTT 2870
2871 ATAAATTTACTAATATCTTCTATGCAAAATATTCCGGTGCCTTTCTAAACTTTAAGGTTTTTATTTGATG 2940
2941 TACACCTAAATTATATTTTATTTTAATCACTCCACTGGACTTGTTTTATTCCTTCATCATATACACCTACT 3010
3011 CCTATTATGACTACAAGTTGGCAAAAGTAATGATATGAATTTCTACTTAAATAAATAATAGTCACCTAGA 3080
3081 TAAATTAATTTAACAAAAGATAAATATCAAACCTTCTCACCTAAATTTTGAGCAAACTTCTCACATAA 3150
3151 ACTTGTGGACTAAACCCGAAAAATCTTCAGAAAAATTAATATTTAGTACTGGAAAAGTCAGATTAAATGTC 3220
3221 GCACAAGACTTTCTATTGTTGGGAATAAAACAAATTAATATTGGATTAAATAGTTGAAATATTTAGGTA 3290
3291 AAATGCTACATGTCATTTATTCATTGGATATTATTTCTTAAATTTAAATTCATTATTTAAAGTTATT 3360
3361 TTTGAATAAGGGCCGATTTTCGTGAAATTCCTTCTAGATATGGGTCTTCTAGACGTAAAGTTGATCTATT 3430
3431 AAATTTTAAATTTATCTTAAATTTTACAAAAGTAAGTATTAATCTTTGTTTCCTTTACTATTTCATTTACA 3500
3501 TTTTGTCTATATTTTCGTTTAAATATGTCATATATTAAAAAAATTAATAATTTTACTTTCTTTTTTTA 3570

3571	CATTATAGCTATATGACGTGACAAAAATCAACTTTACATGCGCCTAGTAGACTTCAAGTTAAAAGGGG	3640
3641	ATAATGGATACTTTGCCTATCTTTTACCATATATTTTAAAAATCCTTAATTATTAAGTTTTCCAATATCTC	3710
3711	TCACCATTCAATTTTCTCCTATCATATATTTTAGGAGTCCTTAATAATTAAGTTTACTAATAAACTTTATT	3780
3781	ATATATTATAGGACTCCTCAATTATTAGTTCTCTTTATGTCTCTCATCGTACATTTTCTCTTGTCTTAT	3850
3851	TTGTTAGGACACTTGAAATTTTCAAAATATATTTTGCCTTTAATATATGAAGTTGTGTTGATTGTAGTT	3920
3921	TTTGTAATATATTTAATTTTTTGAATTTTTATTTTCTAAAAGAAACATAAAATTTAAAAGATTTAAAAG	3990
3991	TATCATTTAACTATTAGAAATAATATATCTATGTTGTTAAAAATGATGGTCTTAATTAAGTGTTTTATT	4060
4061	ATAAAATATCAGATAATTCGTTTTATTTACGCAAAAGTTAAGTGAAGTAACGAAATTATAAATCCCATAG	4130
4131	AATATTGTGTATATACCTTGGCACATGATGATTGTAACATCCTTAATTATTATTAATTCATCGAACCTATT	4200
4201	ATTTCTTCATTGTCTATGTACATTTATCCTTAATAATCCACTTCAGGATTTATTAGTTCTTTGGTTATT	4270
4271	GGTTTAAGTTTATTTTACAACCAAGTGAATTGAATTTGTCTCCATTAATATTTATTGGATTAAAAATA	4340
4341	AATAAATTTGCTCTTATTTGTAGAAAGATTTAGACTTTTAAAAATATTACGTTTTCTGACTCTTTTCTTAT	4410
4411	CAAAATTGGACTCTCTCACTTCCACAAAACCTTAATTACGTGAACAATATCATTAGGGATGGGTACCTATC	4480
4481	TGTCGAGATATCGCATGCCGATCCCCGTCAACGGTGTGAGGGAAGTAGTTTGGATCTTGAAAGATCTTTTA	4550
4551	TCTTTAGAGTTAAGAAGTCTTTCGTATTTGGTGAGGTTTATCCTCTTGAGTTTGGTCATAGACCTAT	4620
4621	TCATGGCTCTGATACCAATTTTTAAGCGGGGGCTTATGCGGATTATTTCTTAAATTGATAAGGGGTATT	4690
4691	AGGGGGTATAGGGTATAAATACAAGCATTCCCTTAGCGTATAGTATAAGTATAGTAGCGTACCTCTATCA	4760
4761	AATTTCCATCTTCTTACCTTGCACAGGGCCTGCAACCTTATCCTTCCTTGTCTTCCCTCCTTCCGTC	4830
4831	CACCTTCATCATATTTAAACCAAACCTACGGGGGAGTCAACGTAACCAACCTGCCTTAGCATCTTTTCCC	4900
4901	TAACGGCTCCTGCCTAAGCGGTACTTCTAGCTTCAACGGCGTCTGGGCTCCAGGTTTAGTCGTCTCGT	4970
4971	GTCTGGTTTATATTACGACAAAGATCTATAGGGACTTTAGGAGATCTGGATTTTAGTACTGGATTTTGG	5040
5041	TTTTAGGAATTAGAAATTTTATTGATAGAAGTATTTTACAAATACAAATACATACTAAGGGTTTCTTATA	5110
5111	TGCTCAACACATGAGCGAAACCTTATAAGAACCCTAATTTCCCTTATCGGGAACTACTCACACATTATT	5180
5181	TATGGAGAAAAATAGAGAGAGATAGATTGTAGAGAGAGACTGGTGATTTTCAGCGTACCGAATTCCCATCG	5250
5251	AACCACTTTGTACAAGAAAGCTGGGTGGATCCAACCTCAAGACTTCATAGTAACTTTCTGAAGAGCTTGT	5320
5321	TTGATCACACTTGCTGATGCAACCTTCAATCCCTTGCCTTGGCTTTTATGGTAATGGATATAATTCCAT	5390
5391	CTCTGTTGGAAGATTGAACAGTTTCGGAATCCAAACTTAGTCGTCTTACGGCTTCCATCACTTCAAGCAA	5460
5461	TACAAACTCCTTGGAAGAACAACCTCACGACAATCAACACATCCTTGTGTAATGTTTACAGTTATATTA	5530
5531	TCTGTTAAGGAATCTTTCAATCGACCTCTGCTATTTACGGCTCCAATCTGTCCGTATCAGAAGCCTTTC	5600
5601	TCTTGTGTTGTCACGGTTTCTTGACATTACTTGTCTTGTGTGCCATAATTATCGGAGGTCCTCTCAAT	5670
5671	GGCATCGTGCAGTTTAGTTTGTAGTTGATTCCCGCCCCCGGCCCTTACCATTTTGTAGATTCCAGC	5740
5741	TCGTCGACTTTCCTCTCAAGCCCTCTCAAGTAATCTATTGTATGGTCTAGTATTGATACTTTGTCAACCT	5810
5811	TGCCACCGGATGGGACTAGGGATGCAAGAATCATAAACCGTTCGTTTATTTCTCTCTGCGTTTTCTCTC	5880
5881	TGACAAGACGTGGTTTTCTATCAATTTTCATCAGCCGTTGGCTTTGCAAGGCAGTCACGTGTGCCTTTTTGT	5950
5951	TTACCAGCATCAAGCCTGGAGTTTTTCATGCATTCTAGCTACTTCAAAAAGTACTTTCTTCAGAAATCTTT	6020
6021	GTGAGGTTCCGCTTCGGGGAACATGAGTACCCGACGATCCATCCTTGTTCCTCAACTAACGAAGCTTGATTC	6090
6091	TCTATTCCCATTTCTGAAGTAGGGACCAAGAACCAACTGATGGGAACCTTCAAAAAGATTGGAAAGTACC	6160
6161	CTTTGATAATGGACCTCATCGCCTTGGACACCCGTGTTTTCTATCTGCTGATTGCATTCTTGAGTGCCTGT	6230
6231	GCATACAATTATTTCGTCTCCGGTGCCCTTTTCCATCAGAAAGTGGGGCAAAGACTCTAGATTTTCATG	6300
6301	AGTTTGAGATATACAGTCACTAGAATTATAGAAGTATTGAGACAATTGCTGATTGCATCATCCATGAAA	6370
6371	GGCCAGCTTTGTGTTTGAGGAACCTCCCCATTGATGCCTTCTGCCAAATTCGATTCTGTCTATGAGTAAAT	6440
6441	TGTCTGCAAAGTCATCCAACTGTTATCAGGAGAACATATGTTTCGTGTCTGGACAATTCAAAAAGCTGATC	6510
6511	AAGATCGTTTTCTGGTATATTAGCATGTTCAAGCGCTTCACAAATGAGGTCAATTGTTGTTTGTAACTG	6580

6581	TTGGAGACATAGTTGGGAATCTTGGGAACGGTGGCAGGACTGTCCAAGAATGAAGTTTTTATATGCTGGA	6650
6651	TTAGATTCAAATCCTCCGGTACTAGCTCTGTTGCTCCCAGCTCAACTACTCCTTCTGCATATGGAAGCA	6720
6721	CACAACGTCTGAATTGACGCACTCTTTGCAAGCAAAGAACGCGAGAAAACCTTGGTGTCCGCACGATGA	6790
6791	GCGTTGCATAGCCATACTGCTTGATTTCTGTGCTAATGTTCTTCCAGGCAACCCCTGGCCAATATTGAATA	6860
6861	TGAAAGACATGCAAACCAAGAAAAACCACTCAGCATCAGTGAGGTCTTCTGGTGATAATGCAGCAGTAGG	6930
6931	CCTTTTAGCTTGTGTGTTGGTTTCACCAAGTGAAAGAGACTCATAAAGTTCTCTCAATTGATCACTTCTC	7000
7001	TGCAATCCCAGCTGATCGGGATTCAATTCGACAGATTGTACAGTTTTTCGAGTTTTAATATCTCCATTGT	7070
7071	AGAACCCATCACCCCACTCCAAGACCCCTGGTTGTGCAACTGAATTGGACCAGAAAAATGCATAACTCCA	7140
7141	TTGGATACTTCTAACAGCAATAGCAAGTTGCTTCTCTCAAATTCTCAGGCACTATCTTTTGGTTTTGGATA	7210
7211	CCAGTAGCCATGGTAGCCTGCTTTTTTGTACAACTTGTTTGATGGGGATCCGTCGACCTGCAGGCGGCC	7280
7281	GCGAATTCAGTAGTGATTCTGGGATCCGCACTGTGAATGATTAGAATAATTTCTAAAAATCCCAATATGAG	7350
7351	GATGCCATATTTATAATAGAATAAAATAAAATGTGAACAAAGAAAGAGATAAAGTAGTTCACTTTTTGAA	7420
7421	ATCTAAGAGAGAAATGGGAACAAGAAAGAGACAAAAGTAGTTTCAAACAACTTCTCTCTAAGTTTAGT	7490
7491	CCCTTTTTAAAAATATGAAACCCAATACGCTCTGATTAAGAATAGAAAAATATCAAATTTTCAATATAATTT	7560
7561	ATACTAATCGTTTTGAATTTTTCATACGTATAGTGACGTTTCATCATAACAACCAAAACGTTGTTGC	7630
7631	TTCAACAATAATATAGTAGTAGTTAATTTATTATTTAGTAATAAGTGGTCCTAAAAATTAGATAAATA	7700
7701	TTACTATGATAATATAAAAAATTTTGAGTCAGTCCTAAAAATTATTTAGTATTCATACATGAATCAAAC	7770
7771	TGATTAGTTAAGTGTCAACAATTGGACAAGTGGCATGGAGGTTGTAAAGAATGACATAAGCCAACCTGCT	7840
7841	ATTTTTATCCAAAAAAGAAGACAACCTTGACAACCTACATTTCTTTTATTTTTATAAATTTACTAATATC	7910
7911	TTCTATGCAAAATTATTCGGTGCCTTTCTAAACTTTAAGGTTTTTATTTGATGTACACCTAAATTATATT	7980
7981	TTATTTTAATCACTCCACTGGACTTGTTTATTCCTTCATCATATACACCTACTCCTATTATGACTACAAG	8050
8051	TTGGCAAAAGTAATGATATGAATTTCTACTTAAATAAATAATAGTCACCTAGATAAAATTAATTTAACAAA	8120
8121	AGATAAATATCAAACCTTCTCACCTAAAAATTTTGAGCAAACTTCTCACTAAAACTTGTGGACTAAACCC	8190
8191	GAAAATCTTCAGAAAATTAAATTTTAGTACTGGAAAAGTCAGATTAAATGTCTGCACAAGACTTTCTATT	8260
8261	GTTGGGAATAAAACAAATTAATATTGGATTAAAATAGTTGAAATATTTAGGTAAAATGCTACATGTCATT	8330
8331	TATTCATTGGATATTATTTCTTAAATTTAAATTCATTATTTAAAAGTTATTTTTGAATAAGGGCCGAT	8400
8401	TTCTGTAAATTCCTTCTAGATATGGGTCTTCTAGACGTAAGTTGATCTATTAAATTTTAAATTTATCT	8470
8471	TAAATCTTTACAAAGTAAGTATTAATCTTTGTTTCTTTACTATTCAATTTACATTTTGTCTATATTTCTG	8540
8541	TTTAAATATGTCATATATTAAAAAAATTAATAATTTTACTTTCTTTTTTACATTATAGCTATATGAC	8610
8611	GTGACAAAAATCAACTTTTACATGCGCCTAGTAGACTTCAAGTTAAAAGGGGATAATGGATACTTTGCC	8680
8681	TATCTTTTACCATATATTTTAAATCCCTTAATTATTAAGTTTTCATATCTCTCACCATTCAATTTCTC	8750
8751	CTATCATATATTTTAGGAGTCCTTAATAATTAAGTTTACTAATAAACTTTATTATATATTATAGGACTCC	8820
8821	TCAATTATTAGTTCTCTTTATGTCTCTCATCGTACATTTTCTCTTGTCTTATTTGTTAGGACACTTGAA	8890
8891	ATTTTCAAATATATTTTGTCTTTAATATATGAAGTTGTGTTGATTGTAGTTTTTGTAAATATATTTAA	8960
8961	TTTTTTGAATTTTTATTTTCTAAAAGAAACATAAAATTTAAAAGATTTAAAAGTATCATTAACCTATTAG	9030
9031	AAATAATATATCTATGTTGTTAAAATGATGGTTCTTAATTAAGTTTATTATATAAAATATCAGATAAT	9100
9101	TCGTTTTATTTACGCAAAAGTTAAGTGAAGTAACGAAATTATAAATCCCATAGAATATTGTGTATATACT	9170
9171	TGGCACATGATGATTGTAACATCCTTAATTATTATTAATTCATCGAACCTATTATTTCTTCATTGCTAT	9240
9241	GTACATTTATCCTTAATAATTCCACTTCAGATTTATTAGTTCCTTTGGTTATTGGTTTAAAGTTATTTTA	9310
9311	CAACCAAGTGAATTGAATTTGTCTCCATTAATATTTATTGGATTAAAAAATAAATAAATTTGCTCTTAT	9380
9381	TTGTAGAAAGATTTAGACTTTTAAATATTACGTTTTCTGACTCTTTTCTTATCAAAATTGGACTCTCTC	9450
9451	ACTTCCACAAAACCTTAATTACGTGAACAATATCATTAGGGATGGGTACCGAGCTCCAATTCGCCCTATAG	9520
9521	TGAGTCGTATTACAATTCAGTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCGTCGTTACCCAA	9590

9591	CTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCC	9660
9661	CTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGG	9730
9731	TGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAAACATCCAACGTCGCTTTCAGGGATCAGCTTT	9800
9801	TGCCATTCTCACCGGATTCACTCGTCACTCATGGTGATTCTCACTTGATAACCTTATTTTGTACGAGGG	9870
9871	GAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATAACCAGGATCTTGCCATCCTA	9940
9941	TGGAAC TGCC TCGGTGAGTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATC	10010
10011	CTGATATGAATAAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTT	10080
10081	GTAACACTGGCAGAGCATTACGCTGACTTGACGGGACGGCGGCTTTGTTGAATAAATCGAACTTTTGCTG	10150
10151	AGTTGAAGGATCAGATCACGCATCTTCCCGACAACGCAGACCGTTCCGTGGCAAAGCAAAAGTTCAAAAT	10220
10221	CACCAACTGGTCCACCTACAACAAAGCTCTCATCAACCGTGGCTCCCTCACTTTCTGGCTGGATGATGGG	10290
10291	GCGATT CAGCCTGGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGCCCCCCCCCCCCCTG	10360
10361	CAGGTCGACCAAGCGGCCATCGTGCC TCCCCACTCCTGCAGTTCGGGGGCATGGATGCGCGGATAGCCG	10430
10431	CTGCTGGTTTCCTGGATGCCGACGGATTTGCACTGCCGGTAGAACTCCGCGAGGTCGTCCAGCCTCAGGC	10500
10501	AGCAGCTGAACCAACTCGCGAGGGGATCGAGCCCC CTGCTGAGCCTCGACATGTTGTGCGAAAATTCGCC	10570
10571	TGGACCCGCCCAACGATTTGTGCTCACTGTCAAGGTTTGACCTGCAC TTCATTTGGGGCCACATACACC	10640
10641	AAAAAATGCTGCATAATTCTCGGGGCAGCAAGTCGGTTACCCGGCGCCGTGCTGGACCGGGTTGAATG	10710
10711	GTGCCCGTAACTTTCGGTAGAGCGGACGGCCAATACTCAACTTCAAGGAATCTCACCCATGCGCGCCGGC	10780
10781	GGGGAACCGGAGTTCCCTTCAGTGAGCGTTATTAGTTCCGCCGCTCGGTGTGTCGTAGATACTAGCCCTG	10850
10851	GGGCAC TTTTGAAATTTGAATAAGATTTATGTAATCAGTCTTTTAGGTTTGACCGGTTCTGCCGCTTTT	10920
10921	TTAAATTTGGATTTGTAATAATAAAACGCAATTGTTTGTTATTGTGGCGCTCTATCATAGATGTCGCTAT	10990
10991	AAACCTATT CAGCACAATATATTGTTTTCATTTTAATATTGTACATATAAGTAGTAGGTACAATCAGTA	11060
11061	AATTGAACGGAGAATATTATT CATAAAAAATACGATAGTAACGGGTGATATATT CATTAGAAATGAACGAA	11130
11131	ACCGGCGGTAAGGATCTGAGCTACACATGCTCAGGTTTTTTACAACGTGCACAACAGAATTGAAAGCAA	11200
11201	TATCATGCGATCATAGGCGTCTCGCATATCTCATTAAGCAG GACTCTAGCGAACCCAGAGTCCCGCT C	11270
11271	AGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCAC	11340
11341	GAGGAAGCGGT CAGCCCATT CGCCGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCTGA	11410
11411	TAGCGGTCCGCCACACCCAGCCGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATAT	11480
11481	TCGGCAAGCAGGCATCGCCATGAGTCACGACGAGATCCTCGCCGTCGGGCATGCGCGCCTTGAGCCTGGC	11550
11551	GAACAGTT CGGCTGGCGCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCC	11620
11621	ATCCGAGTACGTGCTCGCTCGATGCGATGTTTCGCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAAGCG	11690
11691	TATGCAGCCGCCGATTCATCATCAGCCATGATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAGGAG	11760
11761	ATCCTGCCCCGGCACTTCGCCCAATAGCAGCCAGTCCCTTCCCGCTTCAGTGACAACGTCGAGCACAGCT	11830
11831	GCGCAAGGAACGCCCGTCGTGGCCAGCCACGATAGCCGCGCTGCCTCGTCTGCAGTTCATTCAGGGCAC	11900
11901	CGGACAGGTCGGTCTTGACAAAAAGAACCGGGCGCCCCTGCGCTGACAGCCGGAACACGGCGGCATCAGA	11970
11971	GCAGCCGATTGTCTGTTGTGCCAGTCATAGCCGAATAGCCTCTCCACCAAGCGGCCGAGAACCTGCG	12040
12041	TGCAATCCATCTTGTTCAACCAT GGTAATTGTAAATGTAATGTAATGTTGTTGTTGTTGTTGTTGTT	12110
12111	GGTAATTGTTGTAAAAATAGGAGAGTGAATATGAGACTCT TAATTGGATACCGAGGGGAATTTATGGAACG	12180
12181	TCAGTGGAGCATTTTTGACAAGAAATATTTGCTAGCTGATAGTGACCTTAGGCGACTTTTGAACGCGCAA	12250
12251	TAATGGTTTTCTGACGTATGTGCTTAGCTCATTAAC TCCAGAAACCGCGGCTCAGTGGCTCCTTCAACG	12320
12321	TTGCGGTT CTGT CAGTTCCAAACGTAAACGGCTTGTCGCCGCTCATCGCGGGGGTCATAACGTGACTC	12390
12391	CCTTAATTCTCCGCTCATGATCATCCGATGGGTGGATTTATCACAAATGGGACCCGCCGCCGACAGAGGT	12460
12461	GTGATGTTAGGCCAGGACTTTGAAAATTTGCGCAACTATCGTATAGTGGCCGACAAATTGACGCCGAGTT	12530
12531	GACAGACTGCCTAGCATTGAGTGAATTATGTAAGGTAATGGGCTACACTGAATTGGTAGCTCAAACGTGT	12600

12601	CAGTATTTATGTATATGAGTGTATATTTTCGCATAATCTCAGACCAATCTGAAGATGAAATGGGTATCTG	12670
12671	GGAATGGCGAAATCAAGGCATCGATCGTGAAGTTTCTCATCTAAGCCCCCATTTGGACGTGAATGTAGAC	12740
12741	ACGTCGAAATAAAGATTTCCGAATTAGAATAATTTGTTTATTGCTTTCGCCTATAAATACGACGGATCGT	12810
12811	AATTTGTCGTTTTATCAAAATGTACTTTCATTTTATAATAACGCTGCGGACATCTACATTTTGAATTGA	12880
12881	AAAAAATTGGTAATTACTCTTTCTTTTCTCCATatttagttgtgccccgtagcaaattgtttgtcaat	12950
12951	catctctctccctcaaactctcgctctatctctcgttcgcttcctctcgcttttatacaaacacaagtgt	13020
13021	ataaaatttgTTTTtattatataaagcgagagaaaattgtataaatacatatatTTTTgttttcctctgc	13090
13091	cagatctc	13098

Appendix 2: Integrity of T-DNA Insert in Del/Ros1-N Event; absence of vector backbone sequence

To confirm that Event Del/Ros1-N carried no sequences from the plasmid vector outside the T-DNA, the genome of a single Del/Ros1-N MicroTom plant (generation >6) was sent for resequencing by BGI¹⁵, under project number F18FTSSCKF3306_LYCrghR. Illumina HiSeq X Ten sequencing gave >60-fold coverage. The genome sequence of *Del/Ros1-N* has been deposited as a BAM file in the European Nuclear Archive (ENA) under accession number ERR3500875. Details of the sequencing follow.

Whole genome resequencing by BGI for Del_Ros1_MicroTom_2A

Project No.: F18FTSSCKF3306_LYCrghR

1 Sequencing assessment

According to the contract terms of Project No. F18FTSSCKF3306_LYCrghR, 3 paired-end sequencing libraries with the insert size around 300bp were constructed. In total, 37.17G clean data (valid data) was yield after a three-step raw data (sequencing produced data) filtration:

- 1) Adaptor trimming. Any sequencing read with adaptor mapping rate higher than 50% is removed.
- 2) Low quality reads trimming. Any sequencing read which consists of more than 50% of low-quality bases (Q20<50%) is removed.
- 3) Contiguous N bases trimming. Any sequencing read with over 2% of N base read is removed.

Table 1 shows the data quality of filtered data. 98.59% of total clean reads have quality value higher than 20 (Q20 ≥ 98.59%), which indicates a quite high data quality.

Table 1 Quality evaluation of filtered data

Sample	Library	GC _rate(%)	Q20_ rate(%)	Q30_ rate(%)	Reads(M)	Bases (G)	Clean_Data/ Raw_Data(%)
Del_Ros1_Mi croTom_2A	wHAXPI093150-68	36.05	98.59	96.01	247.78	37.17	90.33

Sample: Sample Name

Library: Library Name

GC_rate (%): The proportion of bases that are G or C in the sequencing reads.

Q20_rate (%): The percentage of bases in sequencing reads with a quality score ≥ 20 out of the total bases.

Q30_rate (%): The percentage of bases in sequencing reads with a quality score ≥ 30 out of the total bases.

Reads (M): The total number of sequencing reads (for paired-end sequencing, it refers to the number of read pairs).

Bases (G): The product of the number of reads and the read length, representing the total number of bases sequenced.

Clean_Data/Raw_Data (%): The percentage of data retained after filtering compared to the raw data, indicating the proportion of valid data.

2 Sequencing depth assessment

We used the GCF_000188115.3_SL2.50 genome version of tomato from NCBI as the reference genome and aligned all sequencing data to it using the BWA software. The genome size is 824,257,102 bp, with an effective genome size of 737,791,809 bp (excluding Ns from the reference sequence), and a GC content of 30.48% for the reference genome. Additionally, to facilitate analysis, we merged numerous small scaffold reference sequences into one or several

¹⁵ <https://www.bgi.com/global/science-detail/plant-animal-and-microbe-whole-genome-resequencing>

large chromosomes and renamed them. The alignment rate for the samples is 97.92%, with an effective sequencing depth 45.09X.

Sample	Coverage_rate(%)	Map_reads_rate(%)	Map_bases_rate(%)	Uni_hit_reads_rate(%)	Uni_hit_bases_rate(%)	Sequencing_depth	Effective_depth
Del_Ros1_MicroTom_2A	88.64	97.92	97.92	65.22	63.86	45.09	45.09

Sample: Sample name

Coverage_rate: Coverage rate, indicating the proportion of the genome covered by sequences obtained from sequencing

Map_reads_rate: Read alignment rate, calculated as the number of reads aligned to the reference genome divided by the total number of reads in the clean data. If the sequencing sample is contaminated or has significant differences from the reference genome, a low alignment rate may affect subsequent information analysis

Map_bases_rate: Base alignment rate, calculated as the number of bases aligned to the reference genome divided by the total number of bases in the clean data

Uni_hit_bases_rate (reads): Unique alignment rate by reads, calculated as the number of reads aligned to a unique position in the reference genome divided by the total number of reads aligned to the reference genome

Uni_hit_bases_rate (bases): Unique alignment rate by bases, calculated as the number of bases aligned to a unique position in the reference genome divided by the total number of bases in the clean data

Sequencing_depth: Sequencing depth, calculated as the number of bases aligned to the reference genome divided by the genome size

Effective_depth: Effective sequencing depth, calculated as the number of bases aligned to the reference genome divided by the effective genome size (excluding Ns in the reference sequence)

A bioinformatics approach was applied to determine whether the genome of our Del/Ros1-N plants contained any sequences of the vector backbone (sequences of the original transformation vector that should not be inserted into the genome), by searching for sequences matching the 20 kb of vector backbone with the entire genome of Del_Ros1_MicroTom, in 150 nucleotide windows with 100% match, using default parameters. Specifically, Illumina genome reads from Del_Ros1_MicroTom_2A to Prk290_backbone sequence (the original transformation vector) were mapped using bowtie2 version (2.2.9; <https://github.com/BenLangmead/bowtie2>) and the resulting mapping file was sorted and indexed using samtools version 1.5 (<https://github.com/samtools>) and visualized using Interactive Genomic Viewer (version 2.4.6). Alignments were deposited in the European Nuclear Archive (ENA) under accession numbers ERR3502329, ERR3502327 and ERR3500875. The T-DNA portion of the vector was inserted with 100% match to the expected sequence (no mutations or mismatches).

This search confirmed that no sequences matching the backbone of the vector, including the bacterial tetracycline resistance marker, are present in Del/Ros1-N lines (Figure 1).

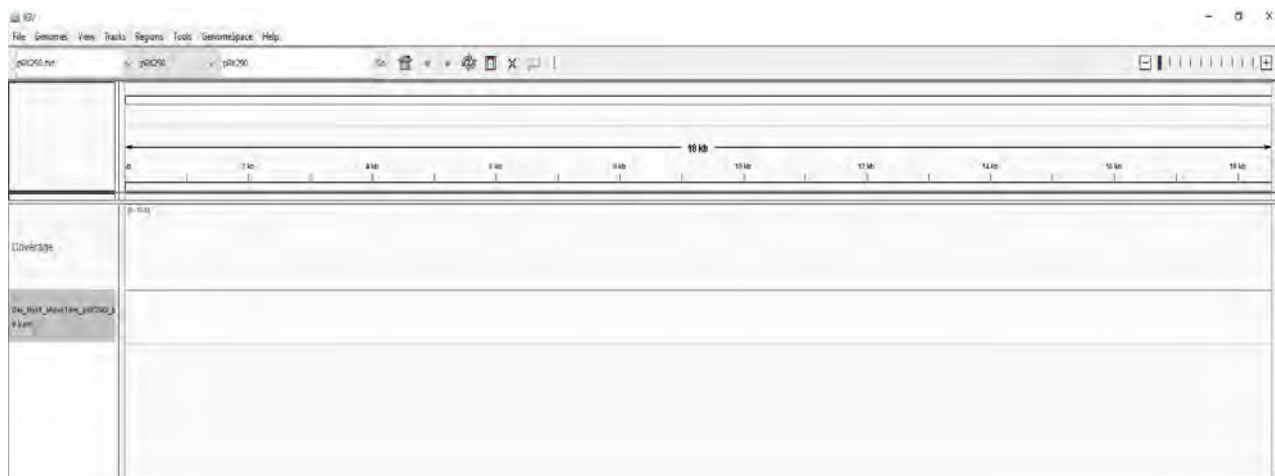


Figure 1: Results of search of *Del/Ros1-N* genome sequence for sequences of 150 nucleotides or greater with identity to pRK290 sequence taken from CLD04541: nucleotides 5854-25385 deposited in NCBI as AF184978.1. (<https://www.ncbi.nlm.nih.gov/nuccore/AF184978.1>). *Del/Ros1-N* genome sequence has been deposited in ENA accession number ERR3500875. No reads aligned to this sequence with 100% identity over 150 nucleotides as shown in the alignment deposited in ENA under accession number ERR3502329.

Using this same genome sequence, the integrity of the *Del* and *Ros1* transgenes inserted in the T-DNA in the *Del/Ros1-N* event was examined. With more than 20 sequences aligned over all parts of each gene no consistent mismatches (mutations) in either transgene were found, confirming that transformation had not involved mutations of the genes encoding the *Del* and *Ros1* proteins. (**Figures 2 and 3**).

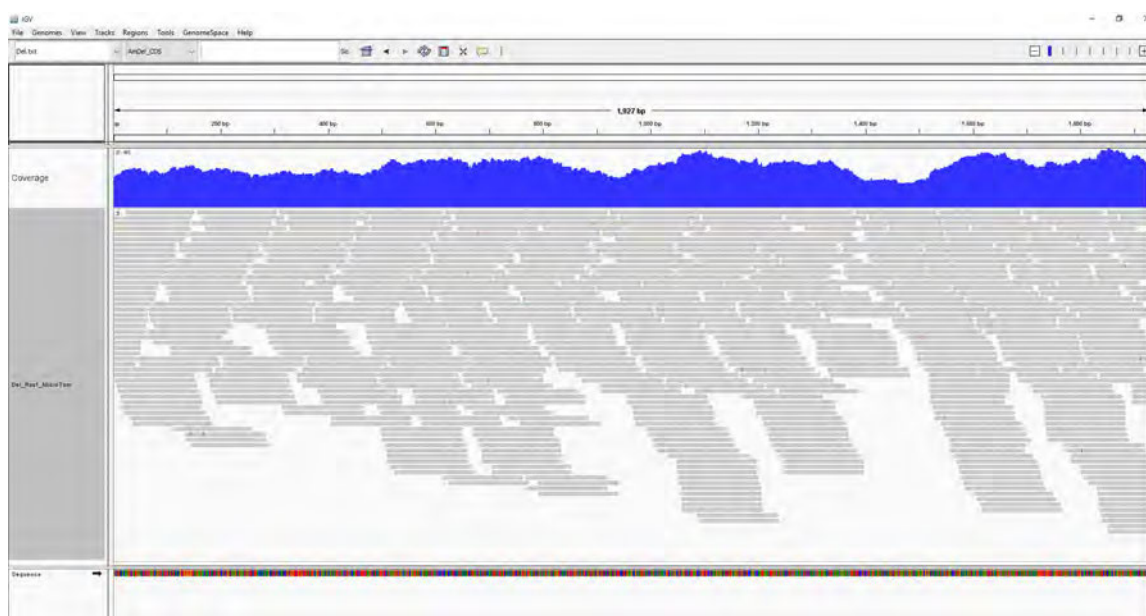


Figure 2: Alignment of reads from Illumina hi-seq over the sequence of the *Delila* (*Del*) gene in the sequenced *Del/Ros1-N* plant. Aligned reads shown as grey bars. Dark lines within grey bars show sequence mis-matches to the sequence of pDEL.ROS in individual reads. The amino acid sequence of the encoded transcription factor is shown in the bottom panel. The dark blue

profile shows the degree of read coverage for each part of the gene sequence. This alignment has been deposited in ENA under accession number ERR3502327.

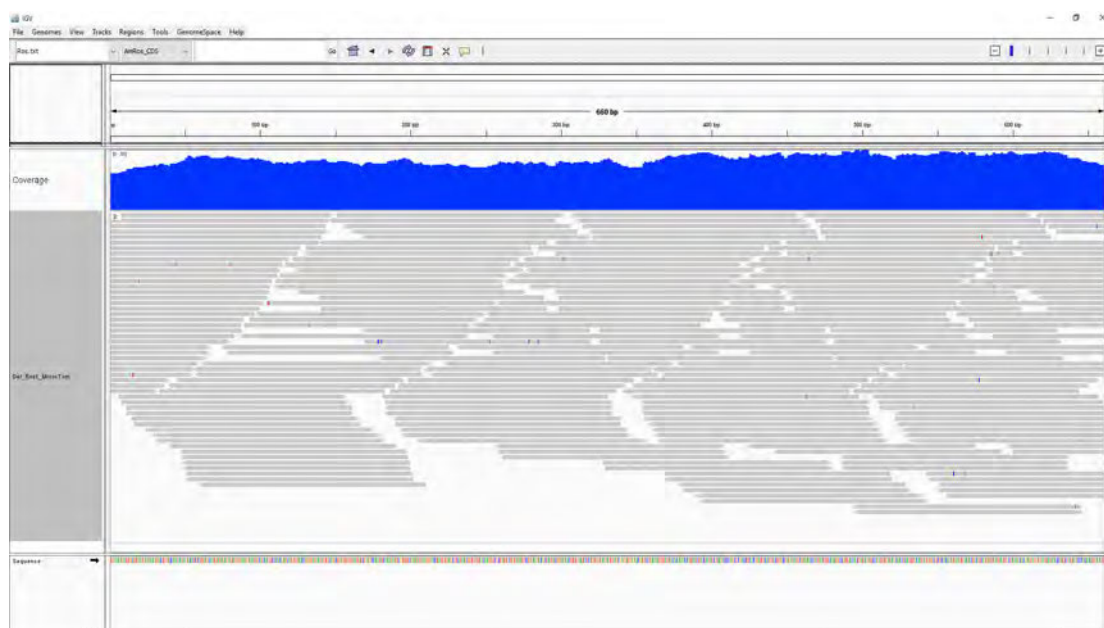


Figure 3: Alignment of reads from Illumina hi-seq over the sequence of the *Rosea1* (*Ros1*) gene in the sequenced *Del/Ros1-N* plant. Aligned reads shown as grey bars. Dark lines within grey bars show sequence mis-matches to the sequence of pDEL.ROS in individual reads. The amino acid sequence of the encoded transcription factor is shown in the bottom panel. The dark blue profile shows the degree of read coverage for each part of the gene sequence. This alignment has been deposited in ENA under accession number ERR3502327

Further bioinformatic analysis was conducted to establish whether any genes had been interrupted during the T-DNA insertion to create unexpected Open Reading Frames or whether any tomato genes had been disrupted following the T-DNA insertion, which might have unpredicted effects.

For the T-DNA insert, sequences homologous to the sequences flanking the T-DNA insertion (FLB and FRB) were searched by BLAST analysis amongst to tomato Open Reading Frames (ORFs) against “Tomato gene models CDS and cDNA (ITAG release 3.20)” <https://solgenomics.net/tools/blast/> and no new functional AA sequences could be detected spanning the T-DNA:FRB junction sequence on chromosome 4.

The T-DNA insertion occurred 797 bp upstream of the start of transcription of Solyc04g077950.2.1, a prediction based upon homology to a fragment of a rice protein Os01g0765200. There is limited annotation for this sequence which could be a pseudogene. The insertion occurred 971 bp upstream of the start of transcription of Solyc04g077940.3.1 but 1,134 bp upstream of the start of translation of this gene and so the T-DNA insertion was not expected to influence its expression. Solyc04g077940.3.1 encodes a flowering time control protein FPA (AHRD V1 *-*- FPA_ARATH) containing the Interpro domain(s) IPR012921. No alteration of flowering or flowering time was observed in purple tomato plants in any genetic background across multiple generations compared to wild type controls indicating that the T-DNA insertion did not affect activity of the protein encoded by Solyc04g077940.3.1. The T-DNA

insertion occurred 2439 bp downstream of Solyc04g077960.1.1, a position which is not expected to influence the expression of this predicted gene.

Conclusion: No ORF was interrupted by the presence of the T-DNA insert on chromosome 4. Putative new ORFs were searched by translating the genomic: T-DNA junction sequences in all possible reading frames (6) and detecting any initiator of translation codons (ATG encoding Methionine). Bioinformatic analysis showed that the T-DNA insertion in the *Del/Ros1-N* transformation event did not interrupt any existing open reading frame (ORFs). Frame 2 flanking the right border reading 3' to 5' created an ORF containing an initiation codon (ATG) that could produce a 32 amino acid peptide. However, this peptide has no homology to any known allergen or protein and there was no evidence (from tomato RNA-seq databases) that this sequence is transcribed in tomato. It was concluded that the T-DNA insertion on chromosome 4 did not create a new functional open reading frame (ORF).

Appendix 3: Analysis of next-generation whole genome sequencing data from a commercial line of The Purple Tomato™

Overview

This report details the bioinformatic analysis and workflow used to 1) identify the genomic integration site(s) of transgenic locus *Del/Ros1* in the tomato line used for all commercial breeding done by Norfolk Healthy Produce, and 2) to identify the presence (*i.e.* verify the absence) of vector backbone sequence in the genome. The analysis for transgene insertion site(s) leverages Illumina next-generation sequencing data to identify junction reads that span the border between the transgene and host genome. The analysis for presence of vector backbone uses an alignment algorithm with sensitive detection to identify the presence of any sequence in the genome ≥ 20 bp derived from the vector.

I. Detection of transgene insertion site(s)

Materials

Reference Sequences

- **Transgene reference:** pdelros.fasta (GenBank Accession: MN580094.1)
- **Genomic reference:** *Solanum lycopersicum* SL4.0 assembly (downloaded from Sol Genomics Network)

Sequencing Data

- **Sample:** Genomic DNA extracted from frozen leaf tissue of a commercial breeding line, derived from MoneyMaker F9 progeny of event *Del/Ros1-N* (OECD unique identifier NPS-01201-8). The commercial breeding line is the result of a cross of MoneyMaker F9 x Goldkrone, with subsequent selection of selfed progeny for purple fruit color and high fruit count per truss. The material used for sequencing is >7 generations of selfing from the MoneyMaker x Goldkrone F1.
- **Platform:** Illumina NovaSeq X Plus
- **Read format:** 150bp paired-end
- **Data files:** NPS_01201_8_1.fq.gz and NPS_01201_8_2.fq.gz (paired-end reads)

Software

- BWA-MEM2 <https://github.com/bwa-mem2/bwa-mem2> (version 2.2.1) (Vasimuddin *et al.* 2019)
- Samtools (version 1.21) <https://htslib.org> (Danecek *et al.* 2021)
- Seqtk (version 1.4) <https://github.com/lh3/seqtk>
- Integrative Genomics Viewer (IGV; version 2.19.2) <https://igv.org/> (Thorvaldsdóttir *et al.* 2013).
- NCBI BLASTn https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

Methods

Genomic DNA sequencing

The quality control report from Novogene is appended to this report, and describes the quality control of the primary sequencing data. The read depth was 40X, based on 210,542,096 paired-end reads of 150bp, for a total of 31,581,314,400bp. The reference genome for tomato

comprises 782.5Mbp ([https://phytozome-next.jgi.doe.gov/info/Slycopersicum ITAG4_0](https://phytozome-next.jgi.doe.gov/info/Slycopersicum_ITAG4_0)), from which we can calculate 40.4X coverage of the genome.

Alignment to Transgene Reference

Paired-end reads were aligned to the transgene reference sequence using BWA-MEM2 with a minimum seed length of 35bp to ensure specific matching. This ensures that reads used to identify junction points to the genome are exact matches to the transgenic insert. [N.B. The reference file used for alignment was the entire vector pDEL.ROS sequence (Accession: MN580094.1), so as a bonus, if any read exactly matched any of the vector backbone for ≥ 35 bp, it also would appear in the analysis. However, we used a more sensitive method for the specific purpose of determining whether any portion of the vector backbone is present in the genome; that method is detailed in the next section “Vector Backbone Analysis”.]

Commands and explanation:

```
# Index transgene reference
bwa-mem2 index pdelros.fasta

# Align reads to transgene
bwa-mem2 mem -k 35 pdelros.fasta NPS_01201_8_1.fq.gz
NPS_01201_8_2.fq.gz | samtools view -bS - | samtools sort -o
althea_reads_aligned_to_pdelros.bam

# Index BAM file
samtools index althea_reads_aligned_to_pdelros.bam
```

Explanation of parameters:

- `bwa-mem2 index pdelros.fasta`: Creates an index of the transgene reference sequence to enable fast searching
- `bwa-mem2 mem`: The alignment algorithm that finds where each sequencing read matches the reference
- `-k 35`: Sets the minimum seed length to 35 bases - this means a read must have at least 35 consecutive matching bases to be considered a valid alignment
- `pdelros.fasta`: The transgene reference sequence file
- `NPS_01201_8_1.fq.gz` and `NPS_01201_8_2.fq.gz`: The paired-end sequencing read files
- `samtools view -bS -`: Converts the alignment results from text format (SAM) to binary format (BAM)
- `samtools sort`: Arranges the alignments in order by their position on the reference sequence
- `-o althea_reads_aligned_to_pdelros.bam`: Names the output file containing the sorted alignments
- `samtools index`: Creates an index of the BAM file to enable fast random access

Extraction of Potential Junction Reads

In order to identify reads that exactly matched the transgene *and* were junctions to genomic DNA, we extracted reads that exactly matched the transgene and that contained ≥ 20 bp that did not match the transgenic sequence. (The portion of the read that doesn't match the transgene

reference is referred to as a “soft-clipped” portion.) These are potential junctions of transgene insert DNA and tomato genomic DNA.

Commands and explanation:

```
# Extract reads with at least 20 bases soft-clipped
samtools view althea_reads_aligned_to_pdelros.bam | awk '$6 ~ /[2-9][0-9]S|[0-9]{3,}S/' | cut -f1 > substantial_junctions.txt

# Extract sequences from original FASTQ files
seqtk subseq NPS_01201_8_1.fq.gz substantial_junctions.txt | seqtk seq -A > substantial_junction_R1.fasta
seqtk subseq NPS_01201_8_2.fq.gz substantial_junctions.txt | seqtk seq -A > substantial_junction_R2.fasta
```

Explanation of parameters:

- `samtools view althea_reads_aligned_to_pdelros.bam`: Displays the contents of the alignment file
- `awk '$6 ~ /[2-9][0-9]S|[0-9]{3,}S/':` Filters for reads that have at least 20 bases "soft-clipped" (meaning these bases didn't match the transgene). The pattern `[2-9][0-9]S|[0-9]{3,}S` looks for reads where the CIGAR string (a code describing the alignment) contains 20 or more bases marked with "S" (soft-clipped)
- `cut -f1`: Extracts just the read names from the filtered alignment data
- `> substantial_junctions.txt`: Saves the list of read names to a file
- `seqtk subseq`: A tool that extracts specific sequences from a FASTQ file based on a list of sequence names
- `seqtk seq -A`: Converts the FASTQ format to FASTA format (removing quality scores)
- `> substantial_junction_R1.fasta` and `> substantial_junction_R2.fasta`: Names the output files for the forward and reverse reads

Alignment to Host Genome

Junction read candidates were aligned to the *Solanum lycopersicum* SL4.0 reference genome. This step aligns the reads that contain both a portion of the transgene sequence and a stretch of sequence that does not match the transgene onto the tomato genome.

Commands and explanation:

```
# Index reference genome
bwa-mem2 index S_lycopersicum_chromosomes.4.00.fa

# Align junction reads to genome
bwa-mem2 mem S_lycopersicum_chromosomes.4.00.fa
substantial_junction_R1.fasta substantial_junction_R2.fasta | samtools
view -bS - | samtools sort -o junction_reads_on_genome.bam

# Index BAM file
samtools index junction_reads_on_genome.bam
```

Explanation of parameters:

- `bwa-mem2 index S_lycopersicum_chromosomes.4.00.fa`: Creates an index of the tomato genome reference sequence
- `bwa-mem2 mem`: The alignment algorithm
- `S_lycopersicum_chromosomes.4.00.fa`: The tomato reference genome file
- `substantial_junction_R1.fasta` and `substantial_junction_R2.fasta`: The input files containing the potential junction reads
- `samtools view -bS -`: Converts the alignment results from text format (SAM) to binary format (BAM)
- `samtools sort`: Arranges the alignments in order by their position on the reference sequence
- `-o junction_reads_on_genome.bam`: Names the output file containing the sorted alignments
- `samtools index`: Creates an index of the BAM file to enable fast random access

Visualization and Manual Curation

To see where the junction reads aligned on the genome, the Integrative Genomics Viewer (IGV) was used. The *S. lycopersicum* reference genome was indexed and loaded into IGV. The Indexed file of `junction_reads_on_genome.bam` was then loaded into IGV and the distribution of reads across the genome inspected.

Explanation of this step:

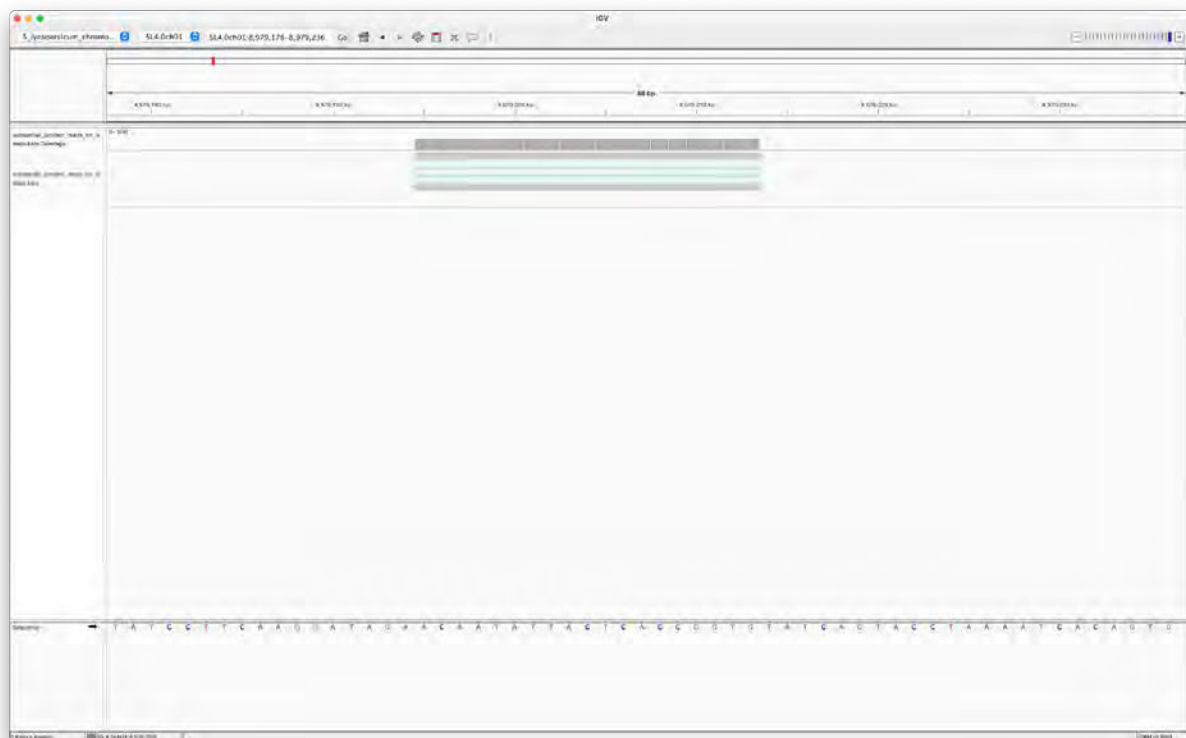
- IGV (Integrative Genomics Viewer) is a visualization tool that displays alignments graphically
- In junction reads, one part matches the transgene while another part matches the host genome
- BLASTn is a nucleotide sequence similarity search tool used to confirm whether parts of reads match the transgene or host genome
- Manual curation involves human inspection to distinguish true junction reads from false positives

Results

Identified Integration Site(s)

Junction reads were found aligned in 2 positions in the genome as described below.

Chromosome 1:



Inspection of the alignment indicated that these junction reads were aligning to a 19bp sequence in the tomato genome. This 19mer was the compared to the sequence of pDEL.ROS using BLASTn and the results appear below:

```
Job Title:19mer
Program: BLASTN
Query: 19mer ID: lcl|Query_6460817(nucleic acid) Length: 19
Subject:Plant binary expression vector pDEL.ROS, complete sequence ID: MN580094.1(nucleic acid) Length: 33236

Sequences producing significant alignments:
```

Description	Scientific Name	Common Name	Taxid	Max Score	Total Score	Query cover	E Value	Per. Ident	Acc. Len	Accession
Plant binary expression vector pDEL.ROS, complete sequence	Plant binary...	NA	2683228	38.2	828	100%	1e-06	100.00	33236	MN580094.1

Alignments:

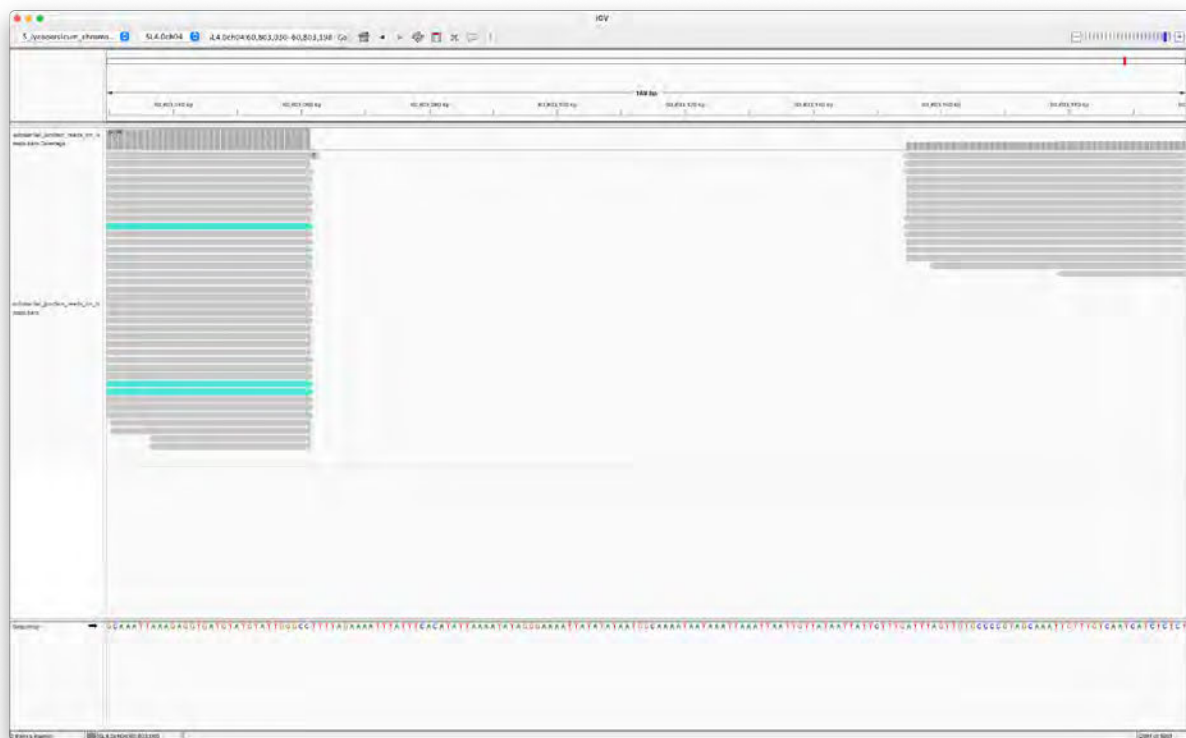
```
>Plant binary expression vector pDEL.ROS, complete sequence
Sequence ID: MN580094.1 Length: 33236
Range 1: 12577 to 12595

Score:38.2 bits(19), Expect:1e-06,
Identities:19/19(100%), Gaps:0/19(0%), Strand: Plus/Minus

Query 1      ACAATATTACTACCGGTG 19
            |||
Subject 12595 ACAATATTACTACCGGTG 12577
```

The alignment shows that this 19bp sequence from the tomato reference genome sequence exactly matches the reverse complement of nucleotides 12577-12595 of the pDEL.ROS sequence. This is a rare chance occurrence, but also not unexpected. It has no relevance to the presence of the transgene; it is simply a coincidence that the reference tomato genome contains a 19mer that matches a portion of pDEL.ROS.

Chromosome 4:



This site corresponds exactly to the previously reported locusB in the MicroTom Purple Tomato genome. The 94 bp gap in the tomato genome at the integration is identical to what was previously reported for locusB in MicroTom. [N.B. The numbering of the nucleotides is different from the alignment reported for the MicroTom sequence, because the *S. lycopersicum* reference genome sequence has since been updated.]

Conclusion

The Purple Tomato line now commercialized and used for breeding of new varieties is now confirmed to contain only a single transgenic locus, identical to the locus B already analyzed in the sibling MicroTom line.

II. Detection of vector backbone sequence in the genome

This report details the bioinformatic analysis performed to determine whether any vector backbone sequences were integrated into the plant genome along with the transgene of interest. The analysis used next-generation Illumina sequencing data from the commercial Purple Tomato transgenic line and aligned it to the complete vector sequence to identify any reads mapping to the backbone region.

Materials

Reference Sequences

- **Transgene and vector reference:** pdelros.fasta (Accession: MN580094.1)

Sequencing Data

- As above in Part I.

Software

- Bowtie2 (version 2.5.4) (Langmead and Salzberg, 2012)
- Samtools (version 1.21) (Danecek *et al.* 2021)

Methods

Alignment to Complete Vector

Paired-end reads were aligned to the complete pDEL.ROS vector sequence (including both transgene and backbone regions) using Bowtie2 with sensitive local alignment parameters designed to detect sequence identity as short as 20bp.

Commands and explanation:

```
#Build Bowtie2 index of the vector reference sequence
bowtie2-build pdelros.fasta pdelros_bt2
```

```
#Align reads to the complete vector:
```

```
bowtie2 --very-sensitive-local -x pdelros_bt2 -1 NPS_01201_8_1.fq.gz -
2 NPS_01201_8_2.fq.gz -S bowtie2_pdelros.sam
```

Explanation of parameters:

- **bowtie2-build**: Creates an index of the vector sequence to enable efficient alignment
- **--very-sensitive-local**: Sets parameters for highly sensitive local alignment to detect even partial matches. Specifically, this preset configures:
 - **-D 20**: Perform up to 20 consecutive seed extensions before giving up
 - **-R 3**: Maximum of 3 re-seed attempts when a good alignment isn't found
 - **-N 0**: No mismatches allowed in seed alignment (stringent matching)
 - **-L 20**: Seed length of 20 bp
 - **-i S,1,0.50**: Interval between seeds scales with read length
 - **--local**: Performs local alignment, allowing soft-clipping of unaligned portions
- **-x pdelros_bt2**: Specifies the index prefix
- **-1** and **-2**: Specify the forward and reverse read files
- **-S**: Specifies the output SAM file

Processing Alignments

The alignment data was processed to identify regions in the reference vector sequence onto which reads had been mapped.

Commands and explanation:

```
#Convert to BAM and sort
```

```
samtools view -bS bowtie2_pdelros.sam > bowtie2_pdelros.bam
```

```
samtools sort bowtie2_pdelros.bam -o bowtie2_pdelros_sorted.bam
```

```
#Count total mapped reads
```

```
samtools view -F 4 bowtie2_pdelros_sorted.bam | wc -l
```

Analyzing Read Distribution

To determine where reads were mapped across the vector sequence, the number of reads was counted in each 1000bp window, starting from nucleotide coordinate 1 (the start of the T-DNA left border).

Commands and explanation:

```
#Count reads mapping to each 1000bp window:  
samtools view -F 4 bowtie2_pdelros_sorted.bam | awk '{print  
int($4/1000)*1000}' | sort | uniq -c > coverage_by_region.txt
```

Results

Read Mapping Statistics

A total of 3,651 reads mapped to the pDEL.ROS sequence. The distribution of reads in the output `coverage_by_region.txt` file was as follows:

Nucleotide coordinates	Number of reads
0-999	153
1000-1999	185
2000-2999	153
3000-3999	429
4000-4999	382
5000-5999	338
6000-6999	234
7000-7999	184
8000-8999	494
9000-9999	404
10000-10999	321
11000-11999	210
12000-12999	164
Total	3651

Analysis of the distribution of reads aligning to the vector

The pDEL.ROS sequence comprises the transgenic insert region, between the left and right T-DNA borders, from nucleotide coordinates 1-12710, and the vector backbone, from 12710-

33236 (see GenBank accession MN580094.1 for full annotation of the sequence; <https://www.ncbi.nlm.nih.gov/nuccore/MN580094.1?report=genbank>).

All the reads that align to pDEL.ROS from the resequencing of The Purple Tomato align to the region containing the transgenic insertion; no reads align to the vector backbone. The result is consistent with the previously reported absence of vector backbone in the sibling MicroTom line.

Conclusion

The analysis provides strong evidence that only the transgene portion (positions 1-12710) of the vector was integrated into the plant genome. No reads were found mapping to the vector backbone region, indicating a clean integration event without vector backbone contamination. The method employed is far more sensitive (able to detect as few as 20bp) than traditional molecular techniques, such as Southern blotting.

This confirms that the transgenic line contains only the intended transgene sequences and is free of vector backbone sequences. As a result, no unintended effects of spurious insertion of DNA are possible.

Overall Conclusion

The Purple Tomato line now commercialized and used for breeding of new varieties is now confirmed to contain only a single transgenic locus, identical to the locus B already analyzed in the sibling MicroTom line. In addition, no portion of the vector outside the intended transgenic locus is present in the genome.

Given that the now-confirmed single transgenic locus neither interrupts any open reading frames in the tomato genome, nor creates any new open reading frames (with the exception of a single short ORF, already analyzed and found negative for similarity to any known proteins, or allergen epitopes), the only conceivable effect of the single transgenic locus is to increase the amounts of anthocyanin pigments in the ripe fruit. No other mechanism that could affect the phenotype of the The Purple Tomato plant can be postulated.

Given this clear conclusion, and the fact that the only change to the tomato fruit of any significance to human health and nutrition is an increase in the same anthocyanin compounds already found in other fruits and vegetables, in amounts similar to or higher than those present in The Purple Tomato, we conclude that The Purple Tomato is as safe for human consumption as any other tomato.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10: 1-4.

Langmead, B., and Salzberg, S. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9:357-359.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2012). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 14: 178-192.

Vasimuddin, M. Misra, S, Li, H, and Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems, 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), Rio de Janeiro, Brazil, 2019, pp. 314-324, doi: 10.1109/IPDPS.2019.00041.

Appendix 4: Further analysis of insert stability: T-DNA flanking sequences in Del/Ros1-N stock lines derived by self-fertilization over >6 generations

Eugenio Butelli & Cathie Martin

John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

Table of Contents

- 1 – ORIGINAL *DEL/ROS1-N* TRANSFORMATION EVENT
- 2 – ANALYSIS OF THE GENOMIC REGION FLANKING THE RB OF INSERT B
- 3 – ANALYSIS OF T-DNA BORDER REGIONS INSERTED IN THE PLANT GENOME (INSERT B)
- 4 – STABILITY OF INSERT B ACROSS GENERATIONS AND GENETIC BACKGROUNDS
- 5 – CONCLUSION

This analysis involved further characterization of plants containing the *Del/Ros1-N* transformation event across many generations following self-pollination or crossing. T-DNA flanking sequences in the tomato genome were retrieved by polymerase chain reaction (PCR) and sequenced to determine the location and characterization of both plant genomic and T-DNA sequences at the point of insertion. This work complemented previous molecular characterization studies (see Appendix 5).

1 – ORIGINAL *DEL/ROS1-N* TRANSFORMATION EVENT

Previously, the analysis of the genomic regions flanking the left border (LB) of the T-DNA confirmed the presence of two distinct T-DNA insertions in event *Del/Ros1-N*. Based on “Tomato WGS chromosomes (SL 2.50)” one T-DNA insertion (insert A) was identified on an unanchored genomic contig SL2.50ch00 at position 14105 and a second T-DNA insertion (insert B) on chromosome 4 (SL2.50ch04) at position 62818821. The T-DNA insertion at locus A was lost in the *Del/Ros1-N* event by segregation, as confirmed by PCR analysis in later generations further work focused only on the remaining locus B insert. Re-sequencing and further annotation of the tomato genome enabled us to update the location of the LB flanking sequence at locus B:

T-DNA insertion	Tomato WGS chromosomes (SL 2.50)	New genome sequence and/or annotations
Insert B	position 62818821 on chromosome 4	position 62904771 on chromosome 4 (Build SL 3.0)

1a – Insert B: updated BLAST analysis of the genomic region FLB-B flanking the left border of the T-DNA using Tomato genome chromosomes Build SL 3.0

FLB-B: 449 nt (see Appendix 5):

TATTTAGTTGTGCCCCGTAGCAAATTGTTTGTCAATCATCTCTCTCCCTCAAACCTCTCGCTCTATCT
CTCGTTCGCTTCCTCTCGCTTTTATACAAACACAAGTGTATAAAATTTGTTTTATTATATAAAGCGA

GAGAAAATTGTATAAATACATATATTTTTGTTTTCTCTGCCAGATCTCGCTGCCACTCTCCCTAAT
 CTCGTTCCGCCACCCTCGCCTTTCTCGCTTATACAAACAGAAGCGAAATGTATAAATTGTGTTTCTG
 TTTGTATAAAGTAAGAGAAAATTATATACACATGCAAATACATATATTTTCGTCCTATACACTTAT
 AGTTATACAATAAAAAATACTCCCCTGCCAGTTTCTTTTGCCTTTCTCTTTCTAATTTTATACAAAT
 TCAAATTGTATATAATTTCTCCCTTTCTCATTTTATACAATTAA

Homology of FLB-B to the tomato genome were searched by BLAST analysis against "Tomato genome chromosomes (Build SL 3.0)" <http://solgenomics.net/tools/blast/>

Untitled vs Tomato genome chromosomes (Build SL3.0)					
SubjectId	id%	Aln	evalue	Score	Description
SL3.0ch04	99.78	447/448	0.0	880	Length = 66557038
SL3.0ch04	88.86	391/440	e-129	464	Length = 66557038
SL3.0ch04	89.81	335/373	e-111	406	Length = 66557038
SL3.0ch04	89.14	353/396	e-111	404	Length = 66557038
SL3.0ch04	88.86	343/386	e-107	392	Length = 66557038
SL3.0ch04	87.79	345/393	e-103	379	Length = 66557038
SL3.0ch04	89.01	332/373	e-101	373	Length = 66557038
SL3.0ch04	88.13	349/396	e-101	373	Length = 66557038
SL3.0ch04	88.89	344/387	e-101	371	Length = 66557038
SL3.0ch04	88.47	330/373	1e-99	367	Length = 66557038

>SL3.0ch04

Length = 66557038

Score = 880 bits (444), Expect = 0.0

Identities = 447/448 (99%)

Strand = Plus / Plus

Query: 2 atttagttgtgccccgtagcaaatgtttgtcaatcatctctctccctcaaactctcgct 61
 |||
 Sbjct: 62904771 atttagttgtgccccgtagcaaatgtttgtcaatcatctctctccctcaaactctcgct 62904830

Query: 62 ctatctctcgcttcgcttccctctcgctttttatacaaacacaaagtgtataaaattgttttt 121
 |||
 Sbjct: 62904831 ctatctctcgcttcgcttccctctcactttttatacaaacacaaagtgtataaaattgttttt 62904890

Query: 122 attatataaagcgagagaaaaattgtataaatacatatatttttgttttcctctgccagat 181

```

|||||
Sbjct: 62904891 attatataaagcgagagaaaattgtataaatacatatatttttgttttcctctgccagat 62904950

Query: 182      ctgctgccactctccctaatactcgttcgccaccctcgccctttctcgcttatacaaacag 241
|||||
Sbjct: 62904951 ctgctgccactctccctaatactcgttcgccaccctcgccctttctcgcttatacaaacag 62905010

Query: 242      aagcgaaatgtataaattgtgtttctgtttgtataaagtaagagaaaattatatatacac 301
|||||
Sbjct: 62905011 aagcgaaatgtataaattgtgtttctgtttgtataaagtaagagaaaattatatatacac 62905070

Query: 302      atgcaaatacatatattttcgtcctatacacttatagttatacaataaaaaatactcccct 361
|||||
Sbjct: 62905071 atgcaaatacatatattttcgtcctatacacttatagttatacaataaaaaatactcccct 62905130

Query: 362      gccagtttcttttgccctttctctcttttctaattttatacaaaattcaaattgtatataatt 421
|||||
Sbjct: 62905131 gccagtttcttttgccctttctctcttttctaattttatacaaaattcaaattgtatataatt 62905190

Query: 422      tctccctttctcattttatacaatttaa 449
|||||
Sbjct: 62905191 tctccctttctcattttatacaatttaa 62905218

```

Search Parameters

Database: SL3.0ch00-12.fasta
 Posted date: Jan 27, 2017 2:55 PM
 Number of letters in database: 828,076,956
 Number of sequences in database: 13

Lambda	K	H
1.37	0.711	1.31

Gapped

Lambda	K	H
1.37	0.711	1.31

Search Statistics

Matrix: blastn matrix:1 -3
 Gap Penalties: Existence: 5, Extension: 2
 Number of Sequences: 13
 Number of Hits to DB: 8,560,534
 Number of extensions: 15733
 Number of successful extensions: 15733
 Number of sequences better than 1.0e-10: 13
 Number of HSP's gapped: 12348
 Number of HSP's successfully gapped: 4788
 Length of query: 449
 Length of database: 828,076,956
 Length adjustment: 20

Effective length of query: 429
 Effective length of database: 828,076,696
 Effective search space: 355244902584
 Effective search space used: 355244902584
 X1: 11 (21.8 bits)
 X2: 15 (29.7 bits)
 X3: 50 (99.1 bits)
 S1: 20 (40.1 bits)
 S2: 36 (71.9 bits)

1b -Conclusion

Analysis of the genomic regions flanking the left border (LB) of the T-DNA confirmed the presence a T-DNA insertion (insert B) on chromosome 4 (Build SL 3.0) at a re-numbered position 62904771 according to Tomato genome chromosomes Build SL 3.0.

Following out-segregation of Insert A from the *Del/Ros1-N* transformation event, tomato plants exhibiting a stable purple fruit phenotype and containing only the transgenic insert B were used for further studies including the transfer by crossing to a range of different tomato genetic backgrounds.

2 – ANALYSIS OF THE GENOMIC REGION FLANKING THE RB OF INSERT B

The sequences flanking the LB of insert B (FLB-B) had been characterized previously and the insertion at position 62904771 on chromosome 4 (Build SL 3.0) had been validated. We characterised the sequence flanking the RB of the T-DNA insert B (FRB-B) in stock plants in both MicroTom and MoneyMaker genetic backgrounds (generations T6 and F9 respectively).

2a - Plant material

Leaf material from progeny plants obtained by self-pollination of *Del/Ros1-N* plants containing only the Insert B from the *Del/Ros1-N* transformation event in both MicroTom and MoneyMaker genetic backgrounds were harvested and kept frozen prior to DNA extraction (See Figure 7)

Plant sample name	Plant pedigree
Stock <i>Del/Ros1-N</i> in MicroTom	<i>Del/Ros1-N</i> T6
Stock <i>Del/Ros1-N</i> in MoneyMaker	<i>Del/Ros1-N</i> in MoneyMaker F9

2b - PCR amplification results

The conditions for PCR amplification are detailed below. Based upon previous analysis of the LB analysis of insert B, PCR primers were design to amplify the genomic region flanking the RB of the T-DNA. A PCR product of approximately 720 bp was retrieved. The band was gel purified, cloned and subsequently sequenced (EF00613497_EF00613497).

PCR run conditions:

Activate	95oC	2 min
37 cycles	95oC	30 s
	65oC	40 s
	72oC	50 s
Final cycle	72oC	10 min

Primer name	Primer sequence
BF (genomic)	GGAAGAAGAGTAAATACGTAAAA

RBF2 (T-DNA RB) GGGAGAGGCGGTTTGCGTATTGG

2c - Sequencing of the genomic:RB T-DNA junction region

Sanger sequencing of the PCR products was undertaken using the BF and RBF2 oligonucleotides (see section 3b).

>EF00613497_EF00613497

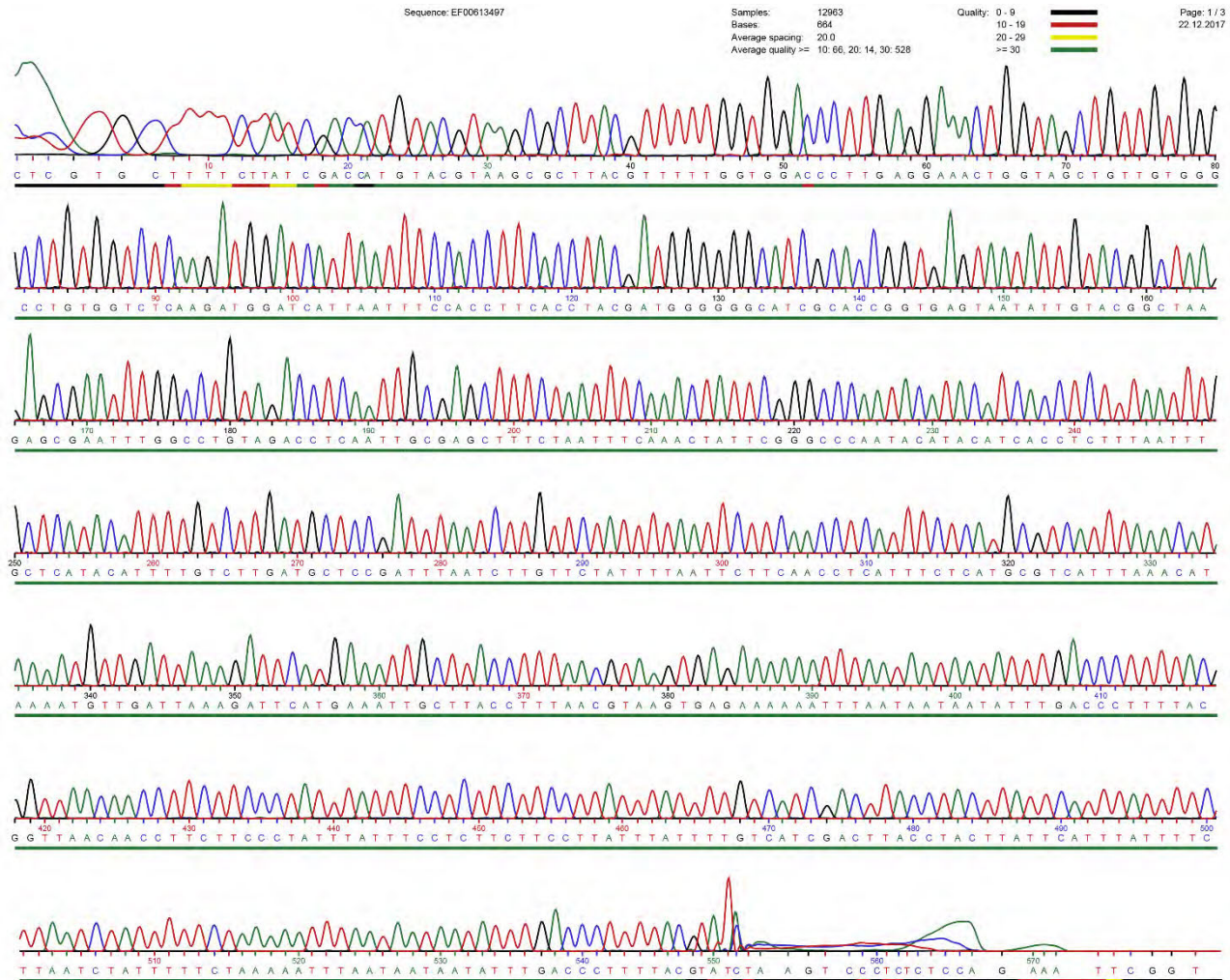
CTCGTGCTTTTCTTATCGACCATGTACGTAAGCGTTACGTTTTTGGTGGACCCTTGAGGAACTGGTAGCTGT
TGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGT
AATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGACCTCAATTGCGAGCTTCTAATTTCAAACCTATTCGGGC
CCAATACATACATCACCTCTTTAATTTGCTCATACATTTTGTCTTGATGCTCCGATTTAATCTTGTTCTATTTT
AATTCTTCAACCTCATTTCTCATGCGTCATTTAAACATAAAATGTTGATTAAAGATTCATGAAATTGCTTACCT
TTAACGTAAGTGAGAAAAAATTTAATAATAATTTGACCCTTTTACGGTTAACAACCTTCTTCCCTATTTATT
TCCTCTTCTCCTTATTTATTTTGTATCGACTTACCTACTTATTCATTTATTTTCTTAATCTATTTTCTAAA
AATTTAATAATAATTTTGACCCTTTTACGTATCTAAGTCCCTCTCTCCAGAAATTCGGT

NNNNNNNN = Low quality sequence

NNNNNNNN = inserted right border region

NNNNNNNN = tomato genome (FRB-B, 332 nt)

NNNNNNNN = Low quality sequence



FRB-B: 332 nt

GGCCCAATACATACATCACCTCTTTAATTTGCTCATACATTTTGTCTTGATGCTCCGATTTAA
TCTTGTTCTATTTTAATTCTTCAACCTCATTCTCATGCGTCATTTAAACATAAAATGTTGATT
AAAGATTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAAATTTAATAATAATTTTGAC
CCTTTTACGGTTAACAACCTTCTTCCCTATTTATTTCTTCTTCTTCTTATTTATTTTGTCATC
GACTTACCTACTTATTCATTTATTTCTTAATCTATTTTCTAAAAATTTAATAATAATTTTGA
CCCTTTTACGTAT

Homology of FRB-B to the tomato genome were searched by BLAST analysis against “Tomato genome chromosomes (Build SL 3.0)” <http://solgenomics.net/tools/blast/>

Untitled_sequence vs Tomato genome chromosomes (Build SL3.0)

SubjectId	id%	Aln	evalue	Score	Description
SL3.0ch04	99.70	331/332	0.0	650	Length = 66557038

[SL3.0ch04](#) 96.00 72/75 1e-26 125 Length = 66557038

>SL3.0ch04

Length = 66557038

Score = 650 bits (328), Expect = 0.0

Identities = 331/332 (99%)

Strand = Plus / Minus

Query: 220 ggccaatacatatcacctctttaatttgctcatatcttgtcttgatgctccgatt 279
 |||
 Sbjct: 62904677 ggccaatacatatcacctctttaatttgctcatatcttgtcttgatgctccgatt 62904618

Query: 280 taatcttggttctattttaattcttcaacctcatttctcatgcggtcatttaaacataaaat 339
 |||
 Sbjct: 62904617 taatcttggttctattttaattcttcaacctcatttctcatgcggtcatttaaacataaaat 62904558

Query: 340 gttgattaaagattcatgaaattgcttacctttaacgtaagtgagaaaaaatttaataat 399
 |||
 Sbjct: 62904557 gttgattaaagattcatgaaattgcttacctttaacgtaagtgagaaaaaatttaataat 62904498

Query: 400 aatatttgacccttttacggttaacaaccttcttcctatttatttcctcttcttcctta 459
 |||
 Sbjct: 62904497 aatatttgacccttttacggttaacaaccttcttcctatttatttcctcttcttcctta 62904438

Query: 460 tttattttgtcatcgacttacctacttattcatttattttcttaattctatttttctaaaa 519
 |||
 Sbjct: 62904437 tttattttgtcatcgacttacctacttattcatttattttcttaattctatttttctaaaa 62904378

Query: 520 atttaataataatatttgacccttttacgtat 551
 |||
 Sbjct: 62904377 atttaataataatatttgacccttttacgtat 62904346

Search Parameters

Parameter	Value
allowgaps	yes
expect	1.0e-10
gapext	2
gapopen	5
matrix	blastn matrix:1 -3

Search Statistics

Statistic	Value
Hits to DB	9,927,005
S1	21
S1_bits	42.1
S2	37
S2_bits	73.8
X1	11
X1_bits	21.8
X2	15

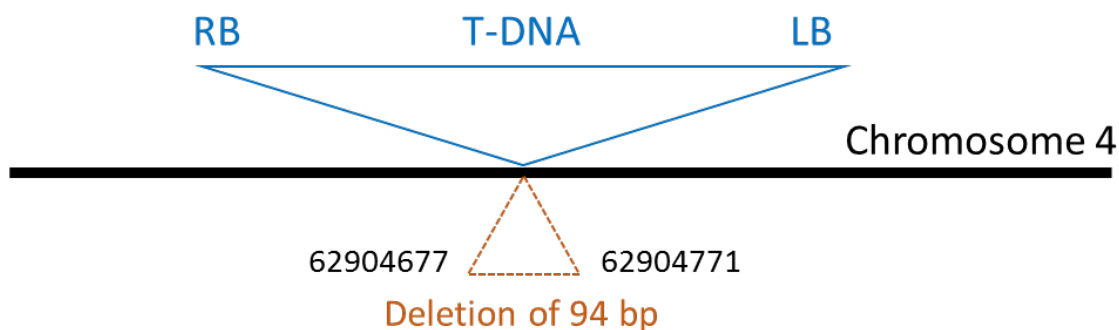
X2_bits	29.7
X3	50
X3_bits	99.1
dbentries	13
dbletters	828076956
effectivedblength	828076696
effectivespace	533281392224
effectivespaceused	533281392224
entropy	1.31
entropy_gapped	1.31
kappa	0.711
kappa_gapped	0.711
lambda	1.37
lambda_gapped	1.37
length_adjustment	20
num_extensions	10
num_successful_extensions	10
number_of_hsps_gapped	8
number_of_hsps_successfully_gapped	2
posted_date	Jan 27, 2017 2:55 PM
querylength	644
seqs_better_than_cutoff	1

Produced by Bioperl module Bio::SearchIO::Writer::HTMLResultWriter on Thu Apr 5 16:42:38 2018

2d - Conclusion

The analysis of the genomic regions flanking the right border of the T-DNA insert B (FRB-B) showed an insertion on chromosome 4 (Build SL 3.0) at position 62904677 suggesting a deletion of the tomato genome of 94 bp at the point of insertion (LB insertion at position 62904771).

Schematic representation of the insert B in the *Del/Ros1-N* event:



3 - ANALYSIS OF T-DNA BORDER REGIONS INSERTED IN THE PLANT GENOME (INSERT B)

The T-DNA border regions of insert B present in the tomato genome were compared to the original plasmid used for transformation using NCBI BLASTN 2.8.0+ (**Zhang et al., 2000**)

3a – T-DNA RB region analysis (insert B)

Alignment of genomic:RB T-DNA junction region with original T-DNA sequence used for transformation

>EF00613497_EF00613497 (see section 3b)

```
CTCGTGCTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCTTGAGGA
AACTGGTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGATCATTAAATTTCCACCTTCACCTA
CGATGGGGGGGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTA
GACCTCAATTGCGAGCTTTCTAATTTCAAACATTCGGGCCCAATACATACATCACCTCTTT
AATTTGCTCATACATTTTGTCTTGATGCTCCGATTTAATCTTGTTCTATTTTAATTCTTCAACC
TCATTTCTCATGCGTCATTTAAACATAAAATGTTGATTAAAGATTTCATGAAATTGCTTACCTT
TAACGTAAGTGAGAAAAAATTTAATAATAATATTTGACCCTTTTACGGTTAACAACCTTCTTC
CCTATTTATTTCTCTTCTTCTTCTTATTTATTTGTCATCGACTTACCTACTTATTCATTTATTT
CTTAATCTATTTTCTAAAAATTTAATAATAATATTTGACCCTTTTACGTATCTAAGTCCTCT
CTCCAGAAATTCGGT
```

NNNNN = Low quality sequence

NNNNN = right border T-DNA region inserted in the genome

NNNNN = tomato genome (FRB-B, 332 nt)

NNNNN = Low quality sequence

> Right Border region of T-DNA in transformation vector

```
GCTTTTGTTCCTTTAGTGAGGGTTAATTTTCGAGCTTGCGGTAATCATGGTCATAGCTGTTT
CCTGTGTGAAATTGTTATCCGCTCACAAATTCACACAACATACGAGCCGGAAGCATAAAGT
GTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCC
CGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGG
GAGAGGCGGTTTGCCTATTGGAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAA
GCGCTTACGTTTTTGGTGGACCCTTGAGGAACTGGTAGCTGTTGTGGGCCTGTGGTCTC
AAGATGGATCATTAAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATA
TTGTACGGCTAAGAGCGAATTTGGCCTGTAGACCTCAATTGCGAGCTTTCTAATTTCAAAC
ATTCGGGCCTAACTTTTGGTGTGATGATGCTGACTGGCAGGATATATACCGTTGTAA
```

NNNNN = T-DNA RB

NNNNN = deleted T-DNA sequence (52nt) following insertion into the plant genome

NCBI - Blast 2 sequence alignment:

366 bits(198) 4e-106 198/198(100%) 0/198(0%) Plus/Plus

Query	297	ATGTACGTAAGCGCTTACGTTTTTGGTGGACCCTTGAGGAACTGGTAGCTGTTGTGGGC	356
Sbjct	1	ATGTACGTAAGCGCTTACGTTTTTGGTGGACCCTTGAGGAACTGGTAGCTGTTGTGGGC	60
Query	357	CTGTGGTCTCAAGATGGATCATTAAATTTCCACCTTCACCTACGATGGGGGGCATCGCACC	416
Sbjct	61	CTGTGGTCTCAAGATGGATCATTAAATTTCCACCTTCACCTACGATGGGGGGCATCGCACC	120
Query	417	GGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGACCTCAATTGCGAGCTTT	476
Sbjct	121	GGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGACCTCAATTGCGAGCTTT	180
Query	477	CTAATTTCAAACATATTCG	494
Sbjct	181	CTAATTTCAAACATATTCG	198

Parameters NCBI - Blast 2 sequences

RID DYSKFYX5114 (Expires on 04-26 00:59 am)

Query ID lcl|Query_138247
Description Right Border region of T-DNA
Molecule type nucleic acid
Query Length 546

Subject ID lcl|Query_138249
Description [EF00613497](#) [_EF00613497](#)
Molecule type nucleic acid
Subject Length 198

Program BLASTN 2.8.0+

3b – T-DNA LB region analysis (insert B)

The conditions for PCR amplification are detailed in section 3b. PCR primers were designed to amplify the genomic region flanking the T-DNA LB region. A PCR product of approximately 1700 bp was retrieved. The band was gel purified, cloned and subsequently sequenced using the same primers (rev-EF00613496).

Primer name	Primer sequence
BR (genomic)	CATTGCAAGGTTATTCGAAAGTAACAAAC
LBR1(T-DNA LB)	GCTTTCGCCTATAAATACGACGGATCG

Alignment of genomic: LB T-DNA junction region with original T-DNA sequence used for transformation

```
>rev-EF00613496
ACAAAAATATATGTATTTATACAATTTTCTCTCGCTTTATATAATAAAAACAAATTTTATACACTTGTGTTTGT
ATAAAAGTGAGAGGAAGCGAACGAGAGATAGAGCGAGAGTTTGAGGGAGAGAGATGATTGACAAACAATTTGCT
ACGGGGCACAACATAATATGGAGAAAAAGAAAGAGTAATTACCAATTTTTTTTCAATTCAAAAATGTAGATGTC
CGCAGCGTTATTATAAAATGAAAGTACATTTTGATAAAACGACAAATTACGATCCGTCGTATTTATAGGCGAAA
GCAATAAACAAATTATTCTAATTCGGAAATCTTTATTTTCGACGTGTCTACATTCACGTCCAAATGGGGGCTTAG
ATGAGAACTTCACGATCGATGCCTTGATTTTCGCCATTTCCAGATACCCATTTTCATCTTCAGATTGGTCTGAGA
TTATGCGAAAATATACACTCATATACATAAATACTGACAGTTTGAGCTACCAATTCAGTGTAGCCCATTTACCTT
ACATAATTCACCTCAAATGCTAGGCAGTCTGTCAACTCGGCGTCAATTTGTGCGCCACTATACGATAGTTGCGCA
AATTTTCAAAGTCCTGGCCTAACATCACACCTCTGTGCGCGGGCGGGTCCCATTTGTGATAAATCCACCCATCGG
ATGATCATGAGCGGAGAATTAAGGGAGTCACGTTATGACCCCCGCCGATGACGCGGGACAAGCCGTTTACGTG
ACTCCCT
```

NNNNN = Low quality sequence

NNNNN = left border T-DNA region inserted in the genome

NNNNN = tomato genome (FLB-2)

NNNNN = Low quality sequence

>Left Border region sequence of T-DNA in transformation vector

```
GCAGGATATATTCAATTGTAAATGGCTTCATGTCCGGGAAATCTACATGGATCAGCAATGA
GTATGATGGTCAATATGGAGAAAAAGAAAGAGTAATTACCAATTTTTTTTCAATTCAAAAATG
```

TAGATGTCCGCAGCGTTATTATAAAATGAAAGTACATTTTGATAAAACGACAAATTACGATC
 CGTCGTATTTATAGGCGAAAGCAATAAACAAATTATTCTAATTCGGAAATCTTTATTTTCGACG
 TGTCTACATTCACGTCCAAATGGGGGCTTAGATGAGAACTTCACGATCGATGCCTTGATT
 TCGCCATTCCCAGATACCCATTTTCATCTTCAGATTGGTCTGAGATTATGCGAAAATATACAC
 TCATATACATAAATACTGACAGTTTGAGCTACCAATTCAGTGTAGCCCATTACCTTACATAAT
 TCACTCAAATGCTAGGCAGTCTGTCAACTCGGCGTCAATTTGTTCGGCCACTATACGATAGT
 TGCGCAAATTTTCAAAGTCCTGGCCTAACATCACACCTCTGTTCGGCGGGCGGGTCCCATTG
 TGATAAATCCACCCATCGGATGATCATGAGCGGAGAATTAAGGGAGTCACGTTATGACCCC
 CGCCGATGACGCGGGACAAGCCGTTTTACGTTTGGAAGTACAGAACCGCAACGTTGAAG
 GAGCCACTGAGCCGCGGGTTTCTGGAGTTTAATGAGCTAAGCACATACGTCAGAAACCATT
 ATTGC

NNNNN = T-DNA LB

NNNNN = deleted T-DNA sequence (75nt) following insertion into the plant genome

NCBI - Blast 2 sequence alignment:

Query	76	ATGGAGAAAAAGAAAGAGTAATTACCAAttttttttCAATTCAAAAATGTAGATGTCCGC	135
Sbjct	1	ATGGAGAAAAAGAAAGAGTAATTACCAATTTTTTTTCAATTCAAAAATGTAGATGTCCGC	60
Query	136	AGCGTTATTATAAAATGAAAGTACATTTTGATAAAACGACAAATTACGATCCGTCGTATT	195
Sbjct	61	AGCGTTATTATAAAATGAAAGTACATTTTGATAAAACGACAAATTACGATCCGTCGTATT	120
Query	196	TATAGGCGAAAGCAATAAACAAATTATTCTAATTCGGAAATCTTTATTTTCGACGTGTCTA	255
Sbjct	121	TATAGGCGAAAGCAATAAACAAATTATTCTAATTCGGAAATCTTTATTTTCGACGTGTCTA	180
Query	256	CATTCACGTCCAAATGGGGGCTTAGATGAGAACTTCACGATCGATGCCTTGATTTCGCC	315
Sbjct	181	CATTCACGTCCAAATGGGGGCTTAGATGAGAACTTCACGATCGATGCCTTGATTTCGCC	240
Query	316	ATTCACGATACCCATTTTCATCTTCAGATTGGTCTGAGATTATGCGAAAATATACACTCA	375
Sbjct	241	ATTCACGATACCCATTTTCATCTTCAGATTGGTCTGAGATTATGCGAAAATATACACTCA	300
Query	376	TATACATAAATACTGACAGTTTGAGCTACCAATTCAGTGTAGCCCATTACCTTACATAAT	435
Sbjct	301	TATACATAAATACTGACAGTTTGAGCTACCAATTCAGTGTAGCCCATTACCTTACATAAT	360
Query	436	TCACTCAAATGCTAGGCAGTCTGTCAACTCGGCGTCAATTTGTTCGGCCACTATACGATAG	495
Sbjct	361	TCACTCAAATGCTAGGCAGTCTGTCAACTCGGCGTCAATTTGTTCGGCCACTATACGATAG	420
Query	496	TTGCGCAAATTTTCAAAGTCCTGGCCTAACATCACACCTCTGTTCGGCGGGCGGGTCCCATT	555
Sbjct	421	TTGCGCAAATTTTCAAAGTCCTGGCCTAACATCACACCTCTGTTCGGCGGGCGGGTCCCATT	480
Query	556	TGTGATAAATCCACCCATCGGATGATCATGAGCGGAGAATTAAGGGAGTCACGTTATGAC	615
Sbjct	481	TGTGATAAATCCACCCATCGGATGATCATGAGCGGAGAATTAAGGGAGTCACGTTATGAC	540
Query	616	CCCCGCCGATGACGCGGGACAAGCCGTTTTACGT	649
Sbjct	541	CCCCGCCGATGACGCGGGACAAGCCGTTTTACGT	574

Parameters NCBI - Blast 2 sequences

RID DYVKNYPA11N (Expires on 04-26 01:50 am)

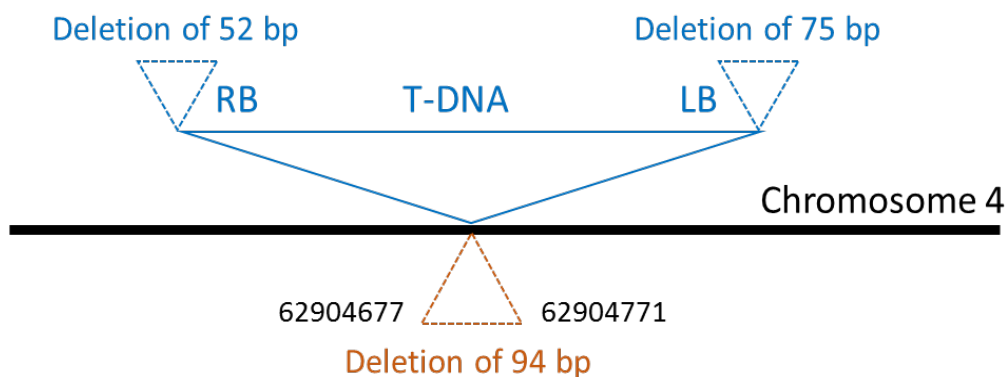
Query ID lcl|Query_29891
Description Left Border region of T-DNA
Molecule type nucleic acid
Query Length 744

Subject ID lcl|Query_29893
Description [rev-EF00613496](#)
Molecule type nucleic acid
Subject Length 582

ProgramBLASTN 2.8.0+

3c - Conclusion

The analysis of the T-DNA border regions inserted on chromosome4 (insert B) showed small deletions of 52 bp of the T-DNA RB region and 75 bp of the T-DNA LB region. Both border repeats at both ends of the T-DNA were also missing in insert B.



4 – STABILITY OF INSERT B ACROSS GENERATIONS AND GENETIC BACKGROUNDS

The Stability of insert B was assessed by retrieving and sequencing the T-DNA:genomic junction sequences at the LB and RB across multiple generations and in different genetic backgrounds. LB integrity was assessed in T1 and T6 plants from the “*Del/Ros1-N* in MoneyMaker” stock line. RB integrity was assessed in T6 plants from the “*Del/Ros1-N* in MoneyMaker” stock line and F9 plants from the “*Del/Ros1-N* in MicroTom” stock line. The breeding scheme for event *Del/Ros1-N* in MicroTom and MoneyMaker genetic backgrounds is detailed in Figure7.

The conditions for PCR amplification are detailed in section 3b and the PCR primers used are similar to those previously used to amplify the LB and RB genomic region flanking the T-DNA at locus B.

4a - Stability of insert B in T1 vs T6 plants (RB analysis)

Alignment of genomic: RB T-DNA junction region of T1 and T6 plants from the “Del/Ros1-N in MoneyMaker” stock line.

RB analysis of T1 plant from the “Del/Ros1-N in MoneyMaker” stock line.

>rev-EF00613488

```
TTGCGTATTGGAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCTTGAGGAAACT
GGTAGCTGTTGTGGGCTGTGGTCTCAAGATGGATCATTAAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAAT
ATTGTACGGCTAAGAGCGAATTTGGCCTGTAGACCTCAATTGCGAGCTTTCTAATTTCAAACCTATTCGGGGCCCAATACATACATCA
CCTCTTTAATTTGCTCATACATTTTGTCTTGATGCTCCGATTTAATCTTGTTCTATTTAATTTCTTCAACCTCATTCTCATGCGT
CATTTAAACATAAAATGTTGATTAAAGATTTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAAATTTAATAATAATATTGAC
CCTTTTACGGTTAACAACTTCTTCCCTATTTATTTCTCTTCTTC
```

RB analysis T6 plant from the “Del/Ros1-N in MoneyMaker” stock line.

>rev-EF00613489

```
TTGCGTATTGGAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCTTGAGGAAACT
GGTAGCTGTTGTGGGCTGTGGTCTCAAGATGGATCATTAAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAAT
ATTGTACGGCTAAGAGCGAATTTGGCCTGTAGACCTCAATTGCGAGCTTTCTAATTTCAAACCTATTCGGGGCCCAATACATACATCA
CCTCTTTAATTTGCTCATACATTTTGTCTTGATGCTCCGATTTAATCTTGTTCTATTTAATTTCTTCAACCTCATTCTCATGCGT
CATTTAAACATAAAATGTTGATTAAAGATTTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAAATTTAATAATT
```

NCBI - BLASTn 2 sequence alignment (Zhang et al., 2000) - 07-11-2019.

Query = T1

Subject = T6

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information

BLAST » blastn suite

Align Sequences Nucleotide BLAST

blastn blasto blastx tblastn tblastx

BLASTn programs search nucleotide subjects using a nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [From](#) [To](#)

TAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATTTGACGGCTAAG
ACCGAATTTGGCCTGTAGACCTCAATTGCGAGCTTTCTAATTTCAAACCTATTCGGGGCCCAATACATACATCA
CCTCTTTAATTTGCTCATACATTTTGTCTTGATGCTCCGATTTAATCTTGTTCTATTTAATTTCTTCAACCTCATTCTCATGCGT
CATTTAAACATAAAATGTTGATTAAAGATTTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAAATTTAATAATAATATTGAC
CCTTTTACGGTTAACAACTTCTTCCCTATTTATTTCTCTTCTTC

Or, upload file [Browse...](#) No file selected.

Job Title [Enter a descriptive title for your BLAST search](#)

☒ Align two or more sequences

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Subject subrange [From](#) [To](#)

TTTTGGGACCCCTTGAGGAACTGGTAGCTGTGGGCTGTGGTCTCAAGATGGATCAT
TAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATTTGACGGCTAAG
ACCGAATTTGGCCTGTAGACCTCAATTGCGAGCTTTCTAATTTCAAACCTATTCGGGGCCCAATACATACATCA
CCTCTTTAATTTGCTCATACATTTTGTCTTGATGCTCCGATTTAATCTTGTTCTATTTAATTTCTTCAACCTCATTCTCATGCGT
CATTTAAACATAAAATGTTGATTAAAGATTTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAAATTTAATAATT

Or, upload file [Browse...](#) No file selected.

Program Selection

Optimize for

☒ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☐ Somewhat similar sequences (blastn)

[Choose a BLAST algorithm](#)

BLAST Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

[Algorithm parameters](#)



U.S. National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite-2sequences » results for RID-W8NBMAKA11N

Edit Search
Save Search
Search Summary ▼

Search Parameters

Program	blastn
Word size	28
Expect value	10
Hitlist size	100
Match/Mismatch scores	1,-2
Gapcosts	0,2,5
Low Complexity Filter	Yes
Filter string	L;m;
Genetic Code	1

Karlin-Altschul statistics

Lambda	1.33271	1.28
K	0.620991	0.46
H	1.12409	0.85

Results Statistics

Effective search space	190650
------------------------	--------

Sequence ID: Query_18641Length: 421Number of Matches: 1
Range 1: 1 to 420Graphics Next Match Previous Match

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand
776 bits(420)	0.0	420/420(100%)	0/420(0%)	Plus/Plus

Query	1	TTGCGTATTGGAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGCTTAC	60
Sbjct	1	TTGCGTATTGGAACATCTACAAATTGCCTTTTCTTATCGACCATGTACGTAAGCGCTTAC	60
Query	61	GTTTTTGGTGGACCCTTGAGGAACTGGTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGA	120
Sbjct	61	GTTTTTGGTGGACCCTTGAGGAACTGGTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGA	120
Query	121	TCATTAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCAGGTGAGTAATATTGTACG	180
Sbjct	121	TCATTAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCAGGTGAGTAATATTGTACG	180
Query	181	GCTAAGAGCGAATTTGGCCTGTAGACCTCAATTGCGAGCTTTCTAATTTCAAACATTTCG	240
Sbjct	181	GCTAAGAGCGAATTTGGCCTGTAGACCTCAATTGCGAGCTTTCTAATTTCAAACATTTCG	240
Query	241	GGCCAATACATACATCACCTCTTTAATTTGCTCATAATTTTGCTTGATGCTCCGATT	300
Sbjct	241	GGCCAATACATACATCACCTCTTTAATTTGCTCATAATTTTGCTTGATGCTCCGATT	300
Query	301	TAATCTGTTCTATTTTAATTCTTCAACCTCATTTCTCATGCGTCATTTAAACATAAAAT	360
Sbjct	301	TAATCTGTTCTATTTTAATTCTTCAACCTCATTTCTCATGCGTCATTTAAACATAAAAT	360

```

Query   361   GTTGATTAAAGATTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAAATTTAATAAT   420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct   361   GTTGATTAAAGATTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAAATTTAATAAT   420

```

We conclude that the insert B has been stably maintained across 5 generations in the MicroTom genetic background.

4b - Stability of insert B in T1 vs T6 plants (LB analysis)

Alignment of genomic:LB T-DNA junction region of T1 and T6 plants from the “Del/Ros1-N in MoneyMaker” stock line.

LB analysis of T1 plant from the “Del/Ros1-N in MoneyMaker” stock line using iPCR.

```

FLB-B: 449 nt
TATTTAGTTGTGCCCCGTAGCAAATTGTTTGTCAATCATCTCTCTCCCTCAAACCTCTCGCTCTATCTCTCGTTCGCTTCCCTCTCGC
TTTTATACAAACACAAGTGTATAAAATTTGTTTTTATTATATAAAGCGAGAGAAAAATGTATAAATACATATATTTTGTTCCT
CTGCCAGATCTCGCTGCCACTCTCCCTAATCTCGTTCGCCACCCCTCGCCTTTCTCGCTTATACAAACAGAAGCGAAATGTATAAAT
TGTGTTTCTGTTTGTATAAAGTAAGAGAAAATTATATATACACATGCAAATACATATATTTTCTGCTATACACTTATAGTTATAC
AATAAAATACTCCCTGCCAGTTTCTTTGCCTTTCTCTCTTTCTAATTTTATACAAATTCAAATTGTATATAATTTCTCCCTTT
CTCATTTTATACAATTTAA

```

LB analysis T6 plant from the “Del/Ros1-N in MoneyMaker” stock line.

```

>rev-EF00613496
AAGGCAAAAGAAACTGGCAGGGGAGTATTTTATTGTATAACTATAAGTGTATAGGACGAAAATATATGTATTTGCATGTGTATAT
ATAATTTTCTCTTACTTTATACAAACAGAAACACAATTTATACATTTTCGCTTCTGTTTGTATAAGCGAGAAAGCGAGGGTGCGA
ACGAGATTAGGGAGAGTGGCAGCGAGATCTGGCAGAGGAAAACAAAAATATATGTATTTATACAATTTTCTCTCGCTTTATATAAT
AAAAACAAATTTTATACACTTGTGTTTGTATAAAAGTGAGAGGAAGCGAACGAGAGATAGAGCGAGAGTTTGAGGGAGAGAGATGA
TTGACAAACAATTTGCTACGGGGCACAACATAATATGGAGAAAAAGAAAGAGTAATTACCAATTTTTCATTCAAAAATGTAG
ATGTCCGCAGCGTTATTATAAAATGAAAGTACATTTTGATAAAACGACAAATTACGATCCGTCGTATTTATAGGCGAAAGCAATAA
ACAAATTATTCTAATTCGGAAATCTTTATTTTCGACGTGTCTACATTCACGTCCAAATGGGGGCTTAGATGAGAACTTCACGATCG
ATGCCTTGATTTTCGCCATTCCCAGATACCCATTTTCATCTTCAGATTGGTCTGAGATTATGCGAAAAATATACACTCATATACATAAA
TACTGACAGTTTGAGCTACCAATTCAGTGTAGCCCATTTACCTTACATAATTCACTCAAATGCTAGGCAGTCTGTCAACTCGGCGTC
AATTTGTCGGCCACTATACGATAGTTGCGCAAATTTTCAAAGTCCTGGCCTAACATCACACCTCTGTGCGCGGCGGGTCCCATTTG
TGATAAATCCACCCATCGGATGATCATGAGCGGAGAATTAAGGGAGTCACGTTATGACCCCGCCGATGACGCGGGACAAGCCGTT
TTACGT

```

NCBI - BLASTn 2 sequence alignment (Zhang et al., 2000) - 07-11-2019.

Query = T1

Subject = T6

NIH

U.S. National Library of Medicine

NCBI

National Center for Biotechnology Information

BLAST®

» blastn suite

Align Sequences Nucleotide BLAST

blastnblasttblasttblastx

BLASTn programs search nucleotide subjects using a nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#)

Clear

Query subrange [?](#)

From

To

ATATAAGCGAGAGAAATTGTATAAATACATATATTTTGTTCCTCTGCCAGATCTCGCTGCCACTCTCCCTAATCTCGTTTCCGACCTTCTCGCTTATACAAACAGAGCGAAATGTATAAATTGTCTTCTGTTTGTATATAAGTAAGAGAAATTATATATACATGCAAATACATATATTTTGGCTATACACTTATAGTTATACAATAAAATACTCCCTGCCAGTTCTTTGCCCTTCTCTCTTCTAATTATACAAATCAAAATGTATATAATTTCTCCCTTCTCATTTTATACAATTAA

Or, upload file

Browse...

No file selected. [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☒ Align two or more sequences [?](#)

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#)

Clear

Subject subrange [?](#)

From

To

TCCAGATAACCAATTTCACTCTCAGATTGSGCTGAGATTATGCGAAATATACACTCATATACATAAATCTGACAGTTTGAGCTACCAATTCAGTGTAGCCCAATACCTTACATAATTCATCTCAATGCTAGCGAGTCTGTCACTCGGCTCAATTTGTGCGGCACTATACGATAGTTGCGCAAATTTTCAAAATCTGCGCTAACATCACACTCTGTGCGGCGCGGCTCCCAATTTGTGATAAATCCACCAATGCGATATCTATGCGGAGATTAGGCGCTCACGTTATGACCGCGCGGATGACGCGGAGACAGCGGTTTACGT

Or, upload file

Browse...

No file selected. [?](#)

Program Selection

Optimize for

☒ Highly similar sequences (megablast)☐ More dissimilar sequences (discontiguous megablast)☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

BLAST

Search nucleotide sequence using Megablast (Optimize for highly similar sequences)☐ Show results in a new window

[Algorithm parameters](#)

BLAST® » blastn suite-2sequences » results for RID-W8SZDP13114

← Edit Search

Save Search

Search Summary ▾

Search Parameters

Program	blastn
Word size	28
Expect value	10
Hitlist size	100
Match/Mismatch scores	1,-2
Gapcosts	0,2.5
Low Complexity Filter	Yes
Filter string	L;m;
Genetic Code	1

Karlin-Altschul statistics

Lambda	1.33271	1.28
K	0.620991	0.46
H	1.12409	0.85

Results Statistics

Effective search space	410780
------------------------	--------

Download Alignment

Sequence Plot

CC

81

Sequence ID: Query_38495Length: 952Number of Matches: 1
Range 1: 1 to 379[Graphics](#) Next Match Previous Match

Alignment statistics for match #1

	Score	Expect	Identities	Gaps	Strand	
	695 bits(376)	0.0	378/379(99%)	0/379(0%)	Plus/Minus	
Query	1	TATTTAGTTGTGCCCCGTAGCAAATTGTTTGTCAATCATCTCTCTCCCTCAAACCTCTCGC	60			
Sbjct	379	TATTTAGTTGTGCCCCGTAGCAAATTGTTTGTCAATCATCTCTCTCCCTCAAACCTCTCGC	320			
Query	61	TCTATCTCTCGTTTCGCTTCCTCTCGCTTTTATACAAACACAAGTGTATAAAATTTGTTTT	120			
Sbjct	319	TCTATCTCTCGTTTCGCTTCCTCTCACTTTTATACAAACACAAGTGTATAAAATTTGTTTT	260			
Query	121	TATTATATAAAGCGAGAGAAAATTGTATAAATACATATATTTTTGTTTTCTCTGCCAGA	180			
Sbjct	259	TATTATATAAAGCGAGAGAAAATTGTATAAATACATATATTTTTGTTTTCTCTGCCAGA	200			
Query	181	TCTCGCTGCCACTCTCCCTAATCTCGTTCGCCACCCTCGCCTTTCTCGCTTATACAAACA	240			
Sbjct	199	TCTCGCTGCCACTCTCCCTAATCTCGTTCGCCACCCTCGCCTTTCTCGCTTATACAAACA	140			
Query	241	GAAGCGAAATGTATAAATTGTGTTTCTGTTTGTATAAAGTAAGAGAAAATTATATATACA	300			
Sbjct	139	GAAGCGAAATGTATAAATTGTGTTTCTGTTTGTATAAAGTAAGAGAAAATTATATATACA	80			
Query	301	CATGCAAATACATATATTTTCGTCTTATACACTTATAGTTATACAATAAAAAATACTCCCC	360			
Sbjct	79	CATGCAAATACATATATTTTCGTCTTATACACTTATAGTTATACAATAAAAAATACTCCCC	20			
Query	361	TGCCAGTTTCTTTTGCCTT	379			
Sbjct	19	TGCCAGTTTCTTTTGCCTT	1			

We conclude that the insert B has been stably maintained across 5 generations in the MicroTom genetic background.

4c - Stability of insert B in T6 vs F9 plants (RB analysis)

Alignment of genomic:RB T-DNA junction region of T6 plants from the “Del/Ros1-N in MoneyMaker” stock line and F9 plants from the “Del/Ros1-N in MicroTom” stock line..

RB analysis of T6 plant from the “Del/Ros1-N in MoneyMaker” stock line.

```
>EF00613497_EF00613497
CTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCTTGAGGAACTGGTAGCTGTTGTGGGCCTGTGGTCTCA
AGATGGATCATTAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCC
TGTAGACCTCAATTGCGAGCTTTCTAATTTCAAACCTATTCGGGGCCAATACATACATCACCTCTTTAATTTGCTCATACATTTTGT
CTTGATGCTCCGATTTAATCTTGTCTATTTTAATTTCTTCAACCTCATTTCTCATGCGTCATTTAAACATAAAATGTTGATTAAAG
ATTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAATTTAATAATAATATTTGACCCTTTT
```

RB analysis T6 plants from the Del/Ros1-N in MicroTom stock line.

```
>EF00613498_EF00613498
CTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCTTGAGGAACTGGTAGCTGTTGTGGGCCTGTGGTCTCA
AGATGGATCATTAATTTCCACCTTCACCTACGATGGGGGGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCC
TGTAGACCTCAATTGCGAGCTTTCTAATTTCAAACCTATTCGGGGCCAATACATACATCACCTCTTTAATTTGCTCATACATTTTGT
CTTGATGCTCCGATTTAATCTTGTCTATTTTAATTTCTTCAACCTCATTTCTCATGCGTCATTTAAACATAAAATGTTGATTAAAG
ATTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAATTTAATAATAATATTTGACCCTTTTACG
```

NCBI - BLASTn 2 sequence alignment (Zhang et al., 2000) - 07-11-2019.

Query = T6

Subject = F9

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information

BLAST » blastn suite

Align Sequences Nucleotide BLAST

blastn blasto blastx tblastn tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear Query subrange

AGCTGTTGGGCTGTGGTCTCAAGATGGATCAITTAATTCACCTTCACCTACGATGGGG
GGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGACCTCAAT
GCGAGCTTTCAATTTCAAACCTATTCGGGGCCAATACATACATCACCTCTTTAATTTGCTCA
TACATTTTGTCTGAGTCCGATTTAATCTTGTCTATTTTAATTTCTTCAACCTCATTTCTCATGCGTCATTTAAACATAAAATGTTGATTAAAG
ATTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAATTTAATAATAATATTTGACCCTTTT

Or, upload file Browse... No file selected.

Job Title

Enter a descriptive title for your BLAST search

☒ Align two or more sequences

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) Clear Subject subrange

AGCTGTTGGGCTGTGGTCTCAAGATGGATCAITTAATTCACCTTCACCTACGATGGGG
GGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGACCTCAAT
GCGAGCTTTCAATTTCAAACCTATTCGGGGCCAATACATACATCACCTCTTTAATTTGCTCA
TACATTTTGTCTGAGTCCGATTTAATCTTGTCTATTTTAATTTCTTCAACCTCATTTCTCATGCGTCATTTAAACATAAAATGTTGATTAAAG
ATTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAATTTAATAATAATATTTGACCCTTTTACG

Or, upload file Browse... No file selected.

Program Selection

Optimize for

☒ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

Algorithm parameters

U.S. National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastn suite-2sequences » results for RID-W8E2NEF1114

Edit Search
Save Search
Search Summary

Search Parameters

Program	blastn
Word size	28
Expect value	10
Hitlist size	100
Match/Mismatch scores	1,-2
Gapcosts	0,2.5
Low Complexity Filter	Yes
Filter string	L;m;
Genetic Code	1

Karlin-Altschul statistics

Lambda	1.33271	1.28
K	0.620991	0.46
H	1.12409	0.85

Results Statistics

Effective search space	159598
------------------------	--------

Download All

t Plot

☐ CD

Sequence ID: Query_5319Length: 412Number of Matches: 1
Range 1: 1 to 409[Graphics](#) Next Match Previous Match

Alignment statistics for match #1

	Score	Expect	Identities	Gaps	Strand	
	756 bits(409)	0.0	409/409(100%)	0/409(0%)	Plus/Plus	
Query	1		CTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCCTTGAGGAACTG			60
Sbjct	1		CTTTTCTTATCGACCATGTACGTAAGCGCTTACGTTTTTGGTGGACCCCTTGAGGAACTG			60
Query	61		GTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACCTTCACCTACGAT			120
Sbjct	61		GTAGCTGTTGTGGGCCTGTGGTCTCAAGATGGATCATTAATTTCCACCTTCACCTACGAT			120
Query	121		GGGGGGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGACC			180
Sbjct	121		GGGGGGCATCGCACCGGTGAGTAATATTGTACGGCTAAGAGCGAATTTGGCCTGTAGACC			180
Query	181		TCAATTGCGAGCTTTCTAATTTCAAACCTATTCGGGCCCAATACATACATCACCTCTTTAA			240
Sbjct	181		TCAATTGCGAGCTTTCTAATTTCAAACCTATTCGGGCCCAATACATACATCACCTCTTTAA			240
Query	241		TTTGCTCATACATTTTGTCTTGATGCTCCGATTTAATCTTGTTCTATTTTAATTCTTCAA			300
Sbjct	241		TTTGCTCATACATTTTGTCTTGATGCTCCGATTTAATCTTGTTCTATTTTAATTCTTCAA			300
Query	301		CCTCATTTTCTCATGCGTCATTTAAACATAAAATGTTGATTAAAGATTCATGAAATTGCTT			360
Sbjct	301		CCTCATTTTCTCATGCGTCATTTAAACATAAAATGTTGATTAAAGATTCATGAAATTGCTT			360

84

```

Query  361  ACCTTTAACGTAAGTGAGAAAAAATTTAATAATAATATTTGACCCTTTT  409
        ||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ACCTTTAACGTAAGTGAGAAAAAATTTAATAATAATATTTGACCCTTTT  409

```

We conclude that the insert B has been stably maintained through mendelian crossing and across generations in different genetic backgrounds.

5 – CONCLUSION

The original *Del/Ros1-N* transformation event contained two distinct T-DNA insertions, one (insert A) on chromosome 2 and a second (insert B) on chromosome 4. Following Mendelian segregation, tomato plants exhibiting a stable purple fruit phenotype and containing only the T-DNA insert at locus B were retained. The combined analysis of the LB and RB T-DNA: genome junction regions of insert B showed that a small deletion of the tomato genome (94 bp), T-DNA RB region (52 bp) and T-DNA LB region (75 bp) occurred at the point of insertion. The flanking sequences on insert B were identical between T1 and T6 generations and in the F9 generation of an outcross to MoneyMaker confirming the stability of the locus across generation and genetic backgrounds.

Appendix 5: Southern blot and PCR analysis of insert copy number and presence of vector backbone

N.B. This information is included only for completeness. It has been exhaustively corroborated and superseded by the complete genome sequencing of the T6 MicroTom and commercial lines described in Appendices 2 and 3.

Southern blots of DNA from the T0 Del/Ros1-N tomato plant, from other tomato lines from the transformation with the pDEL.ROS1 binary vector, and from a wild type control confirmed 2 insertions of the T-DNA containing E8:Del and E8:Ros1 in the plant carrying the Del/Ros1-N event in the T0 generation, in comparison to other transformation events (C,Y,Z) that were not taken forward (Figure A). Following propagation through subsequent generations the T-DNA insertion at locus A was lost from the Del/Ros1-N event, through segregation. Details of this analysis follow.

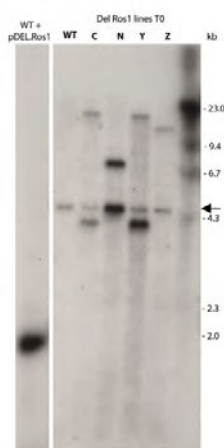
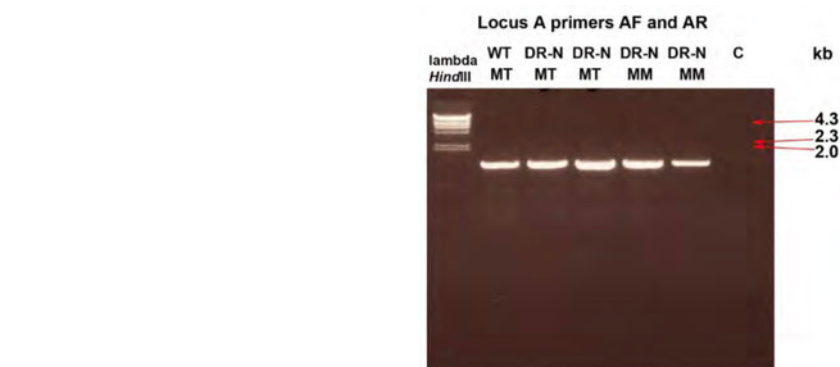


Figure A. Genomic DNA from WT Red tomato and 4 T0 lines of tomato (C,N,Y,Z) transformed with pDEL.ROS1. Genomic DNA was digested with *Bam*HI and probed with the *Ros1* cDNA. The arrow indicates DNA homologous to the *Ros1* cDNA present in tomato genomic DNA (*ANT1* and *ANT2* genes). The track labelled WT+pDEL.ROS1 is a copy number reconstruction for one copy of the T-DNA in pDEL.ROS1 inserted into the tomato genome.

In addition, copy number as assessed by PCR analysis to detect the NPTII selectable marker gene, confirming the presence of 2 copies of the insert in event N (see below). Inverse PCR (iPCR) was performed to identify the sequences flanking both inserts (see below). DNA from plants with purple fruit of the T6 generation of Del/Ros1-N in MicroTom and from plants with purple fruit from an outcross of Del/Ros1-N (T1) x Money Maker (selfed over 9 generations) were analyzed for T-DNA insertions at locus A and locus B by PCR using primers developed from the flanking sequences of both loci A and B. The T-DNA insertion at locus A was not detected in either of these derived stock lines (Figure B) indicating that it had segregated away during the selfing that generated the lines. This was confirmed by attempts to amplify the T-DNA insertion using primers from the left and right borders of the T-DNA and the flanking sites at locus A (Figure B). Using primers from the sequences flanking the original insertion at locus A, a wild type fragment of DNA of 1.25 kb was amplified from both stock lines (Figure B) as well as from the control WT DNA from MicroTom. Sequencing of this fragment of DNA confirmed that it

was the wild type sequence from MicroTom with no footprints and that the T-DNA insertion at locus A had segregated away rather than been lost by recombination.



Primer	Sequence
AR	5'-GGGTGTAATTACGAATTCCTTGC-3'
AF	5'-CATGATCGAATAAACTTTACTGT-3'

Figure B. PCR of genomic DNA from WT MicroTom (MT) and Del/Ros1-N stock plants in MicroTom (MT) or MoneyMaker (MM) genetic backgrounds, using two primers (AF and AR) separated by 1.25kbp on either side of T-DNA insertion A detected in analysis of Del/Ros1-N T1 plants. Only the 1.25 kbp fragment was detected in the stock lines, which was identical to the sequence amplified from the Wild Type (MT) plant. This confirmed that the T-DNA insertion at locus A, detected in some T1 plants, was lost by subsequent segregation in generating the stock Del/Ros1-N lines in both MicroTom and MoneyMaker.

The T-DNA insertion from locus B was detected in both MicroTom and MoneyMaker-derived stock lines (Figure C). Comparison of the sequences flanking the left and right borders of the T-DNA between T1 vs T6 MicroTom plants and F9 MoneyMaker vs T6 MicroTom plants confirmed the stability of the locus B insertion in the tomato genome across generations and genetic backgrounds at position 62904771 on chromosome 4 (see Appendix 4). Note that the whole genome sequence analysis confirmed the conclusions drawn from inverse PCR analysis with respect to the position of locus B in the genome.

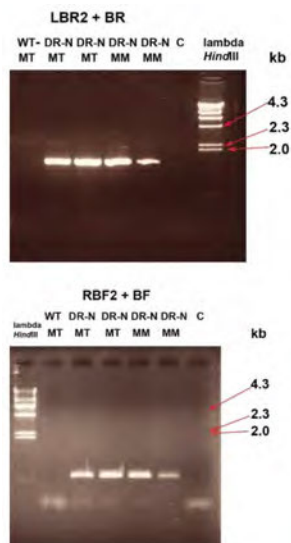


Figure C. PCR amplification of genomic DNA from individual plants of purple tomato *Del/Ros1-N* in MicroTom (MT) and MoneyMaker (MM) stock lines compared to a wild type (WT) MicroTom plant. LBF2 and RBF2 are primers from the end of the T-DNA for the right and left border respectively (see Appendix 4). BR is a primer from the sequence of locus B flanking the left border of the T-DNA in the original iPCR analysis. Between primers LBR2 and BR a fragment of 1.39 kbp would be predicted if there was a T-DNA insertion there. In fact a band of 1.15 kbp was amplified. BF represents sequence at locus B originally detected flanking the right border of the T-DNA insertion. Between primers RBF2 and BF a fragment of 0.83 kbp would be predicted if there was a T-DNA insertion at locus B. In fact, a band of 0.72 kbp was amplified. The DNA fragments encoding the flanking sequences at locus B were sequenced directly and indicated that the T-DNA insertion involves loss of 52 bp from the end of the right border region and 75bp from the end of the left border region together with a loss of 94 bp of locus B at the insertion site, as depicted in Figure 5 of the original submission, and supported by subsequent whole genome sequencing analysis.

Analysis of T-DNA copy and loci number

Pete Isaac¹ & Cathie Martin²

¹ iDNA Genetics Ltd, Norwich Research Park, Norwich NR4 7UH, UK

² John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

Table of Content

- 1 – PLANT MATERIAL
- 2 – Q-PCR PROCEDURE
- 3 – Q-PCR RESULTS
- 4 – CONCLUSION

T-DNA copy and loci number in event *Del/Ros1-N* was estimated by quantification of the number of *nptII* gene copies in progenies obtained by self-pollination of a T0 plant containing the *Del/Ros1-N* transformation event. The *nptII* copy number quantification was sub-contracted to “iDNA Genetics Ltd” (Norwich, UK) which has extensive expertise in PCR-based diagnostics and transgene copy number determination in plant systems.

1 - PLANT MATERIAL

Leaf material from 8 progeny plants obtained by self-pollination of the SCC77 plant containing the *Del/Ros1-N* transformation event was harvested and kept frozen prior to DNA extraction and quantitative polymerase chain reaction (q-PCR) analysis. Samples were named SCC77_1 to SCC77_8.

2 - Q-PCR PROCEDURE

qPCR procedures and raw data are detailed in the “iDNA Genetics Ltd” record file “10M0229_5_070311_archive”. Briefly: the q-PCR conditions using the PowrSybr mix (Applied Biosystems) were:

	1rxn	150 rxns	Batch
water	3	450	Elga070311
2 X power sybr	10	1500	1008236
10 x Primers	2	300	PI070311
DNA	5		PI040311
		2250	
		281.25	in distrib plate
Place	15	ul in wells	
Add	5	uL DNA	
		Plate Make	sarstedt 721984 202
		Plate Batch	S22768
		Seal Make	sarstedt 95.1994
		Seal Batch	2012-08-IA
PCR	Vesuvias		
Denaturation	95C	10 min	
40 cycles	95C	15s	
	60C	60sec	Read
Follow with Dissociation			

Primers:

Primer	Sequence
tn5-1F	GGGCAGGATCTCCTGTCATCT
npt2-2r	GTTTCGCTTGGTGGTCAAT
SILat52F2	GAGACCACGAGAACGATATTTGC
SILat52R2	CAGCATTGTGGATACCAACATTG

3 – Q-PCR RESULTS

SDS 2.2	AQ Results
Filename	10M0029_5_070311
PlateID	
Assay Type	Absolute Quantification
Run DateTime	03/07/2011 16:22
Operator	
ThermalCycleParams	

Table 1: Raw qPCR data

Sample	Ct	AverageCt	SdCt	Sample	Ct	AverageCt	SdCt	DeltaCt	1st est	EstCopies	reported copies
npt2_SCC77_1	27.929525	27.91736375	0.04508179	SILat52_SCC77_1	25.27417	25.354285	0.091357251	2.56307875	0.5	0.468434202	<1
npt2_SCC77_2	26.793041	26.852336	0.045823539	SILat52_SCC77_2	25.311878	25.333372	0.078627622	1.518964	1	0.984565839	1
npt2_SCC77_3	27.203972	27.24655875	0.034410023	SILat52_SCC77_3	26.24801	26.303818	0.08563857	0.94274075	2	1.4834683	2
npt2_SCC77_4	26.554138	26.37915375	0.135689253	SILat52_SCC77_4	26.060724	25.99984675	0.063718402	0.379307	2	2.214931271	2
npt2_SCC77_5	27.782833	27.6942785	0.067651551	SILat52_SCC77_5	26.905806	26.96215375	0.092967273	0.73212475	2	1.72326273	2
npt2_SCC77_6	25.762545	25.86610825	0.142073621	SILat52_SCC77_6	26.081293	26.09856	0.046639886	-0.23245175	3 to 4	3.422733065	3 to 4
npt2_SCC77_7	27.064577	27.0657415	0.057005174	SILat52_SCC77_7	25.62494	25.51034125	0.09098914	1.55540025	1	0.959372341	1
npt2_SCC77_8	27.673742	27.5729405	0.081060861	SILat52_SCC77_8	26.851547	26.85024025	0.092626325	0.72270025	2	1.73485566	2

Table 2: Summary of observed *nptII* copy number

Sample	reported copies
npt2_SCC77_1	<1
npt2_SCC77_2	1
npt2_SCC77_3	2
npt2_SCC77_4	2
npt2_SCC77_5	2
npt2_SCC77_6	3 to 4
npt2_SCC77_7	1
npt2_SCC77_8	2

The observed pattern of distribution of *nptII* transgene copy number (Table 2) is consistent with the integration of two T-DNA copies inserted at two genetically unlinked Mendelian loci (one T-DNA copy per locus, Table 3) in the tomato genome.

Table 3: Theoretical number of gene copies in the progeny of a transgenic plant containing two genetically unlinked transgenic loci with one T-DNA present at each locus.

	1 – 1	1 – 0	0 – 1	0 – 0
1 – 1	4	3	3	2
1 – 0	3	2	2	1
0 – 1	3	1	1	1
0 – 0	2	1	1	0

4 - CONCLUSION

We estimate, based upon the distribution of *nptII* gene copy number in *Del/Ros1-N* progeny plants, that this event is composed of two genetically unlinked transgenic loci each harboring a single T-DNA copy.

Analysis of T-DNA flanking sequences

Pete Isaac¹, Eugenio Butelli² & Cathie Martin²

¹ *iDNA Genetics Ltd, Norwich Research Park, Norwich NR4 7UH, U6*

² *John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK*

Table of Content

- 1 - PLANT MATERIAL
- 2 – i-PCR PROCEDURES & RESULTS
- 3 - PCR VALIDATION PROCEDURES & RESULTS
- 4 - CONCLUSION

T-DNA flanking sequences in the tomato genome of *Del/Rose1-N* (T1) were retrieved by i-PCR and sequenced. The T-DNA:plant genome junction regions were validated by polymerase chain reaction (PCR). The iPCR work was sub-contracted to “iDNA Genetics Ltd” (Norwich, UK) which has extensive expertise in PCR-based diagnostics in plant systems.

1 - PLANT MATERIAL

Leaf material from 2 progeny plants obtained by self-pollination of the T0 plant containing the *Del/Ros1-N* transformation event, were harvested and kept frozen prior to DNA extraction and iPCR.

Samples were named 1_mse & 2_mse.

2 - i-PCR PROCEDURES & RESULTS

iPCR procedures and raw data are detailed in the “iDNA Genetics Ltd” record files: “12M0006_2_ligation_3_iPCR_030212” and “12M0006_package_070212”

Briefly: the conditions for ligation, cleanup and i-PCR amplification are detailed below:

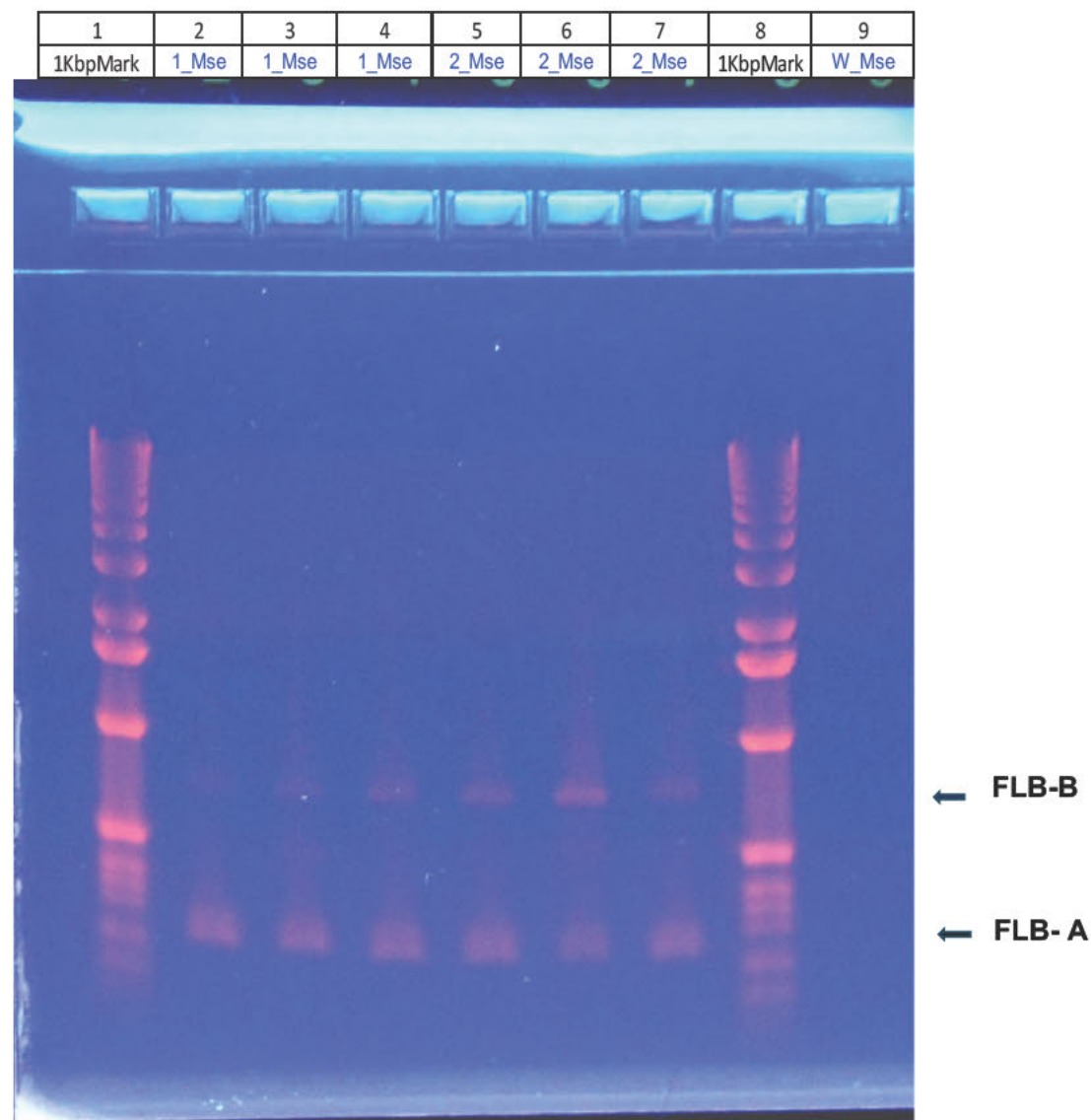
<u>Ligation</u>			
Samples are Mse1 digested, >1000 templates /4uL from 12M0006_1_030211			
Ligate 5000 templates in a 40uL reaction			
Buffer includes ATP			
Ligase	T4DNA		
Supplier	NEB		
Reaction volume 40 uL			
	1 rxn	7 rxnPremix	Batch
water	14	98	Elga030212
10x NEBLigationBuffer	4	28	0031003
Ligase (400u/uL)	2	14	0991004
DNA (50ng)or Water	20		PI201211
Volume	40	140	
	Place	20	uLper tube
	Add	20	uL DNA
Incubate overnight (>960 mins) in Stromboli_96 (ipcr-Ligation)			
12 cycles	Start 16C	decreasing 1C/cycle	
	Each cycle	30 mins	
	Hold at 4C		
Time on	18:12 030212		

Cleanup					
Spin 60 seconds, discard flow through and place column back in same tube					
Add 0.75mL buffer PE to the column and centrifuge for 60 seconds					
Discard flow through, place in same tube and respin additional 1 minute.					
Transfer column to fresh 1.5mL tube					
Add 100uL buffer EB to the centre of the membrane.					
Leave at room temperature 1 minute, then centrifuge 1 minute to collect the DNA					

PCR			
PCR uses typelt			
DNA			
DNA samples are column purified ligations (6 tubes).			
The ligations of DNAs 1 and 2 to be done in triplicate			
Primers			
		TomMseChkF1	TCCTGGCCTAACATCACACCT
		Csp6ILBR1	CATCTAAGCCCCATTGGAC
Make pairs (10 x stock = 2uM each primer 5uL+5uL+240uLwater)			
Master Mix			
MinusQ		12	
	1 Rxn + Q	12 rxns	Batch
water	3	36	Qiagen139280916
Typelt 2x SSR kit	10	120	Qiagen139288369
Primers	2	24	VC060212
DNA	5		LigationMix_Pi030212, post col
Total	20	180	
		360	
Place	15	uL into four wells	
add	5	of DNA or water	
		Plate Make	AbGene_96well non skirted
		Plate Batch	
		Seal Make	Thermo adhesive foil AB-0626
		Seal Batch	523316
Run conditions Stromboli _96 20ul			
activate	95C	5 min	
45 cycles	95C	15 s	
	60C	30.00	
	72.00	60.00	

i-PCR amplification results

Two distinct iPCR products (approximately 250 bp and 750 bp in size) were retrieved following analysis of the regions flanking the left border (LB) of the T-DNA inserts. These two bands were gel purified, cloned and subsequently sequenced (FLB-A & FLB-B respectively). No PCR product could be retrieved following the amplification of the regions flanking the right border (RB) of the T-DNA.



Sequencing of the genomic regions flanking the Left Border of the T-DNA:

FLB-A: 308 nt

```
AATTTATCACTTGTTATACATAAATTATTCTATACTTTCTCTATCATCTTTACAGTCCTTGATGTTGAA
GCTTCATATCTCAAAACGACAAACATGGTCATATCATATAAAATTGAATAAAACTACATGACAAGTT
CTAGTGAGGACACAAAGTGTGACTCATGTGAGCCTATATGCTCTAACCTTGATGCATATGTTGTTG
```

GATCAAACTATATGCACAACTGAAGCCTACAATATTGCTTATAAGAATAATATATATCAATAATATT
CATGTTAGAACACATATAATCATGCTAAAACACATTAA

FLB-B: 449 nt

TATTTAGTTGTGCCCCGTAGCAAATTGTTTGTCAATCATCTCTCTCCCTCAAACCTCTCGCTCTATC
TCTCGTTTCGCTTCCTCTCGCTTTTATACAAACACAAGTGTATAAAATTTGTTTTATTATATAAAG
CGAGAGAAAATTGTATAAATACATATATTTTGTTCCTCTGCCAGATCTCGCTGCCACTCTCCC
TAATCTCGTTCGCCACCCTCGCCTTTCTCGCTTATACAAACAGAAGCGAAATGTATAAATTGTGTTT
CTGTTTGTATAAAGTAAGAGAAAATTATATATACACATGCAAATACATATATTTTCGTCCTATACACT
TATAGTTATACAATAAAAATACTCCCCTGCCAGTTTCTTTTGCCTTTCTCTCTTTCTAATTTTATACA
AATTCAAATTGTATATAATTTCTCCCTTTCTCATTTTATACAATTTAA

The sequence in purple was subsequently validated as junction sequences by PCR analysis
(see section 3)

BLAST analysis of FLB-A & FLB-B:

Homology of FLB-A and FLB-B to the tomato genome were searched by BLAST analysis
against "Tomato WGS chromosomes (SL 2.50)" <http://solgenomics.net/tools/blast/>

FLB-A

BLAST analysis returned a single significant hit on a contig not anchored on the tomato
genome.

[Bioperl](#) Reformatted HTML of BLASTN Search Report for FLB-A

**Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng
Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new
generation of protein database search programs", Nucleic Acids Res. 25:3389-3402. Query=**
FLB-1

(308 letters)

Database: Tomato WGS Chromosomes (SL2.50)
13 sequences; 823,944,041 total letters

Sequences producing significant alignments:	Score	E
	(bits)	value
SL2.50ch00	611	1e-173

> [SL2.50ch00](#)

Length = 21,805,821

Score = 611 bits (308), Expect = 1e-173

Identities = 308/308 (100%), Frame = +1 / -1

```
Query: 1      aatttatcacttggtatacataaattattctatactttctctatcatctttacagtctt 60
              |||
Sbjct: 14105  aatttatcacttggtatacataaattattctatactttctctatcatctttacagtctt 14046
```

```
Query: 61      gatgttgaagcttcatactctcaaaacgacaaacatgggtcatatcatataaaattgaataa 120
              |||
Sbjct: 14045  gatgttgaagcttcatactctcaaaacgacaaacatgggtcatatcatataaaattgaataa 13986
```

```
Query: 121     aactacatgacaagttctagtgaggacacaaagtgtgactcatgtgagcctatatgctct 180
              |||
Sbjct: 13985  aactacatgacaagttctagtgaggacacaaagtgtgactcatgtgagcctatatgctct 13926
```

```

Query: 181   aaccttgatgcatatgttggttgatcaaaactatatgcacaactgaagcctacaatattg 240
            |||
Sbjct: 13925 aaccttgatgcatatgttggttgatcaaaactatatgcacaactgaagcctacaatattg 13866

Query: 241   cttataagaataatatatatcaataatattcatgttagaacacatataatcatgctaaaa 300
            |||
Sbjct: 13865 cttataagaataatatatatcaataatattcatgttagaacacatataatcatgctaaaa 13806

Query: 301   cacattaa 308
            |||
Sbjct: 13805 cacattaa 13798

```

Search Parameters

Parameter	Value
allowgaps	yes
expect	1.0e-10
gapext	2
gapopen	5
matrix	blastn matrix:1 -3

Search Statistics

Statistic	Value
Hits_to_DB	5,050,660
S1	20
S1_bits	40.1
S2	36
S2_bits	71.9
X1	11
X1_bits	21.8
X2	15
X2_bits	29.7
X3	50
X3_bits	99.1
dbentries	13
dbletters	823944041
effectivedblength	823943794
effectivespace	238119756466
effectivespaceused	238119756466
entropy	1.31
entropy_gapped	1.31
kappa	0.711
kappa_gapped	0.711
lambda	1.37
lambda_gapped	1.37
length_adjustment	19
num_extensions	16

num_successful_extensions	16
number_of_hsps_gapped	16
number_of_hsps_successfully_gapped	1
posted_date	Mar 31, 2014 4:41 PM
querylength	289
seqs_better_than_cutoff	1

Produced by Bioperl module Bio::SearchIO::Writer::HTMLResultWriter on Thu Jul 17 08:43:54 2014
Revision: \$Id\$

FLB-B

BLAST analysis returned a series of positive hits among which one exhibited a significantly higher homology (SL2.50ch04) suggesting a T-DNA insertion at position 62818821 on chromosome 4.

[Bioperl](#) Reformatted HTML of BLASTN Search Report for FLB-B

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402. Query=
FLB-2
(185 letters)

Database: Tomato WGS Chromosomes (SL2.50)

13 sequences; 823,944,041 total letters

Sequences producing significant alignments:	Score (bits)	E value
SL2.50ch04	357	4e-97
SL2.50ch10	172	1e-41
SL2.50ch09	170	5e-41
SL2.50ch11	165	3e-39
SL2.50ch03	161	4e-38
SL2.50ch05	157	7e-37
SL2.50ch01	157	7e-37
SL2.50ch02	155	3e-36
SL2.50ch07	153	1e-35
SL2.50ch06	153	1e-35
SL2.50ch12	147	7e-34
SL2.50ch08	121	4e-26
SL2.50ch00	111	4e-23

> [SL2.50ch04](#)

Length = 66,470,942

Score = 357 bits (180), Expect = 4e-97

Identities = 183/184 (99%), Frame = +1 / +1

```
Query: 2          atttagttgtgccccgtagcaaattgtttgtcaatcatctctctccctcaaactctcgct 61
                  |||
Sbjct: 62818821  atttagttgtgccccgtagcaaattgtttgtcaatcatctctctccctcaaactctcgct 62818880

Query: 62          ctatctctcgttcgcttcctctcgctttttatacaaacacaaagtgtataaaatttggtttt 121
                  |||
Sbjct: 62818881  ctatctctcgttcgcttcctctcactttttatacaaacacaaagtgtataaaatttggtttt 62818940

Query: 122         attatataaaagcgagagaaaattgtataaatacatatatattttgttttcctctgccagat 181
                  |||
Sbjct: 62818941  attatataaaagcgagagaaaattgtataaatacatatatattttgttttcctctgccagat 62819000

Query: 182         ctcg 185
                  |||
Sbjct: 62819001  ctcg 62819004
```

Search Parameters

Parameter	Value
allowgaps	yes
expect	1.0e-10
gapext	2
gapopen	5
matrix	blastn matrix:1 -3

Search Statistics

Statistic	Value
Hits_to_DB	3,354,084
S1	20
S1_bits	40.1
S2	36
S2_bits	71.9
X1	11
X1_bits	21.8
X2	15
X2_bits	29.7
X3	50
X3_bits	99.1
dbentries	13
dbletters	823944041
effectivedblength	823943794
effectivespace	136774669804
effectivespaceused	136774669804
entropy	1.31
entropy_gapped	1.31
kappa	0.711
kappa_gapped	0.711
lambda	1.37
lambda_gapped	1.37
length_adjustment	19
num_extensions	5479
num_successful_extensions	5479
number_of_hsps_gapped	4667
number_of_hsps_successfully_gapped	1048
posted_date	Mar 31, 2014 4:41 PM
querylength	166
seqs_better_than_cutoff	13

3 - PCR VALIDATION PROCEDURES & RESULTS

The T-DNA: FLB-B junction sequence was validated following PCR amplification using B-F1 and LB-R2 oligonucleotides present in both the T-DNA (LB) and genomic region (FLB-2). In *Del/Ros1-N* plants of the T6 generation (MicroTom stock line)

Primer sequences

B-F1 AGAAAGGCGAGGGTGGCGAACGA
LB- CAGTGGCTCCTTCAACGTTGCGGT
R2

x38 cycles

94°C 30"

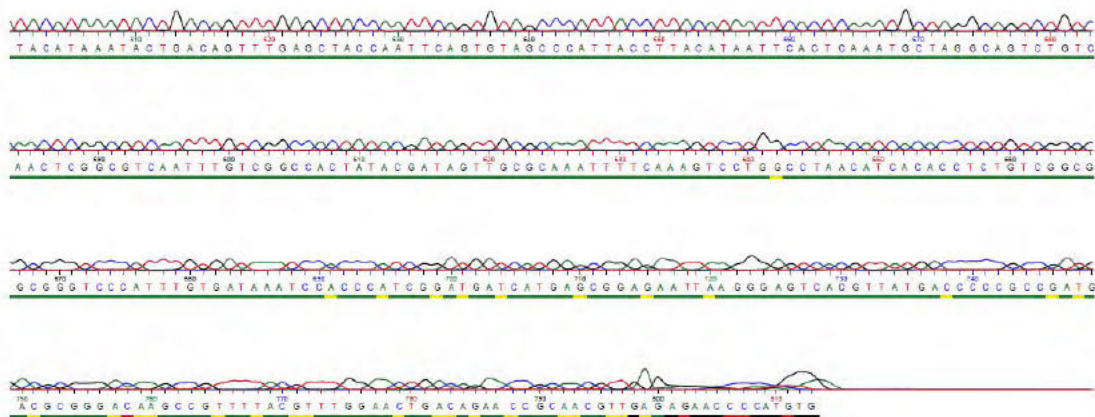
64°C 40"

72°C 50"

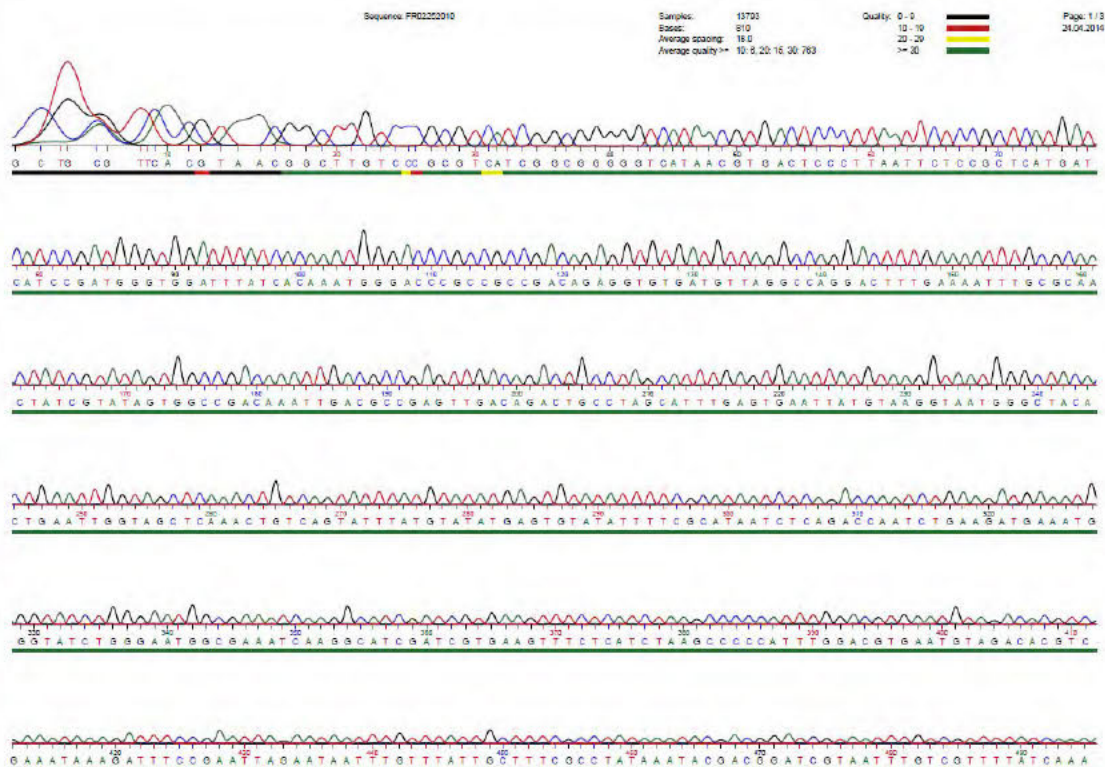
Sanger sequencing of the PCR products was undertaken using the B-F1 and LB-R2 oligonucleotides

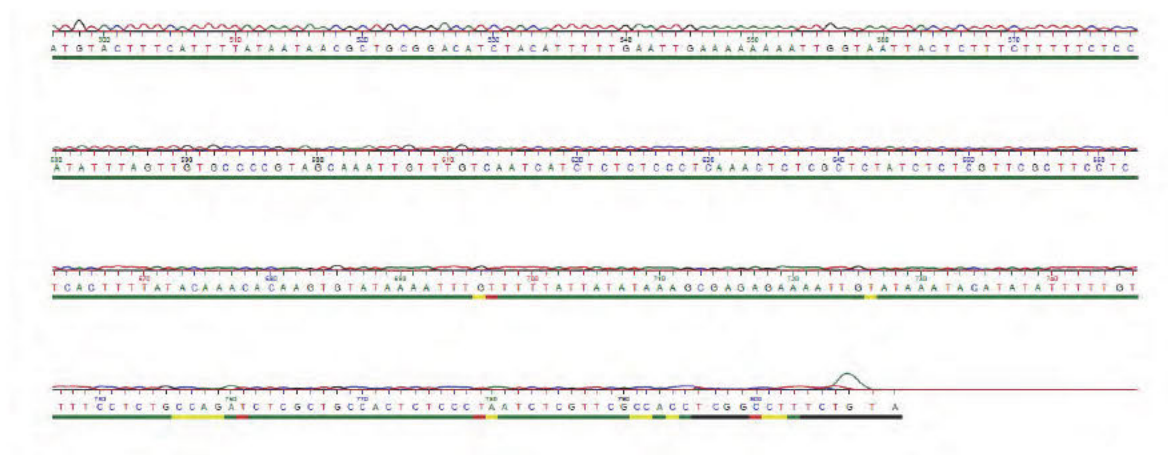
Sequence from T-DNA to flanking region (LB-R2 primer):



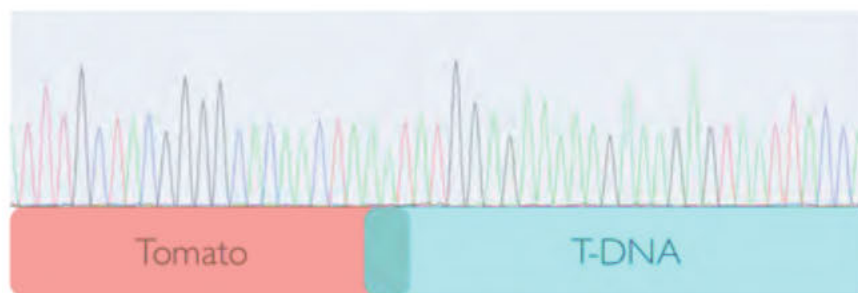


Sequence from flanking region to T-DNA (B-F1 primer):





Assembled validation sequences:



> FLB-2 (highlighted in red) - T-DNA (highlighted in blue)

```
CGAGATCTGGCAGAGGAAAAACAAAATATATGTATTATACAATTTTCTCTCGCTTTATATAATAAAAAACAAATTTTATACACTTGTGTTTGATAAAAGTGAGAG
GAAGCGAACGAGAGATAGAGCGAGAGTTTGAGGGAGAGAGATGATTGACAAACAATTTGCTACGGGGCACAACTAAATATGGAGAAAAAGAGTAATTACCAA
TTTTTTTTCAATTCAAAAATGTAGATGTCCGCGAGCTTATTATAAAATGAAAGTACATTTTGATAAAACGACAAATTACGATCCGTCGTATTATAGGCGAAAGCA
ATAAACAAATTAATCTAATTCGAAATCTTTATTTGACGTGTCTACATTACGTCCTCAAAATGGGGGCTTAGATGAGAACTTCACGATCGATGCCCTTGATTTCGCC
ATTTCCAGATACCCATTTTCATCTTCAGATTGGTCTGAGATTATGCGAAAAATACACTCATATACATAAACTGACAGTTTGAGCTACCAATTCAGTGTAGCCCA
TTACCTTACATAATTAATCAATGCTAGGCAGTCTGTCAACTCGGCGTCAATTTGTGCGGCACTATACGATAGTTGCGCAAAATTTCAAAGTCTGGCCTAACAT
CACACCTCTGTGCGCGGCGGTCCATTTGTGATAAATCCACCATCGGATGATCATGAGCGGAGAATTAAAGGGAGTCACGTTATGACCCCGCCGATGACGCGGG
ACAAGCCGTT
```

BLAST analysis against the tomato genome "Tomato WGS chromosomes (SL 2.50)" <http://solgenomics.net/tools/blast/> returned a validation sequence identical to the one previously identified by i-PCR amplification of the junction sequences (T-DNA insertion at position 62818821 on chromosome 4).

[Bioperl](#) Reformatted HTML of BLASTN Search Report for flanking sequence

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402. **Query=** Flanking Sequence

(185 letters)

Database: Tomato WGS Chromosomes (SL2.50)

13 sequences; 823,944,041 total letters

Sequences producing significant alignments:		Score (bits)	E value
SL2.50ch04		365	2e-99
SL2.50ch10		180	5e-44
SL2.50ch09		178	2e-43
SL2.50ch11		172	1e-41
SL2.50ch03		168	2e-40
SL2.50ch05		165	3e-39
SL2.50ch01		165	3e-39
SL2.50ch02		163	1e-38
SL2.50ch07		161	4e-38
SL2.50ch06		161	4e-38
SL2.50ch12		155	3e-36
SL2.50ch08		125	2e-27
SL2.50ch00		113	9e-24

> [SL2.50ch04](#)

Length = 66,470,942

Score = 365 bits (184), Expect = 2e-99

Identities = 184/184 (100%), Frame = +1 / -1

```
Query: 1          cgagatctggcagaggaaaaacaaaaatatatgtattttatacaattttctctcgctttata 60
                |||
Sbjct: 62819004  cgagatctggcagaggaaaaacaaaaatatatgtattttatacaattttctctcgctttata 62818945

Query: 61          taataaaaaacaaatttttatacacttggtgtttgtataaaaagtgagaggaagcgaacgagag 120
                |||
Sbjct: 62818944  taataaaaaacaaatttttatacacttggtgtttgtataaaaagtgagaggaagcgaacgagag 62818885

Query: 121         atagagcgagagtttgagggagagagatgattgacaaacaatttgctacggggcacaact 180
                |||
Sbjct: 62818884  atagagcgagagtttgagggagagagatgattgacaaacaatttgctacggggcacaact 62818825

Query: 181         aaat 184
                |||
Sbjct: 62818824  aaat 62818821
```

Search Parameters

Parameter	Value
allowgaps	yes
expect	1.0e-10
gapext	2
gapopen	5
matrix	blastn matrix:1 -3

Search Statistics

Statistic	Value
Hits_to_DB	3,430,207
S1	20
S1_bits	40.1
S2	36
S2_bits	71.9
X1	11
X1_bits	21.8
X2	15
X2_bits	29.7
X3	50
X3_bits	99.1
dbentries	13
dbletters	823944041
effectivedblength	823943794
effectivespace	136774669804
effectivespaceused	136774669804
entropy	1.31
entropy_gapped	1.31
kappa	0.711
kappa_gapped	0.711
lambda	1.37
lambda_gapped	1.37
length_adjustment	19
num_extensions	5459
num_successful_extensions	5459
number of hsps gapped	4655
number_of_hsps_successfully_gapped	1045
posted_date	Mar 31, 2014 4:41 PM
querylength	166
seqs_better_than_cutoff	13

Produced by Bioperl module Bio::SearchIO::Writer::HTMLResultWriter on Fri Jul 18 04:13:21 2014
Revision: \$Id\$

Attempts to validate insertion A in the T6 generation were unsuccessful. No evidence of a second insertion at A was obtained from PCR with primers detecting *Del*, *Ros1* or *np11*.

4 – CONCLUSION

The analysis of the genomic regions flanking the left border (LB) of the T-DNA confirmed the presence of two distinct T-DNA insertions in event *Del/Ros1-N* in the T1 generation: one T-DNA insertion on an unanchored genomic contig SL2.50ch00 at position 14105 subsequently identified to lie on chromosome 2 (insertion A) and a second T-DNA insertion on chromosome 4 (SL2.50ch04) at position 62818821 (insertion B). Insertion B was confirmed as present in

subsequent generations of *Del/Ros1-N* and in out crosses introducing the purple trait into other genetic backgrounds. We were unable to validate the presence of insertion A in subsequent generations of *Del/Ros1-N* or in progeny of any outcrosses. This led us to the conclusion that insertion A had segregated away during the generation of tomato lines breeding true for the purple tomato phenotype.

Absence of tetracycline resistance gene

Pete Isaac¹ & Cathie Martin²

¹ *iDNA Genetics Ltd, Norwich Research Park, Norwich NR4 7UH, UK*

² *John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK*

Table of Content

- 1 - PLANT MATERIAL
- 2 - PCR PROCEDURE
- 3 - PCR RESULTS
- 4 - CONCLUSION

The detection of the tetracycline resistance gene was performed by quantitative polymerase chain reaction (q-PCR). This work was sub-contracted to “iDNA Genetics Ltd” (Norwich, UK) which has extensive expertise in PCR-based diagnostics in plant systems.

1 - PLANT MATERIAL

Leaf material from 5 progeny plants obtained by self-pollination of the SCC90 plant (T1) containing the *Del/Ros1-N* transformation event were harvested and kept frozen prior to DNA extraction and q-PCR analysis. Samples were named:

Tet_SCC90_1
Tet_SCC90_2
Tet_SCC90_3
Tet_SCC90_4
Tet_SCC90_5

2 - PCR PROCEDURE

q-PCR procedures and raw data are detailed in the “iDNA Genetics Ltd” record files: “10M0029_8_030511”, “10M0229_6_250311” & “10M0229_7_250311”

Briefly: the conditions for PCR amplification are detailed below:

Selection of best primer pair:

3 forward and 3 reverse oligonucleotides (9 combinations) designed against the Tet sequence were tested using diluted plasmid DNA as a substrate.

DNA was supplied at 0.1pg/uL, have made 1:100 and 1:10000 dilutions from this

Here us 5uL per xn of 0.001pg/uL plasmid ie 5 x 10-3pg plasmid on all 9

pairw ise combinations of primers

Tet04541F1	TCGCGTCCCAGTGAAAGC
Tet04541R1	CGGTCTTCTTCATCATGCAACTT
Tet04541F2	TGGTCGCGTCCCAGTGA
Tet04541R2	TGGCGGTCTTCTTCATCATG
Tet04541F3	AGTGAATGCAGAATGCCAAATG
Tet04541R3	GCCCTGATGGCGGTCTT

Dissolve primers to 100uM then make 10x (2uM) stocks

	conc (uM)	F1R1	F1R2	F1R3	F2R1	F2R2	F2R3	F3R1	F3R2	F3R3
Water		480	480	480	480	480	480	480	480	480
F	100	10	10	10	10	10	10	10	10	10
R	100	10	10	10	10	10	10	10	10	10

Make 2 master mixes with PowerSYBR mix (one with plasmid, one with non-transgenic DNA)

So make for20 reactions

	1rxn	20 rxns	Batch
water	3	30	Elga250311
2 X power sybr	10	100	1008236
10 x Primers	2		VC250311
DNA(plasmid/nontrans)	5	50	1 to 100n diln

Place18 ul in wells

Add2 ul primers

Plate Make	sarstedt 721984 202
Plate Batch	S22768
Seal Make	sarstedt 95.1994
Seal Batch	2012-08-1A

PCR

Vesuvias

Denaturation	95C	10 min	
40 cycles	95C	15s	
	60C	60sec	Read

Follow with Dissociation

Tet04541F2 & Tet04541R1 was selected as the best primer pair for diagnostics.

PCR screening of transgenic plants:

2 forward and 2 reverse oligonucleotides (4 combinations) designed against the Tet sequence were used to screen 5 plants containing the *Del/Ros1-N* transformation event.

Here run assays using SYBR green on DNA from genotype SCC90			
Also run dilution series			
1:10 of 0.1pg/uL		0.01	pg/uL = 0.05pg/assay
and the 1:100 assay from before (0.001pg/uL, 0.005pg/assay).			
	10x (2uM)	Take	
Water		480	Elga250311
Tet04541F2	2	10	TGGTCGCGTCCCAGTGA
Tet04541R1	2	10	CGGTCTTCTTCATCATGCAACTT
	10x (2uM)	Take	
Water		480	Elga250311
s1Lat52F2	2	10	
s1Lat52R2	2	10	
Assay each sample 2 x for each target			
Tet	24 reactions		
Tet	1rxn	24 rxns	Batch
water	3	6	Elga250311
2 X power sybr	10	20	1008236
10 x Primers	2	4	VC250311
DNA	5		PI030311
SlLat	20	reactions	
SlLat	1rxn	20 rxns	Batch
water	3	6	Elga250311
2 X power sybr	10	20	1008236
10 x Primers	2	4	VC250311
DNA	5		PI030311
Place	15	ul in wells	
Add	5	uL DNA	
		Plate Make	sarstedt 721984 202
		Plate Batch	S22768
		Seal Make	sarstedt 95.1994
		Seal Batch	2012-08-1A
PCR	Vesuvias		
Denaturation	95C	10 min	
40 cycles	95C	15s	
	60C	60sec	Read
Follow with Dissociation			

3 - PCR RESULTS

No amplification of the Tet^R gene was observed in transgenic and non-transgenic tomato plants.

Sample	Ct	PCR result
"+ve Control1:10"	33.95704	positive
"+ve Control1:10"	33.489117	positive
+ve Control1:100	33.131016	positive
+ve Control1:100	33.842983	positive
Tet_SCC79_10 non-T	Undetermined	negative
Tet_SCC79_10 non-T	Undetermined	negative
Tet_SCC79_9 non-T	Undetermined	negative
Tet_SCC79_9 non-T	Undetermined	negative
Tet_SCC90_1	39.003906	negative
Tet_SCC90_1	Undetermined	negative
Tet_SCC90_2	36.892727	negative
Tet_SCC90_2	Undetermined	negative
Tet_SCC90_3	Undetermined	negative
Tet_SCC90_3	Undetermined	negative
Tet_SCC90_4	38.05322	negative
Tet_SCC90_4	Undetermined	negative
Tet_SCC90_5	Undetermined	negative
Tet_SCC90_5	Undetermined	negative
Tet_Water	Undetermined	negative
Tet_Water	Undetermined	negative

4 – CONCLUSION

The Tet^R resistance gene could not be detected in plants containing the *Del/Ros1-N* transformation event by quantitative polymerase chain reaction (q-PCR) using a series of standard oligonucleotides designed against the Tet^R sequence.

Appendix 6: Bioinformatics analysis of ORFs

Yang Zhang & Cathie Martin
John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

Table of Content

- 1 - BIOINFORMATIC ANALYSIS OF INTERRUPTED ORFs
- 2 - BIOINFORMATIC ANALYSIS OF PUTATIVE NEW ORFs
- 3 - CONCLUSION

1 - BIOINFORMATIC ANALYSIS OF INTERRUPTED ORFs

Homology of the genomic regions flanking the left (FLB) and right (FRB) borders of the T-DNA to tomato Open Reading Frames (ORFs) were searched by BLAST analysis against "Tomato gene models CDS and cDNA (ITAG release 3.20)" <https://solgenomics.net/tools/blast/>

The purpose of this analysis is to identify putative ORFs that may have been created due to the insertion of the T-DNA into the genome. Typically, putative ORFs of greater than 30 AA that were created by the insertion are analyzed. These ORFs should be searched for homology to known toxins and allergens. The description in this section of the native tomato genes in the area of the insertion belong in the section on the molecular characterization of the insertion.

FLB: 184 nt Sequence validated as junction sequences by PCR analysis, and confirmed by genomic resequencing

```
ATTTAGTTGTGCCCCGTAGCAAATTGTTTGTCAATCATCTCTCTCCCTCAAACCTCTCGCTCT  
ATCTCTCGTTTCGCTTCCTCTCGCTTTTATACAAACACAAGTGTATAAAATTTGTTTTTATTATA  
TAAAGCGAGAGAAAATTGTATAAATACATATATTTTTGTTTTCTCTGCCAGATCTCG
```

Search against Database: Tomato gene models CDS (ITAG release 3.20) = no hit

Overview graph: Not shown because no hits were found.

Coverage Graph: Not shown, because no hits were found.

Search parameters:

The screenshot shows the BLAST search interface with the following parameters:

- BLAST** button
- Advanced Options** (collapsed)
- Expect (e-value) Threshold:** 1e-10
- Max. hits to show:** 20
- Substitution Matrix:** BLOSUM62 (default)
- Filters:** ☒ Filter Low complexity
- Output Options:** Basic, Bioperl, Coverage, Overview, Table

Additional search on a longer region

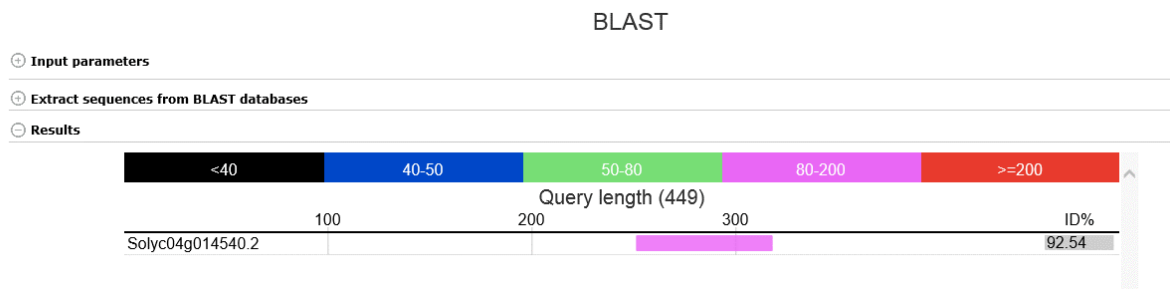
FLB: 449 nt

TATTTAGTTGTGCCCCGTAGCAAATTGTTTGTCAATCATCTCTCTCCCTCAAACCTCTCGCTCTATCT
 CTCGTTTCGCTTCCTCTCGCTTTTATACAAACACAAGTGTATAAAATTTGTTTTTATTATATAAAGCGA
 GAGAAAATTGTATAAATACATATATTTTTGTTTTCCTCTGCCAGATCTCGCTGCCACTCTCCCTAAT
 CTCGTTGCCACCCTCGCCTTTCTCGCTTATACAAACAGAAGCGAAATGTATAAATTGTGTTTCTG
 TTTGTATAAAGTAAGAGAAAATTATATACACATGCAAATACATATATTTTCGTCCTATACACTTAT
 AGTTATACAATAAAAAATACTCCCCTGCCAGTTTCTTTTGCCTTTCTCTCTTTCTAATTTTATACAAAT
 TCAAATTGTATATAATTTCTCCCTTTCTCATTTTATACAATTAA

Untitled_sequence vs Tomato Genome CDS (ITAG release 3.20)

SubjectId	id %	Al n	eval ue	Sc ore	Description
Solyc04g014540.2	92.54	62/67	1e-18	93.7	Zinc finger protein, putative (AHRD V3.3 *** B9T5L7_RICCO) Length = 1095

Overview graph:



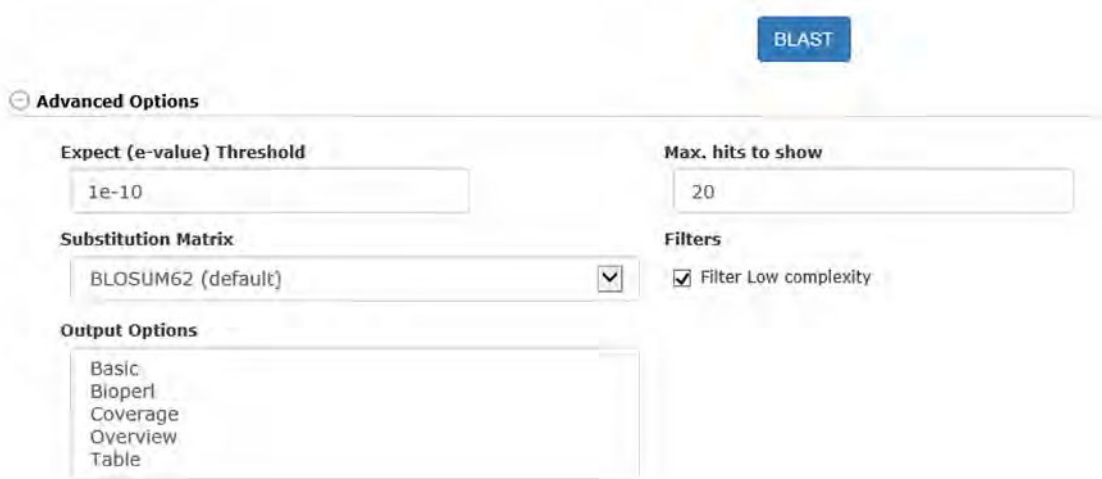
>Solyc04g014540.2 Zinc finger protein, putative (AHRD V3.3 *** B9T5L7_RICCO)
 Length = 1095

Score = 93.7 bits (47), Expect = 1e-18
 Identities = 62/67 (92%)
 Strand = Plus / Plus

Query: 251 gtataaattgtgtttctgtttgtataaagtaagagaaaattatatatacacatgcaaata 310
 |||||
 Sbjct: 764 gtataaattgtgtttctgtttatataaagcgagagaaaattatatatacacatgcaaata 823

Query: 311 catatat 317
 |||||
 Sbjct: 824 tatatat 830

Search parameters:



BLAST

Advanced Options

Expect (e-value) Threshold: 1e-10

Max. hits to show: 20

Substitution Matrix: BLOSUM62 (default)

Filters: ☒ Filter Low complexity

Output Options: Basic, Bioperl, Coverage, Overview, Table

Search result: 251 bp upstream of the T-DNA junction sequence a 67 bp motif of a putative Zinc finger protein is present. No ORF was interrupted.

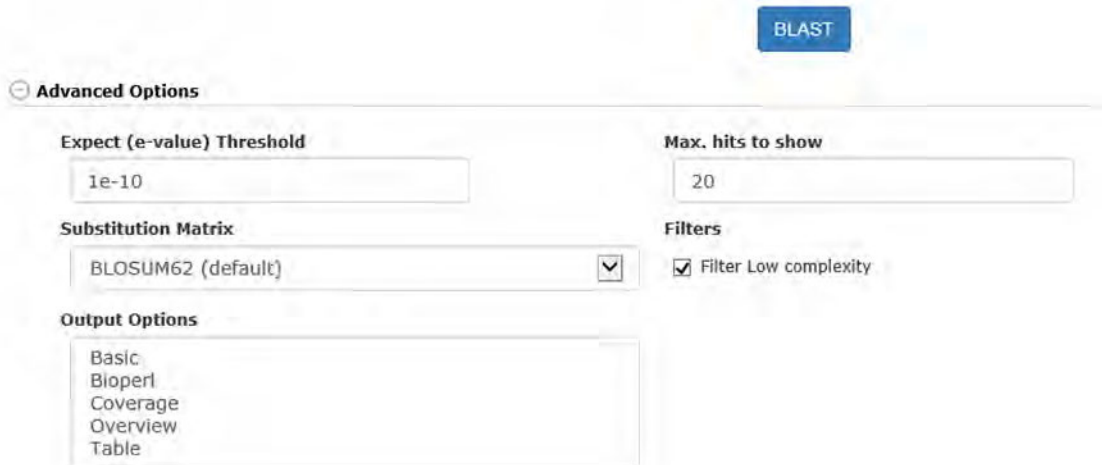
FRB: 332 nt Sequence validated as junction sequences by PCR analysis and by genomic resequencing

GGCCCAATACATACATCACCTCTTTAATTTGCTCATACATTTTGTCTTGATGCTCCGATTTAATCTT
 GTTCTATTTTAATTCTTCAACCTCATTTCTCATGCGTCATTTAAACATAAAATGTTGATTAAAGATTC
 ATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAAATTTAATAATAATTTGACCCTTTTACGGTTA
 ACAACCTTCTTCCCTATTTATTTCTTCTTCTTCTTATTTATTTTGTTCATCGACTTACCTACTTATTCA
 TTTATTTTCTTAATCTATTTTCTAAAAATTTAATAATAATTTGACCCTTTTACGTAT

Overview graph: Not shown because no hits were found.

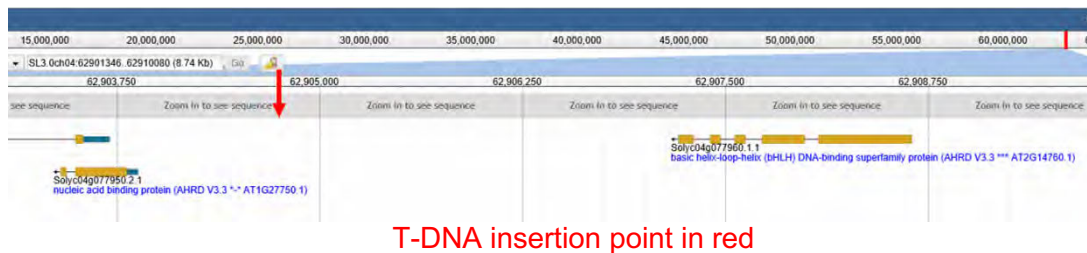
Coverage Graph: Not shown, because no hits were found.

Search parameters:



The image shows the BLAST search parameters interface. At the top right is a blue 'BLAST' button. Below it is a section titled 'Advanced Options' with a minus icon. This section contains several input fields and checkboxes. 'Expect (e-value) Threshold' is set to '1e-10'. 'Max. hits to show' is set to '20'. 'Substitution Matrix' is set to 'BLOSUM62 (default)' with a dropdown arrow. 'Filters' has a checked box for 'Filter Low complexity'. 'Output Options' is a list box containing 'Basic', 'Bioperl', 'Coverage', 'Overview', and 'Table'.

Genome browser visualization of insert (ITAG 3.2 gene model):



The T-DNA insertion on chromosome 4 (SL2.50ch04) between position 62904677..62904771 (94 bp deletion of the genome sequence) did not interrupt any ORF.

The insertion occurred 797 bp upstream of the start of transcription of Solyc04g077950.2.1 which is a prediction based upon homology with a fragment of a rice protein Os01g0765200. There is limited annotation for this sequence which could be a pseudogene.

The insertion occurred 971 bp upstream of the start of transcription of Solyc04g077940.3.1 but 1,134 bp upstream of the start of translation. These distances are relatively large for plant genes and the insertion was not expected to influence the expression of this predicted gene Solyc04g077940.3.1 is a Flowering time control protein FPA (AHRD V1 *- FPA_ARATH); containing the Interpro domain(s) IPR012921 Spen paralogue but no alteration of flowering was observed in purple tomato plants in different genetic backgrounds compared to controls across multiple generations

The insertion occurred 2439 bp downstream of Solyc04g077960.1.1 an insertion which is not expected to influence the expression of this predicted gene.

Conclusion: No ORF was interrupted by the presence of T-DNA insert on chromosome 4.

2 - BIOINFORMATICS ANALYSIS OF PUTATIVE NEW ORFs

Bioinformatics analysis of putative new ORFs was investigated at the LB and RB junctions of the T-DNA insert on chromosome 4 (SL2.50ch04) which occurred between position 62904677..62904771 following a 94 bp deletion of the genome sequence.

2a – LB junction:

Potential new ORFs were searched for by translating the genomic:T-DNA junction sequence in all possible reading frames (6) and detecting initiator of translation codons (M) . No new functional AA sequences of greater than 30 amino acids were detected spanning the FLB: T-DNA junction sequence on chromosome 4.

FLB: T-DNA junction point

AA sequence spanning the junction sequence from stop codon (5') to stop codon (3')

UPPERCASE= FLANKING GENOMIC SEQUENCE (FLB)

Lowercase = T-DNA sequence

```
TACTAACTAATACCAACGTATATTAATTTTCTCAAATTTTAACCAAATACTACACAATTTTAATTTTTTCTTTTCCGTGTTAA
AATTAGAGTCTCAATTAATTTGAATCGCACATTGCAGGTTAATTCAGTGGTGATACTCAAACCTCTGCACATTGCAAGGTTATTCG
AAAGTAACAAACAGAAAGACCTCTAGTTAAGGGTGGAAACGGTATCACCCTACACACAACCTTATGTTGATCACAAACTTATATTTAG
GGCAACTTTTCATACATAGTAAACATAAAATTCATATTTGTATGCTATAACAGAGTTTACATAATTGCTCTCCATAACAAATATATA
TATGTATAATTCGCTATACATATACAATTGAAGCGAATTGTATAAAACAAATTGTATAAAACGAAAAAGAGAAAGAGACTTGGGCA
GAAAATTGTATAAAACAAAGTGTATAATTATAAGTGTATATAACGATTATATACAATTCAAATTGTATAAAATGAGAAAGAGAGAA
AGACAAAAGAGACCTTGGACAGTGAATATACATTTGAGTTAAATTGTATAAAATGAGAAAGGGAGAAATTATATACAATTTGAATTT
GTATAAAATTAGAAAGAGAGAAAGGCAAAAGAACTGGCAGGGGAGTATTTTATTGTATAACTATAAGTGTATAGGACGAAAATA
TATGTATTTGCATGTGTATATATAATTTCTCTTACTTTATACAAACAGAAACACAATTTATACATTTCTGCTTCTGTTGTATAAG
CGAGAAAGGCGAGGGTGGCGAACGAGATTAGGGAGAGTGGCAGCGAGATCTGGCAGAGGAAAACAAAATATATGTATTTATACAA
TTTTCTCTCGCTTTATATAATAAAACAAATTTTATACACTTGTGTTTGTATAAAAGTGAGAGGAAGCGAACGAGAGATAGAGCGA
GAGTTTGAGGGAGAGAGATGATTGACAAACAATTTGCTACGGGGCACAATAAAAGTGGAGAAAAAGAGAGTAATTACCAATTT
TTTTCAATTCAAAAATGtagatgtccgcagcggtattataaaaatgaaagtacattttgataaaaacgacaaattacgatccgctcgt
atztataggcgaaagcaataaacaattattctaatccggaatccttatttcgacgtgtctacattcacgtccaaatgggggctt
agatgagaaacttcacgatcgatgccttgatttcgccattcccagatacccatttcattcttcagattggctctgagattatgcgaaa
atatacactcatatacataaatactgacagtttgagctaccaattcagtgtagccattaccttacataattcactcaaagtctag
gcagtcgtgcaactcggcggtcaatttgcggccactatacgatagttgcgcaaattttcaaagtcctggcctaacatcacacctct
gtcggcgggggtcccatTTgtgataaatccacccatcgatgatcatgagcgggagaattaagggagtcacgttatgacccccgc
gatgacgcgggacaagccgtt
```

Translate Tool - Results of translation

ExPASy Bioinformatics resource Portal by Swiss Institute of Bioinformatics

<https://web.expasy.org/translate/>

5'3' Frame 1

```
tactaactaataccaacgtatattaaatTTTCTCAAATTTTAACCAAATACTACACAATT
Y - L I P T Y I K F S Q I L T K Y Y T I
TTAATTTTTTTTCTTTTCCGTGTTAAATTAGAGTCTCAATTAATTTGAATCGCACATT
L I F F L F R V K I R V S I K F E S H I
gcaggTTAATTCAGTGGTGATACTCAAACCTCTGCACATTGCAAGGTTATTCGAAAGTAAC
A G - F S G D T Q T L H I A R L F E S N
aaacagaagacctctagTTAAGGGTGAACGGTATCACCCTACACACAACCTTATGTTGA
K Q K T S S - G W N G I T T T H N L C -
tcacaaacttatatttagggcaactTTTCATACATAGTAAACATAAAATTCATATTTGTAT
S Q T Y I - G N F H T - - T - N S Y L Y
gctataacagagTTTACATAATTGCTCTCCATAACAAATATATATATGTATAATTCGCTA
A I T E F T - L L S I T N I Y M Y N S L
tacatatacaattgaagcgaattgtataaaacaaattgtataaaacgaaaaagagaaaga
Y I Y N - S E L Y K T N C I K R K R E R
gacttgggagcgaattgtataaaacaaagtgtataattataagtgatatataacgattat
D L G R K L Y K T K C I I I S V Y N D Y
atacaattcaaattgtataaaatgagaaagagagaaagacaaagagacctggacagtga
I Q F K L Y K M R K R E R Q K R P G Q -
```


atatacattttgagttaaattgtataaaatgagaaagggagaaattatatacaattttgaat
 I Y I - V K L Y K M R K G E I I Y N L N
 ttgtataaaattagaaagagagaaaggcaaaagaaactggcaggggagttttttattgt
 L Y K I R K R E R Q K K L A G E Y F Y C
 ataactataagtgtataggacgaaaatatatgtattttgcatgtgtatatataattttctc
 I T I S V - D E N I C I C M C I Y N F L
 ttactttatacaaacagaaacacaattttatacatttcgcttctgtttgtataagcgagaa
 L L Y T N R N T I Y T F R F C L Y K R E
 aggcgaggggtggcgaacgagattagggagagtggcagcgagatctggcagaggaaaacaa
 R R G W R T R L G R V A A R S G R G K Q
 aaatatatgtattttatacaatttttctctcgttttatataataaaaaacaaatttttatacac
 K Y M Y L Y N F L S L Y I I K T N F I H
 ttgtgtttgtataaaagtgaaggaagcgaacgagagatagagcgagagtttgagggaga
 L C L Y K S E R K R T R D R A R V - G R
 gagatgattgacaaacaatttgctacgggggcacaactaaatgatggagaaaaagaaagagt
 E M I D K Q F A T G H N - I W R K R K S
 aattaccaattttttttcaattcaaaaatgtagatgtccgcagcgttattataaaatgaa
 N Y Q F F F N S K M - M S A A L L - N E
 agtacatttttgataaaacgacaaattacgatccgtcgtattttataggcgaaagcaataaa
 S T F - - N D K L R S V V F I G E S N K
 caaattatttctaattcggaaatctttattttcgacgtgtctacattcacgtccaaatgggg
 Q I I L I R K S L F R R V Y I H V Q M G
 gcttagatgagaaacttcacgatcgatgccttgatttcgccattcccagatacccatcttc
 A - M R N F T I D A L I S P F P D T H F
 atcttcagatttggtctgagattatgcgaaaatatacactcatatacataaataactgacag
 I F R L V - D Y A K I Y T H I H K Y - Q
 tttgagctaccaattcagtgtagccattaccttacataattcactcaaagtctaggcag
 F E L P I Q C S P L P Y I I H S N A R Q
 tctgtcaactcggcgtcaattttgtcggccactatacgatagttgcgcaaattttcaaagt
 S V N S A S I C R P L Y D S C A N F Q S
 cctggcctaacatcacacctctgtcggcggcgggtcccattttgtgataaatccacccatc
 P G L T C R L R R R V P F V I N P P I
 ggatgatcatgagcggagaattaagggagtcacgttatgacccccgcgatgacgcggga
 G - S - A E N - G S H V M T P A D D A G
 caagccggtt
 Q A V

5'3' Frame 2

tactaactaataccaacgtatatattaatttttctcaaatttttaaccaaatactacacaattt
 T N - Y Q R I L N F L K F - P N T T Q F
 taatttttttccctttccgtgttaaaattagagtgctcaattaaatttgatcgacattg
 - F F S F S V L K L E S Q L N L N R T L
 cagggttaattcagtggtgataactcaaactctgcacattgcaagggttatttcgaaagtaaca
 Q V N S V V I L K L C T L Q G Y S K V T
 aacagaagacctctagttaagggtggaacgggtatcaccactacacacaacttatgtttgat
 N R R P L V K G G T V S P L H T T Y V D
 cacaaacttatatttagggcaactttcatatagtaaacataaaattcatattttgtatg
 H K L I F R A T F I H S K H K I H I C M
 ctataacagagttttacataattgctctccataacaaatatatatatgtataattcgctat
 L - Q S L H N C S P - Q I Y I C I I R Y
 acatatacaattgaagcgaattgtataaaacaaattgtataaaacgaaaaagagaaagag
 T Y T I E A N C I K Q I V - N E K E K E
 acttgggcagaaaattgtataaaacaaagtgtataattataagtgtatataacgattata
 T W A E N C I K Q S V - L - V Y I T I I
 tacaattcaaattgtataaaatgagaaagagagaaagacaaaagagacctggacagtgaa
 Y N S N C I K - E R E K D K R D L D S E

tatacatttgagttaaattgtataaaatgagaaagggagaaattatatacaatttgaatt
Y T F E L N C I K - E R E K L Y T I - I
tgtataaaattagaaagagagaaaggcaaaagaaactggcaggggagtagtttttattgtat
C I K L E R E K G K R N W Q G S I F I V
taactataagtgtataggacgaaaatatatgtatttgcattgtgtatatataattttctct
- L - V Y R T K I Y V F A C V Y I I F S
tactttatacaaacagaaacacaattttatacatttgcgttctgtttgtataagcgagaaa
Y F I Q T E T Q F I H F A S V C I S E K
ggcgagggtggcgaacgagattagggagagtgaggcagcgagatctggcagaggaaaacaaa
G E G G E R D - G E W Q R D L A E E N K
aatatatgtattttatacaattttctctcgtttatataataaaaaacaaattttatacact
N I C I Y T I F S R F I - - K Q I L Y T
tgtgtttgtataaaagttagaggaagcgaacgagagatagagcgagagtttgaggagag
C V C I K V R G S E R E I E R E F E G E
agatgattgacaaacaatttgcctacggggcacaactaaatgatggagaaaaagaaagagta
R - L T N N L L R G T T K Y G E K E R V
attaccaattttttttcaattcaaaaatgtagatgtccgcagcggttattataaaatgaaa
I T N F F S I Q K C R C P Q R Y Y K M K
gtacattttgataaaacgacaaattacgatccgtcgtattttataggcgaagcaataaac
V H F D K T T N Y D P S Y L - A K A I N
aaattatttctaattcggaaatctttatttgcgacgtgtctacattcacgtccaaatggggg
K L F - F G N L Y F D V S T F T S K W G
cttagatgagaaacttcacgatcgatgccttgatttgcgcattcccagatacccatttca
L R - E T S R S M P - F R H S Q I P I S
tcttcagattgggtctgagattatgcgaaaatatacactcatatacataaatactgacagt
S S D W S E I M R K Y T L I Y I N T D S
ttgagctaccaattcagtgtagccattaccttacataattcactcaaattgctaggcagt
L S Y Q F S V A H Y L T - F T Q M L G S
ctgtcaactcggcgtcaatttgcggccactatacagatagttgcgcaaattttcaaagtc
L S T R R Q F V G H Y T I V A Q I F K V
ctggcctaacatcacacctctgtcggcggggtcccatttgtgataaatccaccatcg
L A - H H T S V G G G S H L - - I H P S
gatgatcatgagcggagaattaagggagtcacggttatgacccccgcgatgacgcgggac
D D H E R R I K G V T L - P P P M T R D
aagcgtt
K P

5'3' Frame 3

tactaactaataccaacgtatatattaatttttctcaaatttttaaccaaatactacacaatttt
L T N T N V Y - I F S N F N Q I L H N F
aatttttttcccttttccgtgtttaaattagagttctcaatttaaatttgaatcgacattgc
N F F P F P C - N - S L N - I - I A H C
agggttaattcagtggtgatactcaaactctgcacattgcaagggttattcgaaagtaacaa
R L I Q W - Y S N S A H C K V I R K - Q
acagaagacctctagtttaaggggtggaacggtatcaccactacacacaacttatgttgatc
T E D L - L R V E R Y H H Y T Q L M L I
acaaacttatattttagggcaactttcatacatagtaaacataaaaattcatatttgtatgc
T N L Y L G Q L S Y I V N I K F I F V C
tataacagagtttacataattgtctctccataacaaatatatatatgtataattcgtata
Y N R V Y I I A L H N K Y I Y V - F A I
catatacaattgaagcgaattgtataaaacaaattgtataaaacgaaaagagaaagaga
H I Q L K R I V - N K L Y K T K K R K R
cttgggcagaaaattgtataaaacaaagtgtataattataagtgtatataacgattatat
L G Q K I V - N K V Y N Y K C I - R L Y
acaattcaaattgtataaaatgagaaagagagaaagacaaaagagacctggacagtgat

T I Q I V - N E K E R K T K E T W T V N
 atacatttgagtttaaattgtataaaatgagaaagggagaaattatatacaatttgaattt
 I H L S - I V - N E K G R N Y I Q F E F
 gtataaaattagaaagagagaaagggcaaaagaaactggcaggggagttttttattgtat
 V - N - K E R K A K E T G R G V F L L Y
 aactataagtgtataggacgaaaatatatgtatttgcattgtgtatatataattttctctt
 N Y K C I G R K Y M Y L H V Y I - F S L
 actttatacaaacagaaacacaattttatacatttctgcttctgtttgtataagcgagaaag
 T L Y K Q K H N L Y I S L L F V - A R K
 gcgaggggtggcgaacgagattagggagagtgccagcgagatctggcagaggaaaaacaaa
 A R V A N E I R E S G S E I W Q R K T K
 atatatgtattttatacaatttttctctcgtttatataataaaaaacaaattttatacattt
 I Y V F I Q F S L A L Y N K N K F Y T L
 gtgtttgtataaaagtggaggaagcgaacgagagatagagcgagagtttgagggagaga
 V F V - K - E E A N E R - S E S L R E R
 gatgattgacaaacaattttgctacggggcacaactaaatattggagaaaaagaaagagtaa
 D D - Q T I C Y G A Q L N M E K K K E -
 ttaccaattttttttcaattcaaaaatgtagatgtccgcagcgttattataaaatgaaag
 L P I F F Q F K N V D V R S V I I K - K
 tacattttgataaaaacgacaaattacgatccgtcgtattttataggcgaaagcaataaaca
 Y I L I K R Q I T I R R I Y R R K Q - T
 aattattctaattcggaaatctttatttgcacgtgtctacattcacgtccaaatgggggc
 N Y S N S E I F I S T C L H S R P N G G
 ttagatgagaaacttcacgatcgatgccttgatttcgccattcccagatacccatattcat
 L D E K L H D R C L D F A I P R Y P F H
 cttcagattgggtctgagattatgcgaaaatatacactcatatacataaataactgacagtt
 L Q I G L R L C E N I H S Y T - I L T V
 tgagctaccaattcagtgtagccattaccttacataattcactcaaagtctagggcagtc
 - A T N S V - P I T L H N S L K C - A V
 tgtcaactcggcgtcaatttgcgcggccactatacgatagttgcgcaaattttcaaagtc
 C Q L G V C N C T S A T I R - L R K F S K S
 tggcctaacatcacacctctgtcggcgggggtccatttgtgataaatccaccatcgg
 W P N I T P L S A A G P I C D K S T H R
 atgatcatgagcggagaaattaagggagtcacgttatgacccccgcgatgacgcgggaca
 M I M S G E L R E S R Y D P R R - R G T
 agccggtt
 S R

3'5' Frame 1

aacggcttgtcccgcgtcatcggcggggggtcataacgtgactcccttaattctccgctca
 N G L S R V I G G G H N V T P L I L R S
 tgatcatccgatgggtggatttatcacaatgggacccgcgcgcgacagaggtgtgatgt
 - S S D G W I Y H K W D P P P T E V - C
 taggccaggactttgaaaatttgcgcaactatcgtatagtggccgacaaattgacgccga
 - A R T L K I C A T I V - W P T N - R R
 gttgacagactgcctagcattttgagtgaattatgtaaggtaatgggctacactgaattgg
 V D R L P S I - V N Y V R - W A T L N W
 tagctcaaactgtcagttatttattgtatatgagtgtatattttcgcataatctcagaccaa
 - L K L S V F M Y M S V Y F R I I S D Q
 tctgaagatgaaatgggtatctgggaatggcgaaatcaaggcatcgatcgtgaagtttct
 S E D E M G I W E W R N Q G I D R E V S
 catctaagccccatttggacgtgaatgtagacacgtcgaaataaagatttccgaattag
 H L S P H L D V N V D T S K - R F P N -
 aataatttgtttattgttttcgcctataaatacgcgagatcgtaatttgtcgtttttatca
 N N L F I A F A Y K Y D G S - F V V L S
 aaatgtacttttattttataataacgctgcggacatctacatttttgaattgaaaaaaaa

K C T F I L - - R C G H L H F - I E K K
 ttggtaattactctttcttttctccatatttagttgtgccccgtagcaaattgtttgtc
 L V I T L S F S P Y L V V P R S K L F V
 aatcatctctctccctcaaactctcgctctatctctcgttcgcttccctctcacttttata
 N H L S P S N S R S I S R S L P L T F I
 caaacacaagtgtataaaattgtttttattatataaaagcgagagaaaattgtataaata
 Q T Q V Y K I C F Y Y I K R E K I V - I
 catatatttttgttttccctctgccagatctcgctgccactctccctaattctcgttcgcca
 H I F L F S S A R S R C H S P - S R S P
 ccctcgcttttctcgcttatacaaacagaagcgaaatgtataaattgtgtttctgtttgt
 P S P F S L I Q T E A K C I N C V S V C
 ataaagtaagagaaaattatatatacacatgcaaatacatatattttctgtcctatacact
 I K - E K I I Y T H A N T Y I F V L Y T
 tatagttatacaataaaaaactccccctgccagtttcttttgcctttctctctttcta
 Y S Y T I K I L P C Q F L L P F S L S N
 tttatacaaaattcaaattgtatataatttctcccttttctcattttatacaatttaactca
 F I Q I Q I V Y N F S L S H F I Q F N S
 aatgtatatattcactgtccaggtctcttttgtcttttctctcttttctcattttatacaattt
 N V Y S L S R S L L S F S L S H F I Q F
 gaattgtatataatcgttatatacacttataattatacactttgttttatacaattttct
 E L Y I I V I Y T Y N Y T L C F I Q F S
 gcccaagtctcttttctcttttctgttttatacaatttggttttatacaattcgcttcaatt
 A Q V S F S F S F Y T I C F I Q F A S I
 gtatatgtatagcgaattatacatatatatatatttggtatggagagcaattatgtaaactc
 V Y V - R I I H I Y I C Y G E Q L C K L
 tgttatagcatatacaaatatgaattttatgtttactatgtatgaaagttgccctaaatata
 C Y S I Q I - I L C L L C M K V A L N I
 agtttgtgatcaacataagttgtgtgtagtggtgataccgttccacccttaactagaggt
 S L - S T - V V C S G D T V P P L T R G
 ctctgtttgttacttttgaataaccttgcaatgtgcagagtttgagtatcaccactgaa
 L L F V T F E - P C N V Q S L S I T T E
 ttaacctgcaatgtgcagttcaaattttaattgagactctaattttaacacggaaaaggaa
 L T C N V R F K F N - D S N F N T E K E
 aaaaattaaaattgtgtagtatatttggttaaaatttgagaaaatttaatatagttggtat
 K N - N C V V F G - N L R K F N I R W Y
 tagttagta
 - L V

3'5' Frame 2

aacggcttgtcccgcgatcatcgggggggtcataacgtgactcccttaattctccgctcat
 T A C P T A S S A G V I T - L P - F S A H
 gatcatccgatgggtggatttatcaciaaatgggacccgcccgcgacagaggtgtgatgtt
 D H P M G G F I T N G T R R R Q R C D V
 aggccaggactttgaaaatttgcgcaactatcgtatagtggccgacaaaattgacgccgag
 R P G L - K F A Q L S Y S G R Q I D A E
 ttgacagactgcctagcattttgagtgaattatgtaaggtaatgggctacactgaattggt
 L T D C L A F E - I M - G N G L H - I G
 agctcaaactgtcagtatatttatgtatatgagtgtatatatttctgcataatctcagaccaat
 S S N C Q Y L C I - V Y I F A - S Q T N
 ctgaagatgaaatgggtatctgggaatggcgaaatcaaggcatcgatcgtgaagtttctc
 L K M K W V S G N G E I K A S I V K F L
 atctaagcccccatgttgacgtgaatgtagacacgtcgaaataaagatttccgaattaga
 I - A P I W T - M - T R R N K D F R I R
 ataatttgtttattgttttgcctataaatacgcacggatcgtaatttgtcgttttatcaa
 I I C L L L S P I N T T D R N L S F Y Q
 aatgtacttttcattttataataacgctgcggacatctacatttttgaattgaaaaaaat

N V L S F Y N N A A D I Y I F E L K K N
 tggtaattactctttctttttctcca**T**tttagttgtgccccgtagcaaattgtttgtca
 W - L L F L F L H I - L C P V A N C L S
 atcatctctctccctcaaactctcgctctatctctcgttcgcttctctcactttttatac
 I I S L P Q T L A L S L V R F L S L L Y
 aaacacaagtgtataaaattgtttttattatataaagcgagagaaaattgtataaatac
 K H K C I K F V F I I - S E R K L Y K Y
 atatatattttgttttccctctgccagatctcgctgccactctccctaattctcgttcgccac
 I Y F C F P L P D L A A T L P N L V R H
 cctcgccctttctcgcttatacaaacagaagcgaaatgtataaattgtgtttctgtttgta
 P R L S R L Y K Q K R N V - I V F L F V
 taaagtaagagaaaattatatacacatgccaaatacatatattttcgctcctatacactt
 - S K R K L Y I H **M** Q I H I F S S Y T L
 atagttatacaataaaaaatactcccctgcagtttcttttgcctttctctcttttctaatt
 I V I Q - K Y S P A S F F C L S L F L I
 ttatacaaaattcaaattgtatataattttctccctttctcattttatacaattttaactcaa
 L Y K F K L Y I I S P F L I L Y N L T Q
 atgtatatttactgtccaggtctcttttgtcttttctctcttttctcattttatacaatttg
M Y I H C P G L F C L S L F L I L Y N L
 aattgtatataatcgttatatacacttataaattatacactttgttttatacaattttctg
 N C I - S L Y T L I I I H F V L Y N F L
 cccaagtctcttttctctttttcgttttatacaatttgttttatacaattcgcttcaattg
 P K S L S L F R F I Q F V L Y N S L Q L
 tatatgtatagcgaattatacatatatatatttggtatggagagcaattatgtaaactct
 Y **M** Y S E L Y I Y I F V **M** E S N Y V N S
 gttatagcatacaaatatgaattttatgtttactatgtatgaaagttgccctaaatataa
 V I A Y K Y E F Y V Y Y V - K L P - I -
 gtttgtgatcaacataagttgtgtgtagtggatgacggtccacccttaactagagggtc
 V C D Q H K L C V V V I P F H P - L E V
 ttctgtttgttactttcgaataaccttgcaatgtgcagagtttgagtatcaccactgaat
 F C L L S N N L A **M** C R V - V S P L N
 taacctgcaatgtgcgattcaaatttaattgagactctaatttttaacacggaaaaggaaa
 - P A **M** C D S N L I E T L I L T R K R K
 aaaattaaaattgtgtagtattttgggttaaaatttgagaaaatttaatacgttgggtatt
 K I K I V - Y L V K I - E N L I Y V G I
 agtttagta
 S -

3'5' Frame 3

aacggcttgtcccgcgtcatcgggggggtcataacgtgactcccttaattctccgctcatg
 R L V P R H R R G S - R D S L N S P L **M**
 atcatccgatgggtggatttatcaciaaatgggacccgcgcgacagaggtgtgatgta
 I I R W V D L S Q **M** G P A A D R G V **M** L
 ggccaggactttgaaaatttgcgcaactatcgatatagtggccgacaaattgacgcccagt
 G Q D F E N L R N Y R I V A D K L T P S
 tgacagactgcctagcattttgagtgaattatgtaaggtaatgggctacactgaattggta
 - Q T A - H L S E L C K V M G Y T E L V
 gctcaaactgtcagttattatgtatatgagtgtatatatttgcataatctcagaccaatc
 A Q T V S I Y V Y E C I F S H N L R P I
 tgaagatgaaatgggtatctgggaatggcgaaatcaaggcatcgatcgatgaagtttctca
 - R - N G Y L G **M** A K S R H R S - S F S
 tctaagccccattttggacgtgaatgtagacacgtcgaaataaagattttccgaattagaa
 S K P P F G R E C R H V E I K I S E L E
 taatttgtttattgttttgcctataaatacgcgagatcgtaatttgtcgttttatcaaa
 - F V Y C F R L - I R R I V I C R F I K
 atgtactttcattttataataacgctgcggacatctacatttttgaattgaaaaaaaatt

M Y F H F I I T L R T S T F L N - K K I
ggtaattactctttcttttctccatcttttagttgtgccccgtagcaaattgtttgtcaa
G N Y S F F F S I F S C A P - Q I V C Q
tcattctctctccctcaaactctcgctctatctctcgcttcgcttctctcacttttataca
S S L S L K L S L Y L S F A S S H F Y T
aacacaagtgtataaaatttggtttttattatataaagcgagagaaaattgtataaataca
N T S V - N L F L L Y K A R E N C I N T
tatatttttgttttctctgcccagatctcgctgccactctccctaattctcgttcgccacc
Y I F V F L C Q I S L P L S L I S F A T
ctcgccctttctcgcttatacaaacagaagcgaaatgtataaattgtgtttctgtttgtat
L A F L A Y T N R S E M Y K L C F C L Y
aaagtaagagaaaattatatatacacatgcaaatacatatattttcgctctatacactta
K V R E N Y I Y T C K Y I Y F R P I H L
tagttatacaataaaaaatactcccctgccagtttcttttgcccttctctctttctaattt
- L Y N K N T P L P V S F A F L S F - F
tatacaaaattcaaattgtatataattttctccctttctcattttatacaatttaactcaaa
Y T N S N C I - F L P F S F Y T I - L K
tgttatattcactgtccaggtctcttttctcttttctcttttctcattttatacaatttga
C I F T V Q V S F V F L S F S F Y T I -
attgtatataatcgttatatacacttataaattatacactttgttttatacaattttctgc
I V Y N R Y I H L - L Y T L F Y T I F C
ccaagtctcttttctctttttcggttttatacaatttggttttatacaattcgcttcaattgt
P S L F L F F V L Y N L F Y T I R F N C
atatgtatagcgaattatacatatatatatttggttatggagagcaattatgtaaactctg
I C I A N Y T Y I Y L L W R A I M - T L
ttatagcatatacaaatatgaattttatgtttactatgtatgaaagttgccctaaatataag
L - H T N M N F M F T M Y E S C P K Y K
tttgtgatcaacataagttgtgtgttagtggtgataccggtccacccttaactagaggtct
F V I N I S C V - W - Y R S T L N - R S
tctgtttgttacttttgaataaccttgcaatgtgcagagtttgagtatcaccactgaatt
S V C Y F R I T L Q C A E F E Y H H - I
aacctgcaatgtgcgattcaaattttaattgagactctaattttaacacggaaaaggaaaa
N L Q C A I Q I - L R L - F - H G K G K
aaattaaaattgtgtagtatttggttaaaatttgagaaaatttaatatagtttggtatta
K L K L C S I W L K F E K I - Y T L V L
gttagta
V S

Summary of AA sequences spanning the LB junction sequence:

5'3' Frame 1

IWRKRKSNYQFFFNSKM

5'3' Frame 2

LTNNLLRGTTKYGEKERVITNFFSIQKCRCPQRYYKMKVHFDKTTNYDPSY

5'3' Frame 3

QTICYGAQLNMEKKKE

3'5' Frame 1

3'5' Frame 2

3'5' Frame 3

None of the potential ORFs contain an initiating Methionine codon that could produce a peptide of >30 amino acids indicating that they are unlikely to be translated.

2b – RB junction:

Putative new ORFs were searched by translating the genomic:T-DNA junction sequence in all possible reading frames (6) and detecting initiator of translation (M) .
AA sequences spanning the genomic:T-DNA junction are highlighted in grey.
No new functional AA sequence could be detected spanning the T-DNA:FRB junction sequence on chromosome 4.

T-DNA : **FRB**: junction point

AA sequence spanning the junction sequence from stop codon (5') to stop codon (3')

UPPERCASE= FLANKING GENOMIC SEQUENCE (FRB)
Lowercase = T-DNA sequence

```
atgtacgtaagcgcttacggtttttggtggacccttgaggaaactggtagctggtgtgggcctgtggtctcaaga
tggatcattaatttccaccttcacctacgatggggggcatcgaccggtgagtaaatattgtacggctaagagcg
aatttggcctgtagacctcaattgcgagcctttctaatttcaaactattcggGCCCAATACATACATCACCTCTT
TAATTTGCTCATAACATTTTGTCTTGATGCTCCGATTTAATCTTGTCTATTTTAAATTCTTCAACCTCATTTCTC
ATGCGTCATTTAAACATAAAAATGTTGATTAAAGATTCATGAAATTGCTTACCTTTAACGTAAGTGAGAAAAAAT
TTAATAATAATATTTGACCCTTTTACGGTTAACAACCTTCTTCCCTATTTATTTCTCTTCTTCTTCTTATTTATT
TTGTCATCGACTTACCTACTTATTCAATTTATTTTCTTAATCTATTTTCTAAAAATTTAATAATAATATTTGAC
CCTTTTACGTAT
```

Translate Tool - Results of translation

ExPASy Bioinformatics resource Portal by Swiss Institute of Bioinformatics

<https://web.expasy.org/translate/>

5'3' Frame 1

```
atgtacgtaagcgcttacggtttttggtggacccttgaggaaactggtagctggtgtgggc
M Y V S A Y V F G G P L R K L V A V V G
ctgtggtctcaagatggatcattaatttccaccttcacctacgatggggggcatcgacc
L W S Q D G S L I S T F T Y D G G H R T
ggtgagtaaatattgtacggctaagagcgaatttggcctgtagacctcaattgcgagcttt
G E - Y C T A K S E F G L - T S I A S F
ctaatttcaaactattcggGCCCAATACATACATCACCTCTTTAATTTGCTCATAACATTT
L I S N Y S G P I H T S P L - F A H T F
tgtcttgatgctccgatttaattctgttctatTTTAAATTTTCAACCTCATTTCTCATGC
C L D A P I - S C S I L I L Q P H F S C
gtcattttaaacataaaaatgTTGATTAAAGATTCATGAAATTGCTTACCTTTAACGTAAGT
V I - T - N V D - R F M K L L T F N V S
gagaaaaaatttaataataatatttgacccttttacggTTAACAACCTTCTTCCCTATTT
E K K F N N N I - P F Y G - Q P S S L F
atttctctctcttcttatttattttgtcatcgacttacctacttattcatttattttct
I S S S S L F I L S S T Y L L I H L F S
taatctatTTTTTCTAAAAATTTAATAATAATATTTGACCCTTTTACGTAT
- S I F L K I - - - Y L T L L R
```

5'3' Frame 2

```
atgtacgtaagcgcttacggtttttggtggacccttgaggaaactggtagctggtgtgggc
C T - A L T F L V D P - G N W - L L W A
tgtggtctcaagatggatcattaatttccaccttcacctacgatggggggcatcgaccg
```


C G L K M D H - F P P S P T M G G I A P
 gtgagtaatatattgtacggcctaagagcgaatttggcctgtagacctcaattgcgagctttc
 V S N I V R L R A N L A C R P Q L R A F
 taatttcaaactatttcgcccccaatacacatcacctctttaatttgctcatacatttt
 - F Q T I R A Q Y I H H L F N L L I H F
 gtcttgatgctccgatttaattctgttctattttaattcttcaacctcatttctcatgcg
 V L M L R F N L V L F - F F N L I S H A
 tcattttaaacataaaatgttgattaaagattcatgaaattgcttacctttaacgtaagt
 S F K H K M L I K D S - N C L P L T - V
 agaaaaaatttaataataatatttgacccttttacggttaacaaccttcttccctatttta
 R K N L I I I F D P F T V N N L L P Y L
 ttctctcttcttcttattttattttgtcatcgacttacctacttattcatttattttctt
 F P L L P Y L F C H R L T Y L F I Y F L
 aatctatttttctaaaaatttaataataatatttgacccttttacgtat
 N L F F - K F N N N I - P F Y V

5'3' Frame 3

atgtacgtaagcgcttacgtttttgggtggacccttgaggaaactggtagctgttgtgggcct
 V R K R L R F W W T L E E T G S C C G P
 gtgggtctcaagatggatcattaatttccaccttcacctacgatggggggcatcgccagg
 V V S R W I I N F H L R W G A S H R
 tgagtaatatgtacggcctaagagcgaatttggcctgtagacctcaattgcgagctttct
 - V I L Y G - E R I W P V D L N C E L S
 aatttcaaactatttcgcccccaatacacatcacctctttaatttgctcatacattttg
 N F K L F G P N T Y I T S L I C S Y I L
 tcttgatgctccgatttaattctgttctattttaattcttcaacctcatttctcatgctg
 S - C S D L I L F Y F N S S T S F L M R
 cattttaaacataaaaatgttgattaaagattcatgaaattgcttacctttaacgtaagtga
 H L N I K C - L K I H E I A Y L - R K -
 gaaaaaatttaataataatatttgacccttttacggttaacaaccttcttccctatttat
 E K I - - - Y L T L L R L T T F F P I Y
 ttctcttcttcttattttattttgtcatcgacttacctacttattcatttattttctta
 F L F F L I Y F V I D L P T Y S F I F L
 atctatttttctaaaaatttaataataatatttgacccttttacgtat
 I Y F S K N L I I I F D P F T Y

3'5' Frame 1

atacgtaaaaggggtcaaataattattatttaaatttttagaaaaatagattaagaaaataaa
 I R K R V K Y Y Y - I F R K I D - E N K
 tgaataagtaggtaagtcgatgacaaaataaataaggaagaagaggaaataaatagggaa
 - I S R - V D D K I N K E E E E I N R E
 gaagggtgttaaccgtaaaaggggtcaaataattattatttaaatttttctcacttacgtta
 E G C - P - K G Q I L L L N F F S L T L
 aaggtaagcaatttcatgaatctttaatcaacattttatgttttaaatgacgcatgagaaa
 K V S N F M N L - S T F Y V - M T H E K
 tgagggtgaagaattaaaatagaacaagattaaatcggagcatcaagacaaaatgtatga
 - G - R I K I E Q D - I G A S R Q N V -
 gcaaattaaagaggtgatgtatttgggcgggaatagtttgaaattagaaagctcgca
 A N - R G G D V C I G P E - F E I R K L A
 attgaggtctacaggccaaattcgctcttagccgtacaatattactcacgggtgcatgac
 I E V Y R P N S L L A V Q Y Y S P V R C
 ccccatcgtaggtgaagggtggaattaatgatccatcttgagaccacaggccacaaca
 P P S - V K V E I N D P S - D H R P T T
 gctaccagtttctcaagggtccacccaaaaacgtaagcgcttacgtacat
 A T S F L K G P P K T - A L T Y

3'5' Frame 2

atacgtaaaagggtcaaatattattatttaaatttttagaaaaatagattaagaaaaataaat
Y V K G S N I I I K F L E K - I K K I N
gaataagtaggtaagtcgatgacaaaataaataaggaagaagaggaaataaatagggaag
E - V G K S M T K - I R K K R K - I G K
aagggttgtaaccgtaaaagggtcaaatattattatttaaattttttctcacttacgttaa
K V V N R K R V K Y Y Y - I F S H L R -
aggtaagcaatttcgatgaatctttaatcaacattttatgttttaaatgacgcgatgagaat
R - A I S - I F N Q H F M F K - R M R N
gagggttgaagaattaaaaatagaacaagattaaatcggagcatcaagacaaaatgtatgag
E V E E L K - N K I K S E H Q D K M Y E
caaattaaagagggtgatgtatgtattgggc ccgaatagtttgaaattagaaagctcgcaa
Q I K E V M Y V L G P N S L K L E S S Q
ttgaggtctacaggccaaattcgcgtcttagccgtacaatattactaccgggtgcgatgcc
L R S T G Q I R S - P Y N I T H R C C D A
ccccctcgtaggttgaagggtgaaattaatgatccattcttagaccacggcccaacag
P H R R - R W K L M I H L E T T G P Q Q
ctaccagtttctcaagggtccacaaaaaacgtaagcgcttacgtacat
L P V S S R V H Q K R K R L R T

3'5' Frame 3

atacgtaaaagggtcaaatattattatttaaatttttagaaaaatagattaagaaaaataaatg
T - K G Q I L L L N F - K N R L R K - M
aataagtaggtaagtcgatgacaaaataaataaggaagaagaggaaataaatagggaaga
N K - V S R - Q N K - G R R G N K - G R
aggttgttaaccgtaaaagggtcaaatattattatttaaattttttctcacttacgttaaa
R L L T V K G S N I I I K F F L T Y V K
ggtaagcaatttcatgaatctttaatcaacattttatgtttaatgacgcatgagaaatg
G K Q F H E S L I N I L C L N D A - E M
aggttgaagaattaaaaatagaacaagattaaatcggagcatcaagacaaaatgtatgagc
R L K N - N R T R L N R S I K T K C M S
aaattaaagggtgatgtatgtattgggcctgaatagtttgaaattagaaagctcgcaat
K L K R - C M Y W A R I V - N - K A R N
tgaggctacaggccaaattcgcctcttagccgtacaatattactcaccgggtgcgatgcc
- G L Q A K F A L S R T I L L T G A M P
cccactcgtaggtgaagggtgaaattaatgatccattctgagaccacaggcccaacagc
P I V G E G G N - - S I L R P Q A H N S
taccagtttcctcaagggtccacaaaaacgtaagcgcttacgtacat
Y Q F P Q G S T K N V S A Y V H

Summary of AA sequences spanning the RB junction sequence:

5'3' Frame 1

5'3' Frame 2

FQTIRAQYIHHLFNLLIHFVL**M**LRFNLVLF

5'3' Frame 3

3'5' Frame 1

3'5' Frame 2

NKIKSEHQDKMYEQIKEVMYVLGPNSLKLESSQLRSTGQIRS

3'5' Frame 3

A 32 amino acid peptide containing an initiating Methionine codon and spanning the RB junction sequence has been identified in the [3'5' Frame 2](#) and has been analyzed further.

32 amino acid from [3'5' Frame 2](#): **MYEQIKEVMYVLGPNSLKLESSQLRSTGQIRS**

2c - In silico assessment of the likely homology and allergenicity of the 32 AA peptide spanning the RB junction sequence

The frame 2 flanking the right border reading 3' to 5' created an ORF containing an initiation codon (ATG) that could produce a 32 amino acid peptide. Since this ORF initiates in the tomato genome and runs across the right border into the end of the T-DNA, sequences homologous to the tomato genomic region were searched amongst all available online RNA-seq data. No evidence of any transcripts carrying the tomato genome sequence of this putative ORF were found, and neither was this region of DNA annotated as a potential gene in the tomato genome sequence. Consequently, there is no evidence that this region encodes a polypeptide in tomato, and the sequence would need to have an active promoter and Kozak consensus sequence around the initiating ATG codon (which it lacks) to produce such a peptide which would read from the tomato genome into the right border region. No evidence for such a transcript was found. Furthermore this 32 AA peptide was analyzed for any putative homology to known allergens or existing protein.

Homology search to any known allergen:

Homology search to any known allergen was conducted on the allergenonline.org database (<http://allergenonline.org/>) using either the whole 32 AA sequence (Fasta35 & Fasta36 search methods) or 8 amino acids epitopes on 05-11-2019. Overall, no homology (0 hit) could be detected.

Latest Version	
Version #	19
Peer Reviewed Sequences	2129
taxonomic-protein groups	852
Released On	February 10, 2019

32 AA query sequence: **MYEQIKEVMYVLGPNSLKLESSQLRSTGQIRS**

- Full Fast35 homology search – NO SIGNIFICANT HITS FOUND:

Sequence Entry

Fasta Sequence

MYEQIKEVMY VLGPNLSKLE SSQLRSTGQI RS

Search Method (FASTA 35 or FASTA 36)

Full Fasta 35

☒ Open results in new window

Search Options

E-Value Cutoff

1

Max Alignments

20

☐ Show Z-Score
☒ Show Bit Score
☐ Show Histogram

SEARCH

AllergenOnline Database v19 (February 10, 2019)

NOTE Addition of Allergenicity* column on the Browse Database page with classification based on Group references was added on 10 May 2018. Please review the "allergenicity" of any matches you find here with the Browse page and look at Group References (gid) if you want to further evaluate relevance of alignments. <="" a=""> <="" a=""> <="" a=""> fasta35.exe -q -H -B -m 9i -w 80 -E 1 -d 20 C:\Windows\Temp\all4829.tmp fasta/version19.fasta

[User Query #1](#) >query

User Query #1

>query

MYEQIKEVMY VLGPNLSKLE SSQLRSTGQI RS

```
# fasta35.exe -q -H -B -m 9i -w 80 -E 1 -d 20 C:\Windows\Temp\all4829.tmp
fasta/version19.fasta
```

FASTA searches a protein or DNA sequence data bank

version 35.04 Jan. 15, 2009

Please cite:

W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

```

Query: C:\Windows\Temp\all4829.tmp
1>>>query - 32 aa
Library: fasta/version19.fasta 507441 residues in 2129 sequences

507441 residues in 2129 sequences
Statistics: Expectation_n fit: rho(ln(x))= 3.4786+/-0.00197; mu= 7.9004+/-
0.104
mean_var=19.2687+/- 4.762, 0's: 1 Z-trim: 1 B-trim: 171 in 1/42
Lambda= 0.292179
Algorithm: FASTA (3.5 Sept 2006) [optimized]
Parameters: BL50 matrix (15:-5) ktup: 1
join: 42, opt: 30, open/ext: -10/-2, width: 16
Scan time: 0.000
!! No sequences with E() < 1.000000

```

```

32 residues in 1 query sequences
507441 residues in 2129 library sequences
Scomplib [35.04]
start: Tue Nov 05 09:29:40 2019 done: Tue Nov 05 09:29:40 2019
Total Scan time: 0.000 Total Display time: 0.000

```

Function used was FASTA [version 35.04 Jan. 15, 2009]

- Full Fast36 homology search – NO SIGNIFICANT HITS FOUND:

Sequence Entry

Fasta Sequence

MYEQIKEVMY VLGPNLSKLE SSQLRSTGQI RS

Search Method (FASTA 35 or FASTA 36)

Full Fasta 36

☒ Open results in new window

Search Options

E-Value Cutoff

1

Max Alignments

20

☐ Show Z-Score
☒ Show Bit Score
☐ Show Histogram

SEARCH

AllergenOnline Database v19 (February 10, 2019)

NOTE Addition of Allergenicity* column on the Browse Database page with classification based on Group references was added on 10 May 2018. Please review the "allergenicity" of any matches you find here with the Browse page and look at Group References (gid) if you want to further evaluate relevance of alignments. <="" a=""><="" a=""><="" a=""> fasta35.exe -q -H -B -m 9i -w 80 -E 1 -d 20
C:\Windows\Temp\all4829.tmp fasta/version19.fasta

User Query #1 >query

```

>query
MYEQIKEVMY VLGPNSLKLE SSQLRSTGQI RS
# fasta36.exe -q -B -m 9i -w 80 -E 1 -d 20 C:\Windows\Temp\allF09F.tmp
version1936.fasta
FASTA searches a protein or DNA sequence data bank
  version 36.3.8g Oct, 2018
Please cite:
Pearson, W.R., and Lipman, D.J. (1988) Proc. Natl, Acad. Sci., 85:2444-2448

Query: C:\Windows\Temp\allF09F.tmp
1>>>query - 32 aa
Library: version1936.fasta
  507441 residues in 2129 sequences

Statistics: Expectation_n fit: rho(ln(x))= 3.9549+/-0.00338; mu= 8.0892+/-
0.171
  mean_var=16.8668+/- 4.790, 0's: 0 Z-trim(89.2): 2 B-trim: 102 in 1/35
  Lambda= 0.312290
  statistics sampled from 761 (761) to 761 sequences
Algorithm: FASTA (3.8 Nov 2011) [optimized]
Parameters: BL50 matrix (15:-5), open/ext: -10/-2
  ktup: 1, E-join: 1 (0.863), E-opt: 0.2 (0.357), width: 16
  Scan time: 0.000
!! No sequences with E() < 1
>>>///

32 residues in 1 query sequences
507441 residues in 2129 library sequences
  Scomplib [36.3.8g Oct, 2018]
  start: Tue Nov 05 09:30:23 2019 done: Tue Nov 05 09:30:23 2019
  Scan time: 0.000 Display time: 0.000

Function used was FASTA [36.3.8g Oct, 2018]

```

- 8mer exact match search – NO SIGNIFICANT HITS FOUND:

Sequence Entry

Fasta Sequence

MYEQIKEVMY VLGPNSLKLE SSQLRSTGQI RS

Search Method (FASTA 35 or FASTA 36)

8mer Exact Match

☒ Open results in new window

SEARCH

- **Note:** Sliding 8mer Searches Prior to September 12, 2007 may have identified matches of exactly 35% identity. However to be consistent with Codex 2005 the calculation of the cutoff value for a match has been changed to **Greater than 35%**.
- **Note 2:** The sequences in the FASTA searchable database might vary from the sequences described in the public literature, as this database is not updated on a daily basis.
- **Note 3:** We do not observe or log any protein sequences submitted through this website.
- **Note 4:** The E score cutoff for the sliding 8mer search was changed from 100 to 10 on 15 January, 2015 as explained on the "About AllergenOnline" page.

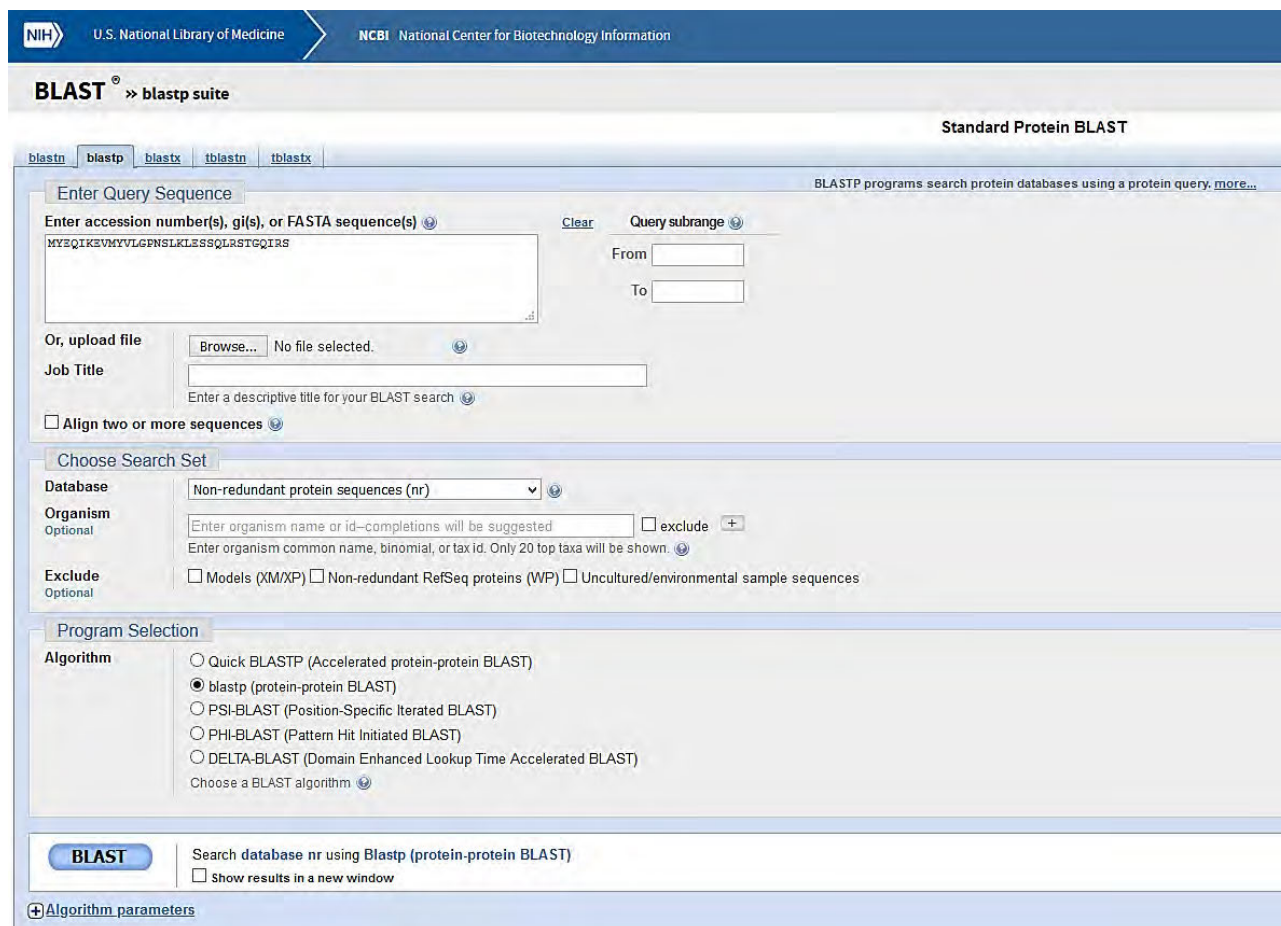
number of 8mer = 25

Number of Sequences with at least one 8mer match = 0

Homology search to any known protein:

Homology search to any known protein was conducted using BLASTp again the NCBI non-redundant protein database (<https://blast.ncbi.nlm.nih.gov/>) on 05-11-2019. No homology (0 hit) could be detected.

32 AA query sequence: **MYEQIKEVMYVLGPNLSKLESSQLRSTGQIRS**



The screenshot displays the NCBI BLASTp web interface. At the top, the NIH and NCBI logos are visible. The main heading is "BLAST® >> blastp suite". The "Standard Protein BLAST" section is active, showing tabs for blastn, blastp, blastx, tblastn, and tblastx. The "Enter Query Sequence" section contains a text box with the query sequence "MYEQIKEVMYVLGPNLSKLESSQLRSTGQIRS". Below this, there are options to "Or, upload file" (Browse...) and "Job Title" (Enter a descriptive title for your BLAST search). The "Choose Search Set" section shows the "Database" set to "Non-redundant protein sequences (nr)". The "Program Selection" section shows the "Algorithm" set to "blastp (protein-protein BLAST)". At the bottom, there is a "BLAST" button and a checkbox for "Show results in a new window".

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information

BLAST® >> blastp suite

Standard Protein BLAST

blastn blastp blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) Query subrange [From](#) [To](#)

MYEQIKEVMYVLGPNLSKLESSQLRSTGQIRS

Or, upload file [Browse...](#) No file selected.

Job Title [Enter a descriptive title for your BLAST search](#)

☐ Align two or more sequences [Align](#)

Choose Search Set

Database [Non-redundant protein sequences \(nr\)](#)

Organism [Optional](#) [Enter organism name or id—completions will be suggested](#) ☐ exclude [+](#)

Exclude [Optional](#) ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm

☐ Quick BLASTP (Accelerated protein-protein BLAST)

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

[Choose a BLAST algorithm](#)

BLAST Search database nr using Blastp (protein-protein BLAST)

☐ Show results in a new window

[Algorithm parameters](#)



U.S. National Library of Medicine
National Center for Biotechnology Information

Log In

BLAST® » blastp suite » results for RID-W36PH0NZ014

HomeRecent ResultsSaved StrategiesHelp

Edit Search
Save Search
Search Summary

How to read this report?
BLAST Help Videos
Back to Traditional Results Page

Job Title

Protein Sequence

RID

W36PH0NZ014

Search expires on 11-06 22:52 pm

Download All

Program

?
Citation

Database

nr

See details

Query ID

Icl|Query_29435

Description

None

Molecule type

amino acid

Query Length

32

Other reports

?

Filter Results

Organism

only top 20 will appear

☐
exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

No significant similarity found. For reasons why,click here



U.S. National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastp suite » results for RID-W376NH0U014

Edit Search
Save Search
Search Summary

Search Parameters

Program	blastp
Word size	6
Expect value	10
Hittlist size	100
Gapcosts	11,1
Matrix	BLOSUM62
Filter string	F
Genetic Code	1
Window Size	40
Threshold	21
Composition-based stats	2

Database

Posted date	Oct 30, 2019 12:42 AM
Number of letters	87,450,932,672
Number of sequences	242,386,682
Entrez query	None

Karlin-Altschul statistics

Lambda	0.311303	0.267
K	0.127315	0.041
H	0.325641	0.14
Alpha	0.7916	1.9
Alpha_v	4.96466	42.6028
Sigma		43.6362

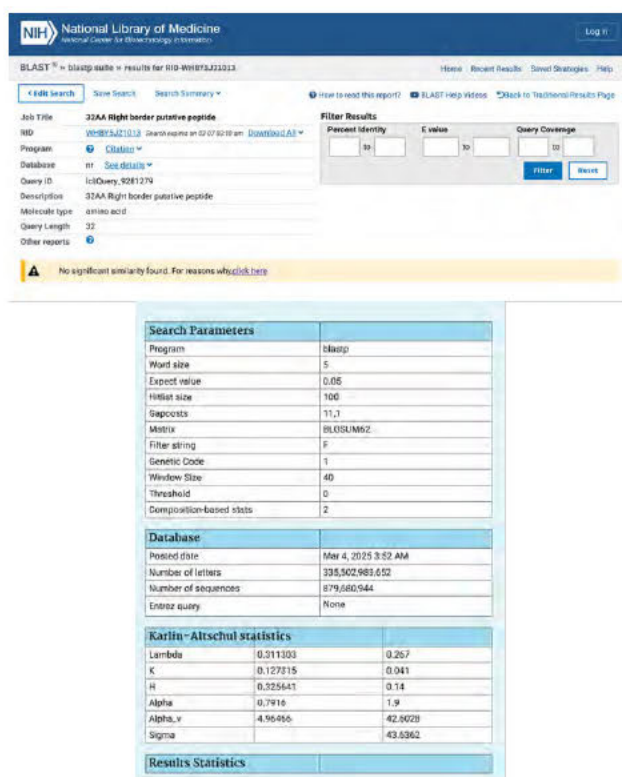
Results Statistics

It was concluded from these combined analyses that there is no evidence that the sequence encoding the 32 AA spanning the RB junction sequence is translated into a functional protein and this putative peptide has no predicted allergenic properties.

2d – Conclusion:

Putative new ORFs were searched by translating the genomic: T-DNA junction sequences in all possible reading frames (6) and detecting any initiator of translation codons (ATG encoding Methionine). Bioinformatic analysis showed that the T-DNA insertion in the *Del/Ros1-N* transformation event did not interrupt any existing open reading frame (ORFs). Frame 2 flanking the right border reading 3' to 5' created an ORF containing an initiation codon (ATG) that could produce a 32 amino acid peptide. However, this peptide has no homology to any known allergen or protein and there was no evidence (from tomato RNA-seq databases) that this sequence is transcribed in tomato. It was concluded that the T-DNA insertion on chromosome 4 did not create any new functional open reading frames (ORF).

To confirm that the potentially translated 32AA sequence spanning the right border from 3' - 5' in frame 2: MYEQIKEVMYVLGPNSLKLESSQLRSTGQIRS has no significant matches in the NCBI database, the BLAST search results were updated on March 4, 2025. The results appear below and confirm that the putative 32AA peptide has no similarity to any protein in the NCBA sequence database.



To confirm that the putative 32AA peptide had no allergenic potential, a search of the online allergenicity database was updated on March 5, 2025. The results appear below and confirm that the putative 32AA peptide has no matches to any known allergens or allergenic epitopes.

3/5/25, 3:19 PM allergenonline.org/search36.php

AllergenOnline Search Results

Note: As of August 2015 we have included gid: groupid in the fasta results that provides detailed information on the allergenicity references for the group, type of allergen, other sequences belonging to the same group and more.

%_id 1 = 100% identity, alen=alignment length

AllergenOnline Database v23 (January 30, 2025)

NOTE Addition of Allergenicity* column on the Browse Database page with classification based on Group references was added on 10 May 2018. Please review the "allergenicity" of any matches you find here with the Browse page and look at Group References (gid) if you want to further evaluate relevance of alignments.

fasta36.exe -q -B -m 9i -w 80 -E 1 -d 20 C:\Windows\Temp\allF6CE.tmp version2336.fasta

User Query #1 >32AA Right border putative peptide
 User Query #1
 >32AA Right border putative peptide
 MYEQIKEVMY VLGPNLSKLE SSQLRSTGQI RS
 # fasta36.exe -q -B -m 9i -w 80 -E 1 -d 20 C:\Windows\Temp\allF6CE.tmp version2336.fasta
 FASTA searches a protein or DNA sequence data bank
 version 36.3.8g Oct, 2018
 Please cite:
 W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448
 Query: C:\Windows\Temp\allF6CE.tmp
 1>>>32AA Right border putative peptide - 32 aa
 Library: version2336.fasta
 567469 residues in 2334 sequences
 Statistics: Expectation

n fit: $\rho(\ln(x)) = 4.0105 \pm 0.0032$; $\mu = 7.9368 \pm 0.163$
 mean

var=17.8057 $\pm$ 4.913, 0's: 0 Z-trim(89.8): 2 B-trim: 108 in 1/35
 Lambda= 0.303945
 statistics sampled from 826 (826) to 826 sequences
 Algorithm: FASTA (3.8 Nov 2011) [optimized]
 Parameters: BL50 matrix (15:-5), open/ext: -10/-2
 ktup: 1, E-join: 1 (0.855), E-opt: 0.2 (0.354), width: 16
 Scan time: 0.000
!! No sequences with E() < 1
 >>>///
 32 residues in 1 query sequences
 567469 residues in 2334 library sequences
 Scomplib [36.3.8g Oct, 2018]
 start: Wed Mar 05 14:19:06 2025 done: Wed Mar 05 14:19:06 2025
 Scan time: 0.000 Display time: 0.000
 Function used was FASTA [36.3.8g Oct, 2018]

www.allergenonline.org/search36.php 1/13/5/25, 3:20 PM 80mer Sliding Window Search Results

80mer Sliding Window Search Results

Database AllergenOnline Database v23 (January 30, 2025)

Input Query >32AA Right border putative peptide

MYEQIKEVMYVLGPNLSKLESSQLRSTGQIRS

Length 32

Number of 80 mers 1

Number of Sequences with hits 0

No Matches of Greater than 35% Identity Found

AllergenOnline Database v23 (January 30, 2025)

www.allergenonline.org/search8036.cgi 1/13/5/25, 3:20 PM Search Results

>32AA Right border putative peptide

number of 8mer = 25

Number of Sequences with at least one 8mer match = 0

www.allergenonline.org/shortmer.php 1/1

3 – CONCLUSION

Bioinformatic analyses show that T-DNA insertion on chromosome 4 did not create any new functional open reading frames (ORFs). One putative 32AA ORF, for which no evidence exists for expression, showed no similarity to any proteins, nor any similarity to any allergens.

Appendix 7: LC/MS analysis of proteins in purple tomato

Cathie Martin and Gerhard Saalbach, John Innes Centre Norwich, UK

Protein was prepared from fully ripe fruit of MoneyMaker F9 Purple Tomatoes according to **Faurobert et al., (2007)** and extracted proteins were analyzed on an Orbitrap LC/MS. More than 2,700 peptides were detected, but none corresponded to the Delila or Rosea1 proteins expressed from the pDEL.ROS1 T-DNA insertion chromosome 4 of tomato. No NPTII protein was detected either.

To establish the detection limit of this method, two standard proteins (BSA and Phosphorylase B) were spiked at different amounts into the purple tomato protein extract, and the result showed that there must be less than 29 fmoles of the Antirrhinum proteins in the total sample (from 2mL of extract). This corresponds to 0.75 ng Rosea and 2 ng of Delila. 2 mL of extract corresponds to 0.5gfw of tomato. The molecular weight of NPTII protein is 29.05kda. As a result, the following values for Limit of Detection are obtained:

DELILA: <4ng/gfw

ROSEA1: <1.5ng/gfw

NPTII: <1.7ng/gfw

The procedure used to estimate the detection limit was as follows:

A standard protein digest of BSA and Phosphorylase from Waters, claimed there was 1 nmole per vial. This was checked using 2 different protein assays, and both assays indicated that there was much less. Consequently, the protein samples were prepared based on the lab measurements and spiked in 100/10/1/0.1 fmoles per run. BSA/Phosphorylase was spiked into 360 ng tomato sample which was 1/290 of the total amount of protein from purple tomato extract provided. Even the lowest amount of 0.1 fmole standard proteins were detected nicely. The Del and Ros1 proteins must be lower than 0.1 fmole as they were not detected which means they must be lower than 29 fmoles in the total sample, as mentioned above.

The quantitative value calculated for the standard proteins using Scaffold was not at the very low end of the detection range. There were many tomato proteins with much lower levels, going down to 1/10 of the lowest values for the standard proteins. This means that the limit of detection was only 1/10 of what was measured, and the amounts of the Del and Ros1 proteins in the total sample were below 2.9 fmoles or 0.075 and 0.2 ng from 2 mL purple tomato extract.

A detailed description of the proteomic LC/MS method used follows. In addition, we will transmit as a separate document an Excel sheet that contains all of the data for the peptides detected.

Proteins were extracted from tomato fruits (three replicates) using a phenol extraction method according to Faurobert et al., (2007). The protein pellets were dissolved in ca. 100 µl of 40 mM sodium phosphate-buffer pH8.0 with 0.2% Rapigest detergent (Waters, Wilmslow, UK). Protein concentrations were estimated using the Direct Detect[®] Spectrometer (Merck). TEAB buffer (0.1 vol, Merck) was added, and cysteine residues were reduced in the presence of 5 mM dithiothreitol and alkylated in the presence of 12.5 mM iodoacetamide (Merck). Aliquots of 20 µg protein were then digested with 1 µg of trypsin (Promega) for 18 h at 37°C. The reaction was stopped by adding 0.1 volume of 20% TFA. One µl aliquots of this reaction corresponding the 360 ng protein were analysed by LC-MS/MS. For quantification, 2 standard proteins (MassPREP Phosphorylase B and BSA, Waters) were spiked in at 0.1, 1, 10, and 100 fmole per injection of 360 ng sample.

For LC-MS/MS analysis, an Orbitrap-Fusion™ mass spectrometer (Thermo Fisher, Hemel Hempstead, UK) equipped with an UltiMate™ 3000 RSLCnano System using an Acclaim PepMap C18 column (2 µm, 75 µm x 500mm, Thermo) was used. Aliquots of the tryptic digests were loaded and trapped using a pre-column which was then switched in-line to the analytical column for separation. The column was connected to a 10 µm SilicaTip™ nanospray emitter (New Objective, Woburn, MA, USA) for infusion into the mass spectrometer. For separation, the following gradient of solvents A (0.1% formic acid in water) and B (0.1% formic acid, 80% acetonitrile, 19.9% water) was used at a flow rate of 300 nl min⁻¹: solvent B at start: 4%; 0 to 3 min: linear ramp to 9% B; 3 to 81 min: ramp to 50% B; followed by a ramp to 90% B and re-equilibration at 4% B for 20 min.

Data dependent analysis was performed using parallel HCD/CID fragmentation with following parameters: MS1 in positive ion mode, orbitrap MS resolution = 120k, mass range (quadrupole) = 300-1900 m/z, AGC 2e5, max. inject time 50 ms; MS2 was triggered with an intensity threshold of 1.5e4, MS2 isolation window was 1.6 Da, HCD/CID NCE=30%, AGC target 100%, max. injection time 500 ms, dynamic exclusion 1 count, 30 s exclusion, exclusion mass window ±5 ppm, the option to inject ions for all available parallelizable time was selected. MS scans were saved in profile mode while MSMS scans were saved in centroid mode.

The mass spectrometry raw data were processed in MaxQuant 1.6 (maxquant.org) to generate peaklists. The peaklists were used for a database search using Mascot (Matrix Science, London, UK; version 2.4.1). Mascot was set up to search the *Solanum lycopersicum* database (uniprot.org, 2015, 35218 entries) together with a custom database containing the sequences of the two *Antirrhinum* enzymes (Q2VAZ7, Q38736), NPTII, and a database for common contaminants (maxquant.org, 245 entries).

Mascot was searched for strictly tryptic peptides with 2 missed cleavages allowed, with a fragment ion mass tolerance of 0.50 Da and a parent ion tolerance of 6.0 PPM. Carbamidomethyl of cysteine was specified in Mascot as a fixed modification. Deamidated asparagine and glutamine, oxidation of methionine and acetyl of the N-terminus were specified in Mascot as variable modifications. Scaffold (version Scaffold_5.1.2, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 99.0% probability and contained at least 1 identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, Al et al Anal. Chem. 2003;75(17):4646-58). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

To determine the limits of detection, a standard protein digest of BSA and Phosphorylase were purchased from Waters, and samples of 100/10/1/0.1 fmoles were spiked per run of 360 ng tomato protein sample (1/290 of the total protein extracted). A 1:1 mixture of 2 replicate protein extracts was used. Under these conditions 0.1 fmole of the standard proteins could be detected nicely.

The *Antirrhinum* proteins must be present at lower than 0.1 fmole and lower than 29 fmoles in the total sample, as described above. There were many tomato proteins with a much lower quantitative values than those of the lowest value for the standards (0.1 fmole) going down to 1/10 of the lowest value of the standard proteins. This would mean that the limit of detection

was $< 1/10$ of what was measured, and the amount of Antirrhinum proteins in the total sample was below 2.9 fmole or 0.075ng Rosea1 and 0.2 ng Delila in the total extract.

Appendix 8: Search of DELILA and ROSEA1 protein sequences against all toxin sequences in UniProt

April 28, 2025

All sequences in UniProt containing the keyword “toxin” as a molecular function (KW800) were downloaded and converted into a searchable BLAST database, named toxins_db. The total number of sequences was 109,442. This list of sequences is available at: <https://www.uniprot.org/uniprotkb?query=%28keyword%3AKW-0800%29>

The DELILA and ROSEA1 protein sequences were searched against this database using the following BLASTp query, which is considered extremely lenient/sensitive (see below for description of search parameters).

```
blastp -query DEL[or ROSEA1].fasta -db toxins_db -out results.txt -  
evaluate 1e-2 -word_size 2 -matrix PAM30 -seg no -comp_based_stats no
```

This query resulted in no hits in the toxin database with either DEL or ROS1. The search reports appear below:

Report for DELILA search:

BLASTP 2.16.0+

Database: /Users/eward/toxins.fasta

109,442 sequences; 38,712,067 total letters

Query= tr|Q38736|Q38736_ANTMA DEL OS=Antirrhinum majus OX=4151 GN=delila

PE=2 SV=1

Length=644

***** No hits found *****

Effective search space used: 21856023135

Database: /Users/eward/toxins.fasta

Posted date: Apr 28, 2025 11:12 AM

Number of letters in database: 38,712,067

Number of sequences in database: 109,442

Matrix: PAM30

Report for ROSEA1 search:

BLASTP 2.16.0+

Database: /Users/eward/toxins.fasta

109,442 sequences; 38,712,067 total letters

Query= tr|Q2VAZ7|Q2VAZ7_ANTMA ROSEA1 OS=Antirrhinum majus OX=4151 GN=ROS1

PE=2 SV=1

Length=220

***** No hits found *****

Effective search space used: 6901126669

Database: /Users/eward/toxins.fasta
Posted date: Apr 28, 2025 11:12 AM
Number of letters in database: 38,712,067
Number of sequences in database: 109,442
Matrix: PAM30

ToxinPred 3.0 Search

In addition, and as a further confirmation that neither DEL nor ROS1 is related to any known toxins, we ran both the DEL and ROS1 sequences against the ToxinPred 3.0 database (Rathore et al. 2024), using BLAST at the webserver hosted by the creators of the database: <https://webs.iitd.edu.in/raghava/toxinpred3/blast.php>

Using an E-value as large as $10e-1$ resulted in no hits.

Inspection of BLAST results

In addition to these analyses, bioinformatic searches were carried out, querying the amino acid sequences for each protein against the UniProt database, and inspecting the top 500 hits for each protein (down to 35% identity level for Delila, and 40% for Rosea1). The only database hits that show any similarity to either protein are either proteins of the same class—bHLH-type transcription factors for Delila, and Myb-type transcription factors for Rosea1—or uncharacterized or putative proteins from other plant species (likely from larger scale sequencing efforts such as whole genomes, where not every open reading frame has yet been annotated). Although these searches are less definitive than the direct comparison to all known toxin sequences (above), for completeness we include the output from each search below.

Literature search

We searched Google using keywords and queries including “toxin” and “bHLH (or Myb) transcription factor”, “evidence for toxicity of transcription factor proteins upon ingestion”, etc., and found no relevant hits. This search was re-run on January 3, 2025.

In addition, we queried the large language models Perplexity.ai, Claude Sonnet 3.5 (Anthropic), and Gemini 1.5 Flash (Google), with queries such as “Please find any evidence for toxicity upon ingestion of bHLH- (or Myb-) type transcription factors.” None of these found any relevant data. An exemplary reply (from Claude Sonnet 3.5) is:

Based on available research literature, there is no clear evidence indicating that bHLH (basic helix-loop-helix) transcription factor proteins are toxic when ingested. These proteins are naturally present in many foods we eat, as they are important regulators of gene expression in all eukaryotic organisms including plants and animals. When consumed as part of normal food, these proteins would typically be broken down into amino acids during digestion like other dietary proteins.

A few key points about bHLH proteins and oral consumption:

- 1. They are susceptible to denaturation by stomach acid and degradation by digestive enzymes*
- 2. As transcription factors, their primary biological function is within the nucleus of cells, not as circulating proteins*
- 3. Even if intact, they would have very low bioavailability when taken orally due to their large size and chemical properties*

That said, research specifically examining the toxicological effects of purified bHLH proteins upon ingestion appears to be quite limited. Given that these are common cellular proteins broken down during normal digestion, there hasn't been a strong scientific rationale to extensively study their toxicity via oral exposure.

Again, this literature search is not definitive, as contrasted with the direct search for sequence similarity between DEL and ROS1 and any known toxic proteins (above), but is included for completeness.

Conclusions

The lack of similarity between either DEL or ROS1 and known protein toxins, even when using highly lenient/sensitive BLAST parameters, provides a comprehensive analysis that conclusively shows that these proteins i) contain no known toxin motifs or domains, and ii) have no significant structural or evolutionary relationship to known toxins. This result, coupled with their exceedingly low expression, allows the conclusion that in The Purple Tomato, neither protein poses a food safety risk.

Reference

Rathore, A.S., Choudhury, S., Arara, A., Tijare, P., and Raghava, G.P.S. (2024) ToxinPred 3.0: An improved method for predicting toxicity of peptides. Computers Biol. Med. 179: 108926. [doi: 10.1016/j.compbio.2024.108926](https://doi.org/10.1016/j.compbio.2024.108926).

Explanation of search parameters chosen

E-value (Expectation Value)

Parameter: `-evalue`

Definition: The E-value represents the number of hits one would expect to see by random chance when searching a database of a particular size. It answers the question: "How many random matches would I expect to see with this score?"

Significance:

- Lower E-values (e.g., $1e-10$) indicate more significant matches
- Higher E-values (e.g., $1e-2$) indicate less significant matches that could occur by chance
- An E-value of 1 means you'd expect one random match in a database of that size
- When no matches are found even at high E-values (e.g. $1e-2$), it strongly suggests no meaningful sequence similarity exists

Word Size

Parameter: `-word_size`

Definition: The length of the initial exact matches that BLAST uses to begin its alignments (measured in amino acids for protein searches).

Significance:

- Default for protein BLAST (blastp) is 3 amino acids
- Smaller word sizes (e.g. 2) increase sensitivity, allowing detection of more distant relationships

- Smaller word sizes also increase processing time and can introduce more chance matches
- Larger word sizes make searches faster but can miss subtle similarities

Scoring Matrix

Parameter: `-matrix`

Definition: The substitution matrix used to score alignments between amino acids, accounting for the likelihood of various amino acid substitutions.

Significance:

- BLOSUM62 (default) is optimized for sequences with approximately 62% identity
- PAM30 is more sensitive for highly divergent sequences with low similarity
- PAM70 is intermediate in sensitivity
- Choice of matrix affects which alignments are considered significant

Sequence Filtering

Parameter: `-seg`

Definition: Controls filtering of low-complexity regions that might produce spurious matches.

Significance:

- Default is "yes" (filtering on)
- Setting to "no" allows all sequence regions to be compared
- Low-complexity regions often produce matches by chance rather than evolutionary relationship

Composition-Based Statistics

Parameter: `-comp_based_stats`

Definition: Adjusts scores based on sequence composition to reduce false positives.

Significance:

- Helps correct for biased amino acid composition in certain proteins
- Disabling this can increase sensitivity but may produce spurious alignments
- Particularly important when comparing sequences with unusual amino acid distributions

2025-01-03 14:26 EST

The job with ID ncbiblast-R20250103-192620-0192-3389812-plm has been submitted with these raw input values:

program

Unset

blastp

matrix

Unset

BLOSUM62

alignments

Unset

500

scores

Unset

500

exp

Unset

10

filter

Unset

F

gapalign

Unset

true

compstats

Unset

F

align

Unset

0

stype

Unset

protein

sequence

Unset

```
>QGV13004.1 Delila [Plant binary expression vector pDEL.ROS]
MATGIQNQKIVPENLRKQLAIAVRSIQWSYAIFWSNSVAQPGVLEWGDGFYNGDIKTRKT
VQSVELNPDQLGLQRSDQLRELYESLSLGETNTQAKRPTAALSPEDLTDAEWFFLVCMSF
IFNIGQGLPGRTLARNQAVWLCNAHRADTKVFSRSLAKSASIQTVVCFPYAEGVVELGA
TELPEDLNLIQHIKTSFLDSPATVPKIPNYVSNSITNNNDLICEALEHANIPENDLDQL
LNCPTDNICSPDNSLDDFADNLLIDESNLAEINGEVPQTQSWPFMDDAISNCLNSSMNS
SDCISQTHENLESFAPLSDGKGAPETNNCMHSTQECNQQIENTGVQGDEVHYQRVLSNLL
KSSHQLVLGPYFRNGNRESSFVSWNKDGSSGTHVPRSGTSQRFLKKVLFEVARMHENSRL
DAGKQKGNSDCLAKPTADEIDRNHVLSEKRRREKINERFMILASLVPSGGKVDKVSILDH
```

```
TIDYLRGLERKVDELESNKMVKGREGRESTTKTKLHDAIERTSDNYGTTRTSNVKKPLTNK
RKASDTDKIGAVNSRGLKDSLTDNITVNITNKDVLIVVSCSSKEFVLLLEVMEAVRRLSL
DSETVQSSNRDGIISITIKAKCKGLKVASASVIKQALQKVTMKS
```

database

Unset

```
uniprotkb_refprotswissprot
```

BLASTP 2.16.0+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Reference for composition-based statistics: Alejandro A. Schaffer, L. Aravind, Thomas L. Madden, Sergei Shavirin, John L. Spouge, Yuri I. Wolf, Eugene V. Koonin, and Stephen F. Altschul (2001), "Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements", Nucleic Acids Res. 29:2994-3005.

Database: uniprotkb_refprotswissprot
91,211,443 sequences; 36,979,232,029 total letters

Query= QGV13004.1 Delila [Plant binary expression vector pDEL.ROS]

Length=644

Sequences producing significant alignments:	Score (Bits)	E Value
TR:A0AAE1YRU8 A0AAE1YRU8_9LAMI Transcription factor GLABRA 3 OS=Se...	877	0.0
TR:A0A6I9T9F3 A0A6I9T9F3_SESIN Transcription factor EGL1 isoform X...	863	0.0
TR:A0A6I9STS8 A0A6I9STS8_SESIN Transcription factor EGL1-like OS=S...	833	0.0
TR:A0A8M8UV68 A0A8M8UV68_SESIN Transcription factor GLABRA 3 isofo...	813	0.0
TR:A0AAD1YSG9 A0AAD1YSG9_9LAMI BHLH domain-containing protein OS=F...	788	0.0

TR:A0A8S0V9U4	A0A8S0V9U4_OLEEU	Myc-like anthocyanin regulatory OS=...	786	0.0
TR:A0A8S0UEL7	A0A8S0UEL7_OLEEU	Transcription factor EGL1 OS=Olea e...	785	0.0
TR:A0A8S0V9B8	A0A8S0V9B8_OLEEU	Myc-like anthocyanin regulatory OS=...	772	0.0
TR:A0A8S0UJ42	A0A8S0UJ42_OLEEU	Transcription factor EGL1 OS=Olea e...	771	0.0
TR:A0AAD2ABW8	A0AAD2ABW8_9LAMI	BHLH domain-containing protein OS=F...	766	0.0
TR:A0A022RE76	A0A022RE76_ERYGU	BHLH domain-containing protein OS=E...	757	0.0
TR:A0A8X8ZMH1	A0A8X8ZMH1_SALSN	BHLH domain-containing protein OS=S...	756	0.0
TR:A0A2Z7BHD0	A0A2Z7BHD0_9LAMI	Basic helix-loop-helix DNA-binding ...	756	0.0
TR:A0AAE2CHT4	A0AAE2CHT4_9LAMI	Transcription factor GLABRA 3 OS=Se...	755	0.0
TR:A0A5J4ZPL9	A0A5J4ZPL9_9ASTE	BHLH domain-containing protein OS=N...	717	0.0
TR:A0A6J1A585	A0A6J1A585_9ROSI	Transcription factor EGL1 isoform X...	717	0.0
TR:A0A8X8ZKM2	A0A8X8ZKM2_SALSN	BHLH domain-containing protein OS=S...	712	0.0
TR:A0AAE2BW4	A0AAE2BW4_9LAMI	Transcription factor EGL1 OS=Sesamu...	711	0.0
TR:A0AAD4JJI5	A0AAD4JJI5_PERFH	BHLH domain-containing protein OS=P...	710	0.0
TR:A0A061GCJ1	A0A061GCJ1_THECC	Basic helix-loop-helix DNA-binding ...	709	0.0
TR:A0A5J4ZUK8	A0A5J4ZUK8_9ASTE	BHLH domain-containing protein OS=N...	707	0.0
TR:A0A6P5YE92	A0A6P5YE92_DURZI	Transcription factor EGL1-like isof...	698	0.0
TR:A0A8X8ZIC7	A0A8X8ZIC7_SALSN	BHLH domain-containing protein OS=S...	690	0.0
TR:A0A1R3GKN1	A0A1R3GKN1_9ROSI	BHLH domain-containing protein OS=C...	683	0.0
TR:A0A5E4G9W2	A0A5E4G9W2_PRUDU	PREDICTED: mRNAion factor OS=Prunus...	682	0.0
TR:A0A6J5XEN3	A0A6J5XEN3_PRUAR	BHLH domain-containing protein OS=P...	681	0.0
TR:A0A6P5YF69	A0A6P5YF69_DURZI	Transcription factor EGL1-like isof...	680	0.0
TR:A0A6P6A5X8	A0A6P6A5X8_DURZI	Transcription factor EGL1-like OS=D...	679	0.0
TR:A0A251P9S8	A0A251P9S8_PRUPE	BHLH domain-containing protein OS=P...	679	0.0
TR:A0AA39AHR6	A0AA39AHR6_VITRO	BHLH domain-containing protein OS=V...	679	0.0
TR:A0A8J4Q5L7	A0A8J4Q5L7_9ROSI	BHLH domain-containing protein OS=C...	677	0.0
TR:A0A6P3Z877	A0A6P3Z877_ZIZJJ	Transcription factor EGL1-like OS=Z...	676	0.0
TR:A0A8X8ZC95	A0A8X8ZC95_SALSN	BHLH domain-containing protein OS=S...	675	0.0
TR:A0A6P5T1W3	A0A6P5T1W3_PRUAV	Transcription factor EGL1 OS=Prunus...	674	0.0
TR:F6HUC4	F6HUC4_VITVI	BHLH domain-containing protein OS=Vitis vin...	671	0.0
TR:A0A5D2YCI7	A0A5D2YCI7_GOSMU	BHLH domain-containing protein OS=G...	670	0.0
TR:A0A0B0NQ03	A0A0B0NQ03_GOSAR	Transcription factor GLABRA 3-like ...	669	0.0
TR:A0A6P6TS56	A0A6P6TS56_COFAR	Transcription factor EGL1-like isof...	668	0.0
TR:A0A6P6TRQ8	A0A6P6TRQ8_COFAR	Transcription factor EGL1-like isof...	668	0.0
TR:A0A1R3IAD6	A0A1R3IAD6_COCAP	BHLH domain-containing protein OS=C...	667	0.0
TR:A0A9D3W1K1	A0A9D3W1K1_9ROSI	BHLH domain-containing protein OS=G...	662	0.0
TR:A0A068UUT4	A0A068UUT4_COFCA	BHLH domain-containing protein OS=C...	659	0.0
TR:A0AA39AHG3	A0AA39AHG3_VITRO	BHLH domain-containing protein OS=V...	659	0.0
TR:A0A5D2TZ36	A0A5D2TZ36_GOSMU	BHLH domain-containing protein OS=G...	659	0.0
TR:A0A1U8MDT8	A0A1U8MDT8_GOSHI	Transcription factor EGL1-like OS=G...	659	0.0
TR:A0A4S4E0X4	A0A4S4E0X4_CAMSI	BHLH domain-containing protein OS=C...	658	0.0
TR:A0A0D2PFY9	A0A0D2PFY9_GOSRA	Basic helix-loop-helix protein 062D...	658	0.0
TR:A0A7N2N4M7	A0A7N2N4M7_QUELO	Uncharacterized protein OS=Quercus ...	656	0.0
TR:A0A6P3Z0Y4	A0A6P3Z0Y4_ZIZJJ	Transcription factor GLABRA 3-like ...	653	0.0
TR:A0A6J1A4E9	A0A6J1A4E9_9ROSI	Transcription factor EGL1 isoform X...	653	0.0
TR:A0A6P5XYR8	A0A6P5XYR8_DURZI	Transcription factor EGL1-like isof...	650	0.0
TR:A0A5B6VIS0	A0A5B6VIS0_9ROSI	Basic helix-loop-helix protein 062A...	649	0.0
TR:A0AAD5NR45	A0AAD5NR45_ACENE	BHLH domain-containing protein OS=A...	649	0.0
TR:A0A5B6VIS6	A0A5B6VIS6_9ROSI	Basic helix-loop-helix protein 062A...	648	0.0
TR:A0A7J9L3D3	A0A7J9L3D3_GOSSC	BHLH domain-containing protein OS=G...	647	0.0
TR:A0A8T1NWQ6	A0A8T1NWQ6_CARIL	BHLH domain-containing protein OS=C...	647	0.0
TR:A0A061G5B2	A0A061G5B2_THECC	Basic helix-loop-helix DNA-binding ...	645	0.0
TR:A0A5N5FZP5	A0A5N5FZP5_9ROSA	Transcription factor EGL1-like OS=P...	644	0.0
TR:A0A9W7HF92	A0A9W7HF92_HIBTR	GLABRA 3, GLABROUS 3 OS=Hibiscus tr...	644	0.0
TR:A0A2R6QNC0	A0A2R6QNC0_ACTCC	Transcription factor GLABRA like OS...	644	0.0

TR:A0AAE0AZX2	A0AAE0AZX2_9ROSI	BHLH domain-containing protein OS=D...	643	0.0
TR:A0A1Q3CUIY7	A0A1Q3CUIY7_CEPFO	HLH domain-containing protein/bHLH-...	643	0.0
TR:A0A6A1VQY7	A0A6A1VQY7_9ROSI	Transcription factor GLABRA 3 OS=Mo...	642	0.0
TR:A0A2I4H8L9	A0A2I4H8L9_JUGRE	Transcription factor EGL1-like OS=J...	642	0.0
TR:A0A9Q1MAQ3	A0A9Q1MAQ3_9SOLA	BHLH domain-containing protein OS=A...	641	0.0
TR:A0A5N5FZS4	A0A5N5FZS4_9ROSA	Transcription factor EGL1-like OS=P...	641	0.0
TR:A0AAD9XQQ9	A0AAD9XQQ9_9ROSI	BHLH domain-containing protein OS=D...	641	0.0
TR:A0A7J8X055	A0A7J8X055_GOSAI	BHLH domain-containing protein OS=G...	640	0.0
TR:A0AA39SBG4	A0AA39SBG4_ACESA	BHLH domain-containing protein OS=A...	640	0.0
TR:A0A7J8ZDD6	A0A7J8ZDD6_9ROSI	BHLH domain-containing protein OS=G...	639	0.0
TR:A0A2K2B670	A0A2K2B670_POPTR	BHLH domain-containing protein OS=P...	639	0.0
TR:A0A2K2BVJ7	A0A2K2BVJ7_POPTR	BHLH domain-containing protein OS=P...	638	0.0
TR:A0A8K0HB17	A0A8K0HB17_9ROSA	BHLH domain-containing protein OS=R...	637	0.0
TR:A0A5N6R694	A0A5N6R694_9ROSI	BHLH domain-containing protein OS=C...	637	0.0
TR:A0A6A2ZUN8	A0A6A2ZUN8_HIBSY	Transcription factor GLABRA 3 OS=Hi...	637	0.0
TR:A0A6P5XXI2	A0A6P5XXI2_DURZI	Transcription factor EGL1-like isof...	636	0.0
TR:A0A8T1P5U4	A0A8T1P5U4_CARIL	BHLH domain-containing protein OS=C...	636	0.0
TR:A0A7J9DT35	A0A7J9DT35_9ROSI	BHLH domain-containing protein OS=G...	636	0.0
TR:A0AAJ6U490	A0AAJ6U490_POPEU	Transcription factor GLABRA 3 isofo...	635	0.0
TR:A0A1S3XEW8	A0A1S3XEW8_TOBAC	Transcription factor EGL1-like isof...	635	0.0
TR:A0AAD6RKF9	A0AAD6RKF9_9ROSI	Transcription factor GLABRA 3-like ...	634	0.0
TR:A0A7J9BM73	A0A7J9BM73_GOSGO	BHLH domain-containing protein OS=G...	633	0.0
TR:A0A2P6P6Z3	A0A2P6P6Z3_ROSCH	Putative transcription factor bHLH ...	633	0.0
TR:A0A4U5PMG8	A0A4U5PMG8_POPAL	Regulator of anthocyanin biosynthes...	633	0.0
TR:A0A067JN60	A0A067JN60_JATCU	BHLH domain-containing protein OS=J...	633	0.0
TR:A0A498JIJ8	A0A498JIJ8_MALDO	BHLH domain-containing protein OS=M...	633	0.0
TR:A0AAJ6UG00	A0AAJ6UG00_POPEU	Transcription factor GLABRA 3 OS=Po...	631	0.0
TR:A0A2I4DDZ7	A0A2I4DDZ7_JUGRE	Transcription factor EGL1-like OS=J...	631	0.0
TR:A0A8S0R833	A0A8S0R833_OLEEU	Transcription factor GLABRA 3-like ...	631	0.0
TR:A0A5C7H115	A0A5C7H115_9ROSI	BHLH domain-containing protein OS=A...	630	0.0
TR:A0A9Q1QUQ4	A0A9Q1QUQ4_9SOLA	BHLH domain-containing protein OS=A...	629	0.0
TR:A0A8T2ZWN5	A0A8T2ZWN5_POPDE	BHLH domain-containing protein OS=P...	629	0.0
TR:A0A1J6J1G8	A0A1J6J1G8_NICAT	Transcription factor glabra 3 OS=Ni...	629	0.0
TR:A0A9Q0PWM8	A0A9Q0PWM8_9ROSI	TRANSCRIPTION FACTOR EGL1 OS=Salix ...	628	0.0
TR:A0A7J7GGQ9	A0A7J7GGQ9_CAMSI	BHLH domain-containing protein OS=C...	628	0.0
TR:A0A9Q0Q852	A0A9Q0Q852_9ROSI	TRANSCRIPTION FACTOR EGL1 OS=Salix ...	628	0.0
TR:A0A6A6MK62	A0A6A6MK62_HEVBR	BHLH domain-containing protein OS=H...	627	0.0
TR:A0A9W7JGS7	A0A9W7JGS7_HIBTR	GLABRA 3, GLABROUS 3 OS=Hibiscus tr...	625	0.0
TR:A0A8B8QZ79	A0A8B8QZ79_9MYRT	Transcription factor EGL1-like OS=R...	625	0.0
TR:A0AA88VCM2	A0AA88VCM2_9ASTE	BHLH domain-containing protein OS=E...	624	0.0
TR:A0A2R6Q8J1	A0A2R6Q8J1_ACTCC	Transcription factor GLABRA like OS...	624	0.0
TR:A0A9W7MRC4	A0A9W7MRC4_HIBTR	GLABRA 3, GLABROUS 3 OS=Hibiscus tr...	624	0.0
TR:A0A1U7ZMP0	A0A1U7ZMP0_NELNU	Transcription factor GLABRA 3-like ...	624	0.0
TR:A0A9Q0T937	A0A9Q0T937_SALPP	TRANSCRIPTION FACTOR EGL1 OS=Salix ...	622	0.0
TR:A0A1U7WZT6	A0A1U7WZT6_NICSY	Transcription factor EGL1-like isof...	622	0.0
TR:A0A1S4D823	A0A1S4D823_TOBAC	Transcription factor EGL1-like OS=N...	622	0.0
TR:A0A7J6IA04	A0A7J6IA04_CANSA	BHLH domain-containing protein OS=C...	621	0.0
TR:A0A1S3XEX3	A0A1S3XEX3_TOBAC	Transcription factor EGL1-like isof...	621	0.0
TR:A0AAJ6XL47	A0AAJ6XL47_POPEU	Transcription factor GLABRA 3 isofo...	621	0.0
TR:A0A2P5EGA3	A0A2P5EGA3_TREOI	Basic helix-loop-helix transcriptio...	620	0.0
TR:A0AAF0UDC2	A0AAF0UDC2_SOLVR	BHLH domain-containing protein OS=S...	620	0.0
TR:A0AA88X230	A0AA88X230_9ASTE	BHLH domain-containing protein OS=E...	619	0.0
TR:A0A6A2Y7E0	A0A6A2Y7E0_HIBSY	Anthocyanin regulatory R-S protein ...	617	0.0
TR:A0A2C9WCF8	A0A2C9WCF8_MANES	BHLH domain-containing protein OS=M...	617	0.0
TR:A0A8S0R8A9	A0A8S0R8A9_OLEEU	Transcription factor GLABRA 3-like ...	616	0.0

TR:A0A4V6A815	A0A4V6A815_POPAL	Regulator of anthocyanin biosynthes...	616	0.0
TR:A0A2P5CPB1	A0A2P5CPB1_PARAD	Basic helix-loop-helix transcriptio...	616	0.0
TR:A0A9Q0PVZ8	A0A9Q0PVZ8_9ROSI	TRANSCRIPTION FACTOR TT8 OS=Salix k...	614	0.0
TR:A0A5N5NAB2	A0A5N5NAB2_9ROSI	BHLH domain-containing protein OS=S...	612	0.0
TR:A0A8X8A9I3	A0A8X8A9I3_POPTO	BHLH domain-containing protein OS=P...	610	0.0
TR:A0A8X8A8B2	A0A8X8A8B2_POPTO	BHLH domain-containing protein OS=P...	609	0.0
TR:A0A9Q0QF81	A0A9Q0QF81_SALPP	TRANSCRIPTION FACTOR TT8 OS=Salix p...	608	0.0
TR:A0A8S0R8W9	A0A8S0R8W9_OLEEU	Transcription factor GLABRA 3-like ...	608	0.0
TR:A0A1U7WR04	A0A1U7WR04_NICSY	Transcription factor EGL1-like isof...	608	0.0
TR:A0A9Q0TIY3	A0A9Q0TIY3_SALPP	TRANSCRIPTION FACTOR EGL1 OS=Salix ...	607	0.0
TR:A0A835K9W9	A0A835K9W9_9ROSI	BHLH domain-containing protein OS=S...	607	0.0
TR:A0A9N7NGU8	A0A9N7NGU8_STRHE	Transcription factor GLABRA 3 (Frag...	605	0.0
TR:A0A5A7QQX0	A0A5A7QQX0_STRAF	Basic helix-loop-helix (BHLH) DNA-b...	604	0.0
TR:A0A498HZI2	A0A498HZI2_MALDO	BHLH domain-containing protein (Fra...	603	0.0
TR:A0A5N5P1L3	A0A5N5P1L3_9ROSI	BHLH domain-containing protein OS=S...	601	0.0
TR:A0AAD7KSE4	A0AAD7KSE4_QUISA	Basic helix-loop-helix transcriptio...	601	0.0
TR:A0A835NAG1	A0A835NAG1_9ROSI	BHLH domain-containing protein OS=S...	599	0.0
TR:A0AAD6WB27	A0AAD6WB27_9ROSI	Transcription factor GLABRA 3-like ...	598	0.0
TR:A0A8B8KF88	A0A8B8KF88_ABRPR	Transcription factor EGL1-like OS=A...	594	0.0
TR:A0A2G3ADX9	A0A2G3ADX9_CAPAN	Transcription factor MYC1 OS=Capsic...	593	0.0
TR:A0A6A4LBC4	A0A6A4LBC4_9ERIC	BHLH domain-containing protein (Fra...	593	0.0
TR:A0A1U7YSA1	A0A1U7YSA1_NELNU	Transcription factor GLABRA 3-like ...	591	0.0
TR:A0AAD7KSN9	A0AAD7KSN9_QUISA	Basic helix-loop-helix transcriptio...	589	0.0
TR:A0AAD6W9F4	A0AAD6W9F4_9ROSI	Transcription factor GLABRA 3-like ...	588	0.0
TR:A0AAE1J6Y5	A0AAE1J6Y5_9FABA	BHLH domain-containing protein OS=A...	587	0.0
TR:A0A445J8F5	A0A445J8F5_GLYSO	Transcription factor GLABRA 3 isofo...	587	0.0
TR:ILKPA9	ILKPA9_SOYBN	BHLH domain-containing protein OS=Glycine m...	587	0.0
TR:A0A7J0DKP6	A0A7J0DKP6_9ERIC	Basic helix-loop-helix (BHLH) DNA-b...	586	0.0
TR:A0AA86S0H2	A0AA86S0H2_9FABA	Uncharacterized protein OS=Sphenost...	586	0.0
TR:A0A9D3VE03	A0A9D3VE03_9ROSI	BHLH domain-containing protein OS=G...	584	0.0
TR:A0A1S2YFV2	A0A1S2YFV2_CICAR	Transcription factor EGL1 isoform X...	583	0.0
TR:A0A8J6CMY3	A0A8J6CMY3_9ROSI	BHLH domain-containing protein OS=G...	583	0.0
TR:A0A7J7C7B4	A0A7J7C7B4_TRIWF	Transcription factor EGL1 OS=Tripte...	583	0.0
TR:A0A834SWW6	A0A834SWW6_9FABA	Transcription factor GLABRA 3 OS=Se...	580	0.0
TR:A0A7J8T6S4	A0A7J8T6S4_GOSDV	BHLH domain-containing protein OS=G...	580	0.0
TR:A0A1U7WIG4	A0A1U7WIG4_NICSY	Transcription factor GLABRA 3-like ...	579	0.0
TR:A0A7J7CKA3	A0A7J7CKA3_TRIWF	Transcription factor GLABRA 3-like ...	578	0.0
TR:A0A7J9GGR4	A0A7J9GGR4_9ROSI	BHLH domain-containing protein (Fra...	578	0.0
TR:A0A7J8ZRA0	A0A7J8ZRA0_9ROSI	BHLH domain-containing protein OS=G...	577	0.0
TR:A0A9Q0JBI9	A0A9Q0JBI9_9ROSI	BHLH domain-containing protein OS=T...	577	0.0
TR:A0A0D2T7P4	A0A0D2T7P4_GOSRA	Basic helix-loop-helix protein 110D...	577	0.0
TR:A0A8X8AF77	A0A8X8AF77_POPTO	BHLH domain-containing protein OS=P...	576	0.0
TR:A0A7J9LJV3	A0A7J9LJV3_GOSSC	BHLH domain-containing protein OS=G...	576	0.0
TR:A0A0G2UY89	A0A0G2UY89_GOSHI	Basic helix-loop-helix protein 110A...	576	0.0
TR:Q9ATD8	Q9ATD8_GOSHI	Basic helix-loop-helix protein 110D OS=Goss...	576	0.0
TR:A0AAE1S8X4	A0AAE1S8X4_9SOLA	BHLH domain-containing protein OS=A...	576	0.0
TR:A0A7J9E6X1	A0A7J9E6X1_9ROSI	BHLH domain-containing protein OS=G...	575	0.0
TR:A0A835CWM6	A0A835CWM6_TETSI	BHLH domain-containing protein OS=T...	575	0.0
TR:A0A166JB04	A0A166JB04_DAUCS	Uncharacterized protein OS=Daucus c...	575	0.0
TR:A0A2P5WPR2	A0A2P5WPR2_GOSBA	BHLH domain-containing protein OS=G...	575	0.0
TR:A0A1S3UWV2	A0A1S3UWV2_VIGRR	Transcription factor EGL1 OS=Vigna ...	574	0.0
TR:W9QPL7	W9QPL7_9ROSA	BHLH domain-containing protein OS=Morus not...	573	0.0
TR:A0A5D2X135	A0A5D2X135_GOSMU	BHLH domain-containing protein OS=G...	572	0.0
TR:A0A835D4N5	A0A835D4N5_TETSI	BHLH domain-containing protein OS=T...	571	0.0
TR:A0A0S3QXK1	A0A0S3QXK1_PHAAN	BHLH domain-containing protein OS=V...	570	0.0

TR:A0A9Q0GUU2	A0A9Q0GUU2_9MAGN	BHLH domain-containing protein OS=P...	570	0.0
TR:A0A445KS20	A0A445KS20_GLYSO	Transcription factor GLABRA 3 OS=Gl...	569	0.0
TR:I1K6H9	I1K6H9_SOYBN	BHLH domain-containing protein OS=Glycine m...	569	0.0
TR:V7CPJ5	V7CPJ5_PHAVU	BHLH domain-containing protein OS=Phaseolus...	569	0.0
TR:A0AA87ZG91	A0AA87ZG91_FICCA	BHLH domain-containing protein OS=F...	568	0.0
TR:A0A9Q0QBK3	A0A9Q0QBK3_SALVM	TRANSCRIPTION FACTOR EGL1 OS=Salix ...	568	0.0
TR:A0A834HKU7	A0A834HKU7_RHOSS	BHLH domain-containing protein OS=R...	567	0.0
TR:A0AA88UGZ0	A0AA88UGZ0_9ASTE	BHLH domain-containing protein OS=E...	566	0.0
TR:A0AA39DGM9	A0AA39DGM9_VITRO	BHLH domain-containing protein OS=V...	565	0.0
TR:A0A061G4I5	A0A061G4I5_THECC	Basic helix-loop-helix DNA-binding ...	564	0.0
TR:A0A2R6PHY5	A0A2R6PHY5_ACTCC	Transcription factor GLABRA like OS...	562	0.0
TR:F6I629	F6I629_VITVI	BHLH domain-containing protein OS=Vitis vin...	562	0.0
TR:A0A151TF17	A0A151TF17_CAJCA	Transcription factor GLABRA 3 (Frag...	562	0.0
TR:A0A251P6B0	A0A251P6B0_PRUPE	BHLH domain-containing protein OS=P...	561	0.0
TR:V4TMA4	V4TMA4_CITCL	BHLH domain-containing protein OS=Citrus cl...	561	0.0
TR:A0A5J4ZE91	A0A5J4ZE91_9ASTE	BHLH domain-containing protein OS=N...	560	0.0
TR:A0A8X8IV72	A0A8X8IV72_POPTO	BHLH domain-containing protein OS=P...	558	0.0
TR:A0A7J8XF04	A0A7J8XF04_GOSAI	BHLH domain-containing protein OS=G...	558	0.0
TR:A0A1U7Z3P7	A0A1U7Z3P7_NELNU	Transcription factor GLABRA 3-like ...	557	0.0
TR:A0A0L9VQT8	A0A0L9VQT8_PHAAN	BHLH domain-containing protein OS=P...	557	0.0
TR:A0A6A4P7Z4	A0A6A4P7Z4_LUPAL	Putative transcription factor bHLH ...	556	0.0
TR:A0A2I0J5E7	A0A2I0J5E7_PUNGR	BHLH domain-containing protein OS=P...	555	0.0
TR:A0A200QU96	A0A200QU96_MACCD	Myc-type OS=Macleaya cordata OX=568...	554	0.0
TR:A0A2C9WI35	A0A2C9WI35_MANES	BHLH domain-containing protein OS=M...	553	0.0
TR:A0A6A4QK93	A0A6A4QK93_LUPAL	Putative transcription factor bHLH ...	552	0.0
TR:A0A6P5ZX88	A0A6P5ZX88_DURZI	Transcription factor EGL1-like OS=D...	552	0.0
TR:A0AAF0W436	A0AAF0W436_DAUCS	BHLH domain-containing protein OS=D...	551	0.0
TR:A0A6D2JT09	A0A6D2JT09_9BRAS	BHLH domain-containing protein OS=M...	550	0.0
TR:A0A9N7MQZ8	A0A9N7MQZ8_STRHE	Transcription factor GLABRA 3 OS=St...	550	0.0
TR:A0A6P4CQC1	A0A6P4CQC1_ARADU	Transcription factor GLABRA 3 OS=Ar...	550	0.0
TR:A0A2C9WI97	A0A2C9WI97_MANES	BHLH domain-containing protein OS=M...	550	0.0
TR:A0A6A6LDX8	A0A6A6LDX8_HEVBR	BHLH domain-containing protein OS=H...	549	0.0
TR:A0A7N0V772	A0A7N0V772_KALFE	BHLH domain-containing protein OS=K...	548	0.0
TR:A0A4S4EZC5	A0A4S4EZC5_CAMSI	Uncharacterized protein OS=Camellia...	547	0.0
TR:A0A1J7HAM8	A0A1J7HAM8_LUPAN	BHLH domain-containing protein OS=L...	547	0.0
TR:A0A7J8UNF8	A0A7J8UNF8_9ROSI	BHLH domain-containing protein OS=G...	545	0.0
TR:A0AA38Z3B5	A0AA38Z3B5_VITRO	BHLH domain-containing protein OS=V...	545	0.0
TR:A0AAD9ZYI8	A0AAD9ZYI8_9ROSI	BHLH domain-containing protein OS=D...	544	0.0
TR:A0AAD6PE91	A0AAD6PE91_9ROSI	BHLH domain-containing protein OS=S...	543	0.0
TR:A0A7N0V7A1	A0A7N0V7A1_KALFE	BHLH domain-containing protein OS=K...	543	0.0
TR:A0AA39RZK3	A0AA39RZK3_ACESA	BHLH domain-containing protein OS=A...	542	0.0
TR:A0A7J9JCI3	A0A7J9JCI3_9ROSI	BHLH domain-containing protein (Fra...	542	0.0
TR:A0AAD5J6B1	A0AAD5J6B1_ACENE	BHLH domain-containing protein OS=A...	542	0.0
TR:A0A9D4VGY4	A0A9D4VGY4_PEA	BHLH domain-containing protein OS=Pis...	540	0.0
TR:A0A3S4N375	A0A3S4N375_9MAGN	Transcription factor GLABRA 3-like ...	539	0.0
TR:A0A8B8LW61	A0A8B8LW61_ABRPR	Transcription factor EGL1-like OS=A...	539	0.0
TR:A0A9J5XXZ9	A0A9J5XXZ9_SOLCO	BHLH domain-containing protein OS=S...	537	0.0
TR:A0A087GT19	A0A087GT19_ARAAL	BHLH domain-containing protein OS=A...	537	0.0
TR:R0F473	R0F473_9BRAS	BHLH domain-containing protein OS=Capsella ...	536	0.0
TR:A0A067GLQ4	A0A067GLQ4_CITSI	BHLH domain-containing protein OS=C...	535	0.0
TR:A0A2G9GK97	A0A2G9GK97_9LAMI	BHLH domain-containing protein OS=H...	532	0.0
TR:A0A8J4QYW5	A0A8J4QYW5_9ROSI	BHLH domain-containing protein OS=C...	536	1e-180
TR:A0A5N5HXW5	A0A5N5HXW5_9ROSA	Transcription factor EGL1 OS=Pyrus ...	532	2e-180
TR:A0A8T1RIX0	A0A8T1RIX0_CARIL	BHLH domain-containing protein OS=C...	535	4e-180
TR:A0AAD9UA74	A0AAD9UA74_9ROSI	BHLH domain-containing protein OS=D...	533	1e-179

TR:A0AA39AG57	A0AA39AG57_VITRO	BHLH domain-containing protein OS=V...	526	2e-179
TR:A0A0D2S5L4	A0A0D2S5L4_GOSRA	BHLH domain-containing protein OS=G...	530	3e-179
TR:D7MIW9	D7MIW9_ARALL	BHLH domain-containing protein OS=Arabidops...	531	3e-179
TR:A0A1U8J7P4	A0A1U8J7P4_GOSHI	Transcription factor EGL1-like isof...	530	4e-179
TR:A0A1U8L4Y0	A0A1U8L4Y0_GOSHI	Transcription factor EGL1-like isof...	530	4e-179
TR:A0A072U5D1	A0A072U5D1_MEDTR	BHLH transcription factor OS=Medica...	529	1e-178
TR:A0A444ZYZ7	A0A444ZYZ7_ARAHY	BHLH domain-containing protein OS=A...	529	1e-178
TR:A0A6J1A427	A0A6J1A427_9ROSI	Transcription factor GLABRA 3 isofo...	526	1e-178
TR:A0A5D2SMW4	A0A5D2SMW4_GOSMU	BHLH domain-containing protein OS=G...	528	2e-178
TR:A0A2I4DLE3	A0A2I4DLE3_JUGRE	Transcription factor EGL1-like OS=J...	530	2e-178
TR:A0A2G5E0Y9	A0A2G5E0Y9_AQUCA	BHLH domain-containing protein OS=A...	529	8e-178
TR:A0A9W7LJN7	A0A9W7LJN7_HIBTR	GLABRA 3, GLABROUS 3 OS=Hibiscus tr...	528	9e-178
TR:A0A6J0ZZT1	A0A6J0ZZT1_9ROSI	Uncharacterized protein LOC11041365...	548	1e-177
TR:A0A565C7A6	A0A565C7A6_9BRAS	BHLH domain-containing protein OS=A...	527	2e-177
TR:A0A9Q0GVJ6	A0A9Q0GVJ6_9MAGN	BHLH domain-containing protein OS=P...	529	2e-177
TR:A0A679KKB7	A0A679KKB7_BRACM	BHLH domain-containing protein OS=B...	526	3e-177
SP:Q9FN69	GL3_ARATH	Transcription factor GLABRA 3 OS=Arabidopsis t...	524	2e-176
TR:A0A1P8BHU1	A0A1P8BHU1_ARATH	Basic helix-loop-helix (BHLH) DNA-b...	524	3e-176
TR:A0A178UFB0	A0A178UFB0_ARATH	Basic helix-loop-helix (BHLH) DNA-b...	524	3e-176
TR:A0A8S0UBL7	A0A8S0UBL7_OLEEU	Myc-like anthocyanin regulatory OS=...	518	7e-176
TR:A0A0D2QPW9	A0A0D2QPW9_GOSRA	Basic helix-loop-helix protein 123D...	523	7e-176
TR:A0A1U8L1M3	A0A1U8L1M3_GOSHI	Transcription factor MYC1-like isof...	522	1e-175
TR:A0A978UGJ1	A0A978UGJ1_ZIZJJ	BHLH domain-containing protein OS=Z...	524	1e-175
TR:A0A8X7VAT0	A0A8X7VAT0_BRACI	BHLH domain-containing protein OS=B...	522	2e-175
TR:A0AAD3Y7M8	A0AAD3Y7M8_NEPGR	BHLH domain-containing protein OS=N...	520	3e-175
TR:A0A061DV61	A0A061DV61_THECC	Basic helix-loop-helix DNA-binding ...	541	3e-175
TR:V4TXI7	V4TXI7_CITCL	BHLH domain-containing protein OS=Citrus cl...	522	3e-175
TR:A0A067H5G2	A0A067H5G2_CITSI	BHLH domain-containing protein OS=C...	522	3e-175
TR:A0A5D2SQT9	A0A5D2SQT9_GOSMU	Uncharacterized protein OS=Gossypiu...	521	4e-175
TR:A0A5C7I2M6	A0A5C7I2M6_9ROSI	BHLH domain-containing protein OS=A...	521	5e-175
TR:A0A0G2US56	A0A0G2US56_GOSHI	Basic helix-loop-helix protein 123D...	520	6e-175
TR:A0A8K0IG95	A0A8K0IG95_COCNU	Transcription factor GLABRA 3 OS=Co...	522	6e-175
TR:A0A7N2MBQ9	A0A7N2MBQ9_QUELO	BHLH domain-containing protein OS=Q...	521	9e-175
TR:A0A061G6F1	A0A061G6F1_THECC	Basic helix-loop-helix DNA-binding ...	515	2e-174
TR:A0A8X7RCW1	A0A8X7RCW1_BRACI	BHLH domain-containing protein OS=B...	519	2e-174
TR:A0A9D3ZZH4	A0A9D3ZZH4_9ROSI	BHLH domain-containing protein OS=G...	519	2e-174
TR:A0AAE1J2P8	A0AAE1J2P8_9FABA	BHLH domain-containing protein OS=A...	516	1e-173
TR:A0A8B8Q3D0	A0A8B8Q3D0_9MYRT	Transcription factor EGL1-like isof...	518	1e-173
TR:A0A1U8L567	A0A1U8L567_GOSHI	Transcription factor MYC1-like isof...	516	2e-173
TR:A0A0B0P3H4	A0A0B0P3H4_GOSAR	Basic helix-loop-helix protein 123A...	516	3e-173
TR:A0AAD3TGS0	A0AAD3TGS0_NEPGR	BHLH domain-containing protein OS=N...	514	4e-173
TR:A0A6I9R2J5	A0A6I9R2J5_ELAGV	Transcription factor GLABRA 3 OS=El...	518	4e-173
TR:A0A2R6R9T2	A0A2R6R9T2_ACTCC	Transcription factor GLABRA like OS...	515	5e-173
TR:A0A6J0NFF1	A0A6J0NFF1_RAPSA	Transcription factor GLABRA 3 isofo...	515	5e-173
TR:A0A1Q3CLU9	A0A1Q3CLU9_CEPFO	BHLH-MYC_N domain-containing protei...	516	1e-172
TR:A0A6J0NE76	A0A6J0NE76_RAPSA	Transcription factor GLABRA 3 isofo...	515	1e-172
TR:M5XIR9	M5XIR9_PRUPE	Uncharacterized protein OS=Prunus persica O...	514	3e-172
TR:A0A9W7LJ77	A0A9W7LJ77_HIBTR	GLABRA 3, GLABROUS 3 OS=Hibiscus tr...	514	3e-172
TR:A0A314Y2C5	A0A314Y2C5_PRUYE	BHLH33 OS=Prunus yedoensis var. nud...	514	4e-172
TR:A0A0A7W5E0	A0A0A7W5E0_PRUAV	BHLH33 OS=Prunus avium OX=42229 GN=...	513	6e-172
TR:A0A484LERO	A0A484LERO_9ASTE	BHLH domain-containing protein OS=C...	511	9e-172
TR:A0AAD6NTK4	A0AAD6NTK4_9ROSI	BHLH domain-containing protein OS=S...	511	5e-171
TR:A0A5E4FTJ7	A0A5E4FTJ7_PRUDU	PREDICTED: mRNAion factor OS=Prunus...	510	9e-171
TR:A0A5A7PKD1	A0A5A7PKD1_STRAF	Basic helix-loop-helix (BHLH) DNA-b...	506	4e-170
TR:A0A7J8NH28	A0A7J8NH28_9ROSI	Uncharacterized protein OS=Gossypiu...	506	3e-169

TR:A0A7J9JBD9	A0A7J9JBD9_9ROSI	BHLH domain-containing protein OS=G...	506	3e-169
TR:A0A2R6Q8X5	A0A2R6Q8X5_ACTCC	Transcription factor GLABRA like OS...	506	3e-169
TR:A0A9Q0ZCX6	A0A9Q0ZCX6_SALVM	TRANSCRIPTION FACTOR TT8 OS=Salix v...	506	3e-169
TR:A0A7J9E8F6	A0A7J9E8F6_9ROSI	Uncharacterized protein (Fragment) ...	505	5e-169
TR:A0A9Q0TEG7	A0A9Q0TEG7_9ROSI	TRANSCRIPTION FACTOR TT8 OS=Salix k...	505	1e-168
TR:A0A6A3CVK1	A0A6A3CVK1_HIBSY	Anthocyanin regulatory Lc protein O...	504	1e-168
TR:A0A9Q0PDJ8	A0A9Q0PDJ8_SALPP	TRANSCRIPTION FACTOR TT8 OS=Salix p...	504	3e-168
TR:A0AAE1IZ17	A0AAE1IZ17_9FABA	BHLH domain-containing protein OS=A...	502	3e-168
TR:A0A7J8UQL9	A0A7J8UQL9_9ROSI	BHLH domain-containing protein (Fra...	503	4e-168
TR:A0A6P6TT62	A0A6P6TT62_COFAR	Transcription factor GLABRA 3-like ...	498	6e-168
TR:A0A9P0Z4G9	A0A9P0Z4G9_CUSEU	BHLH domain-containing protein OS=C...	500	8e-168
TR:A0A484KIG3	A0A484KIG3_9ASTE	BHLH domain-containing protein OS=C...	501	1e-167
TR:A0A8B8Q1S2	A0A8B8Q1S2_9MYRT	Transcription factor EGL1-like isof...	501	2e-167
TR:A0A7J8XG93	A0A7J8XG93_GOSAI	Uncharacterized protein OS=Gossypiu...	500	2e-167
TR:A0A078H9M5	A0A078H9M5_BRANA	(rape) hypothetical protein OS=Bras...	499	5e-167
TR:A0A6A1VNT1	A0A6A1VNT1_9ROSI	Transcription factor GLABRA 3 OS=Mo...	501	7e-167
TR:B9GQ94	B9GQ94_POPTR	BHLH family protein OS=Populus trichocarpa ...	500	8e-167
TR:B9I8J2	B9I8J2_POPTR	Uncharacterized protein OS=Populus trichoca...	500	1e-166
TR:A0A7J8ZUG6	A0A7J8ZUG6_9ROSI	Uncharacterized protein (Fragment) ...	498	2e-166
TR:A0AAJ6U559	A0AAJ6U559_POPEU	Transcription factor EGL1-like isof...	499	2e-166
TR:A0AAD4S8J0	A0AAD4S8J0_9MAGN	BHLH domain-containing protein OS=P...	499	4e-166
TR:A0A7J7P6Q0	A0A7J7P6Q0_9MAGN	BHLH domain-containing protein OS=K...	498	5e-166
TR:A0A445DTS7	A0A445DTS7_ARAHY	BHLH domain-containing protein OS=A...	495	8e-166
TR:A0A679KB97	A0A679KB97_BRACM	BHLH domain-containing protein OS=B...	496	8e-166
TR:A0A7J0FRN6	A0A7J0FRN6_9ERIC	Basic helix-loop-helix (BHLH) DNA-b...	493	1e-165
TR:A0A565AYM7	A0A565AYM7_9BRAS	BHLH domain-containing protein OS=A...	496	1e-165
TR:A0A8B9A1B0	A0A8B9A1B0_PHODC	Transcription factor GLABRA 3 OS=Ph...	498	3e-165
TR:A0A9Q1GK13	A0A9Q1GK13_9CARY	BHLH domain-containing protein OS=C...	495	3e-165
TR:A0A8X7R0X5	A0A8X7R0X5_BRACI	BHLH domain-containing protein OS=B...	494	3e-165
TR:A0A834ZT73	A0A834ZT73_TETSI	BHLH domain-containing protein OS=T...	497	3e-165
TR:A0AAD8LY53	A0AAD8LY53_9APIA	Transcription factor EGL1-like OS=H...	493	4e-165
TR:A0A8X7SC39	A0A8X7SC39_BRACI	BHLH domain-containing protein OS=B...	494	4e-165
SP:Q9CAD0	EGL1_ARATH	Transcription factor EGL1 OS=Arabidopsis thal...	493	7e-165
TR:D7KTF4	D7KTF4_ARALL	BHLH domain-containing protein OS=Arabidops...	493	7e-165
TR:A0A2P6SHC4	A0A2P6SHC4_ROSCH	Putative transcription factor bHLH ...	495	9e-165
TR:A0A8K0E326	A0A8K0E326_9ROSA	Uncharacterized protein OS=Rhamnelli...	494	2e-164
TR:A0A6D2KLA8	A0A6D2KLA8_9BRAS	BHLH domain-containing protein OS=M...	493	2e-164
TR:A0A8B8PV48	A0A8B8PV48_9MYRT	Transcription factor EGL1-like isof...	493	3e-164
TR:A0A6J0KZP3	A0A6J0KZP3_RAPSA	Transcription factor EGL1 OS=Raphan...	492	4e-164
TR:V4JUJ6	V4JUJ6_EUTSA	BHLH domain-containing protein OS=Eutrema s...	492	5e-164
TR:A0A8X7SAT3	A0A8X7SAT3_BRACI	BHLH domain-containing protein OS=B...	491	6e-164
TR:A0A6J0KZK8	A0A6J0KZK8_RAPSA	Transcription factor EGL1 OS=Raphan...	490	1e-163
TR:R0GDR6	R0GDR6_9BRAS	BHLH domain-containing protein (Fragment) O...	491	2e-163
TR:A0A328E2R3	A0A328E2R3_9ASTE	BHLH domain-containing protein OS=C...	491	2e-163
TR:A0A4Y7JAE1	A0A4Y7JAE1_PAPSO	BHLH domain-containing protein OS=P...	491	4e-163
TR:A0A9W3CEX8	A0A9W3CEX8_RAPSA	Transcription factor EGL1-like OS=R...	489	6e-163
TR:A0A1J7GZK9	A0A1J7GZK9_LUPAN	BHLH domain-containing protein OS=L...	490	1e-162
TR:A0A6A2WVT7	A0A6A2WVT7_HIBSY	Transcription factor MYC1 OS=Hibisc...	489	1e-162
TR:A0A5N5K976	A0A5N5K976_9ROSI	Uncharacterized protein OS=Salix br...	489	2e-162
TR:A0A7J8RU19	A0A7J8RU19_GOSDV	Uncharacterized protein OS=Gossypiu...	489	3e-162
TR:A0A8B8PTT1	A0A8B8PTT1_9MYRT	Transcription factor EGL1-like isof...	488	3e-162
TR:A0A8X7YN83	A0A8X7YN83_POPTO	Transcription factor MYC/MYB N-term...	488	4e-162
TR:A0A067K9Y9	A0A067K9Y9_JATCU	BHLH domain-containing protein OS=J...	486	9e-162
TR:A0A8T2ZH99	A0A8T2ZH99_POPDE	BHLH domain-containing protein OS=P...	487	1e-161
TR:A0A8T2X500	A0A8T2X500_POPDE	Transcription factor MYC/MYB N-term...	487	1e-161

TR:A0A498JJ01	A0A498JJ01_MALDO	BHLH domain-containing protein OS=M...	486	2e-161
TR:A0A2Z6LN10	A0A2Z6LN10_TRISU	BHLH domain-containing protein OS=T...	484	2e-161
TR:A0A9Q0TEE6	A0A9Q0TEE6_9ROSI	TRANSCRIPTION FACTOR MYC1 OS=Salix ...	486	3e-161
TR:A0A078GLN8	A0A078GLN8_BRANA	BnaC09g11380D protein OS=Brassica n...	483	4e-161
TR:A0A6P4A4B8	A0A6P4A4B8_ZIZJJ	Transcription factor EGL1-like OS=Z...	485	4e-161
TR:A0AAJ6XLR4	A0AAJ6XLR4_POPEU	Transcription factor EGL1-like isof...	486	8e-161
TR:A0A4U5QU66	A0A4U5QU66_POPAL	Basic helix-loop-helix regulatory f...	484	1e-160
TR:A0A835N824	A0A835N824_9ROSI	BHLH domain-containing protein OS=S...	484	3e-160
TR:A0AAJ6UVU2	A0AAJ6UVU2_POPEU	Transcription factor EGL1-like OS=P...	483	3e-160
TR:A0A3P5YF86	A0A3P5YF86_BRACM	BHLH domain-containing protein OS=B...	481	4e-160
TR:A0A6A3A0N7	A0A6A3A0N7_HIBSY	Transcription factor MYC1 OS=Hibisc...	494	6e-160
TR:A0A4V6A9P1	A0A4V6A9P1_POPAL	Basic helix-loop-helix regulatory f...	481	2e-159
TR:A0A7J0G7T0	A0A7J0G7T0_9ERIC	Basic helix-loop-helix (BHLH) DNA-b...	479	2e-159
TR:A0A078F5I3	A0A078F5I3_BRANA	BnaC09g12820D protein OS=Brassica n...	479	3e-159
TR:A0A9W7IKP3	A0A9W7IKP3_HIBTR	GLABRA 3, GLABROUS 3 OS=Hibiscus tr...	479	5e-159
TR:A0A9P0ZGB2	A0A9P0ZGB2_CUSEU	BHLH domain-containing protein OS=C...	478	7e-159
TR:B9RDS0	B9RDS0_RICCO	Transcription factor, putative OS=Ricinus c...	479	2e-158
TR:A0A6I9QKZ2	A0A6I9QKZ2_ELAGV	Anthocyanin regulatory Lc protein i...	479	2e-158
TR:A0A5N5G836	A0A5N5G836_9ROSA	Transcription factor EGL1-like OS=P...	478	3e-158
TR:A0A5N5NQ63	A0A5N5NQ63_9ROSI	BHLH domain-containing protein OS=S...	478	7e-158
TR:A0A8X7UG27	A0A8X7UG27_BRACI	BHLH domain-containing protein OS=B...	475	1e-157
TR:A0A0D3E4K4	A0A0D3E4K4_BRAOL	BHLH domain-containing protein OS=B...	475	1e-157
TR:A0A8X8ZKM6	A0A8X8ZKM6_SALSN	BHLH domain-containing protein OS=S...	475	1e-157
TR:A0AAD3Y0C6	A0AAD3Y0C6_NEPGR	BHLH domain-containing protein OS=N...	476	1e-157
TR:A0A4Y7KKH2	A0A4Y7KKH2_PAPSO	BHLH domain-containing protein OS=P...	478	2e-157
TR:A0A540N6T8	A0A540N6T8_MALBA	BHLH domain-containing protein OS=M...	477	3e-157
TR:A0AAD3P8P8	A0AAD3P8P8_NEPGR	Uncharacterized protein OS=Nepenthe...	475	4e-157
TR:A0A078HPG4	A0A078HPG4_BRANA	BnaA04g11060D protein OS=Brassica n...	472	4e-157
TR:A0A8K0IM21	A0A8K0IM21_COCNU	Anthocyanin regulatory Lc protein O...	476	5e-157
TR:V4L8S4	V4L8S4_EUTSA	BHLH domain-containing protein OS=Eutrema s...	471	9e-157
TR:A0A8X8CCN6	A0A8X8CCN6_POPTO	Transcription factor MYC/MYB N-term...	475	1e-156
TR:A0A6P4DBH2	A0A6P4DBH2_ARADU	Transcription factor EGL1-like OS=A...	471	4e-156
TR:A0A6P4A5D5	A0A6P4A5D5_ZIZJJ	Transcription factor EGL1-like isof...	472	5e-156
TR:A0A8T1NXJ6	A0A8T1NXJ6_CARIL	BHLH domain-containing protein OS=C...	468	7e-156
TR:A0A8X8DCH2	A0A8X8DCH2_POPTO	BHLH domain-containing protein OS=P...	473	1e-155
TR:A0A978V5H0	A0A978V5H0_ZIZJJ	BHLH domain-containing protein OS=Z...	471	1e-155
TR:A0A078J763	A0A078J763_BRANA	BnaCnng33960D protein OS=Brassica n...	467	9e-155
TR:A0A540N6U2	A0A540N6U2_MALBA	BHLH domain-containing protein OS=M...	469	5e-154
TR:A0A9Q0Z667	A0A9Q0Z667_9ROSI	TRANSCRIPTION FACTOR TT8 OS=Salix k...	466	1e-153
TR:A0A835CA66	A0A835CA66_9FABA	Transcription factor GLABRA 3 OS=Se...	471	1e-153
TR:W9RSW9	W9RSW9_9ROSA	Uncharacterized protein OS=Morus notabilis ...	466	3e-153
TR:A0A9P0Z4R2	A0A9P0Z4R2_CUSEU	BHLH domain-containing protein OS=C...	462	3e-153
TR:A0A067HE87	A0A067HE87_CITSI	BHLH domain-containing protein OS=C...	463	7e-153
TR:A0A498KNS2	A0A498KNS2_MALDO	BHLH domain-containing protein OS=M...	464	8e-153
TR:A0A445D674	A0A445D674_ARAHY	BHLH domain-containing protein OS=A...	462	9e-153
TR:A0A251KST3	A0A251KST3_MANES	Uncharacterized protein OS=Manihot ...	461	6e-152
TR:A0A804K5C8	A0A804K5C8_MUSAM	(wild Malaysian banana) hypotherica...	461	1e-151
TR:A0AA88QVJ0	A0AA88QVJ0_9ASTE	BHLH domain-containing protein (Fra...	461	1e-151
TR:A0A2I0WX50	A0A2I0WX50_9ASPA	Transcription factor GLABRA 3 OS=De...	461	2e-151
TR:A0AAD7L426	A0AAD7L426_QUISA	Basic helix-loop-helix transcriptio...	461	2e-151
TR:A0A9Q0Z6H4	A0A9Q0Z6H4_9ROSI	TRANSCRIPTION FACTOR TT8 OS=Salix k...	459	8e-151
TR:A0A2P5F2R5	A0A2P5F2R5_TREOI	Basic helix-loop-helix transcriptio...	458	2e-150
TR:A0A9E7HTA9	A0A9E7HTA9_9LILI	Anthocyanin regulatory Lc protein O...	457	6e-150
TR:A0A8T3AC79	A0A8T3AC79_DENNO	BHLH domain-containing protein OS=D...	456	2e-149
TR:A0A6J1DTS9	A0A6J1DTS9_MOMCH	Transcription factor EGL1-like isof...	456	2e-149

TR:A0AAD6RGA7	A0AAD6RGA7_9ROSI	Basic helix-loop-helix regulatory f...	455	1e-148
TR:A0A5P1F092	A0A5P1F092_ASPOF	BHLH domain-containing protein OS=A...	454	1e-148
TR:A0A443PLU8	A0A443PLU8_9MAGN	Transcription factor GLABRA 3-like ...	453	2e-148
TR:A0A4S8K2V0	A0A4S8K2V0_MUSBA	BHLH domain-containing protein OS=M...	454	2e-148
TR:A0A6J1DXY7	A0A6J1DXY7_MOMCH	Transcription factor EGL1-like isof...	452	4e-148
TR:A0A0D2SAP3	A0A0D2SAP3_GOSRA	Transcription factor MYC/MYB N-term...	449	4e-148
TR:A0AA88V1G3	A0AA88V1G3_9ASTE	BHLH domain-containing protein (Fra...	451	1e-147
TR:A0A835MPT9	A0A835MPT9_9ROSI	Transcription factor MYC/MYB N-term...	452	1e-147
TR:A0AA87ZW66	A0AA87ZW66_FICCA	BHLH domain-containing protein OS=F...	450	2e-147
TR:A0A6J1DV54	A0A6J1DV54_MOMCH	Transcription factor EGL1-like isof...	450	2e-147
TR:A0A9Q0Z6A5	A0A9Q0Z6A5_9ROSI	TRANSCRIPTION FACTOR MYC1 OS=Salix ...	451	2e-147
TR:A0A2K3N2J9	A0A2K3N2J9_TRIPR	Transcription factor GLABRA 3-like ...	445	1e-146
TR:A0A7J6V7H9	A0A7J6V7H9_THATH	Basic helix-loop-helix transcriptio...	445	2e-146
TR:A0AA38Z3A7	A0AA38Z3A7_VITRO	Transcription factor MYC/MYB N-term...	442	4e-146
TR:A0A022Q465	A0A022Q465_ERYGU	BHLH domain-containing protein (Fra...	438	5e-146
TR:A0AAD7L4B3	A0AAD7L4B3_QUISA	Basic helix-loop-helix transcriptio...	446	6e-146
TR:A0A9D5HBH7	A0A9D5HBH7_9LILI	BHLH domain-containing protein OS=D...	444	1e-145
TR:A0AAD7LQK8	A0AAD7LQK8_QUISA	Basic helix-loop-helix transcriptio...	443	4e-145
TR:A0A9Q0PW20	A0A9Q0PW20_9ROSI	TRANSCRIPTION FACTOR EGL1 OS=Salix ...	440	5e-145
TR:A0A5D2X4M7	A0A5D2X4M7_GOSMU	Transcription factor MYC/MYB N-term...	439	1e-144
TR:A0A6J1KY06	A0A6J1KY06_CUCMA	Transcription factor EGL1-like isof...	439	4e-143
TR:A0A7J8LQD6	A0A7J8LQD6_9ROSI	BHLH domain-containing protein (Fra...	433	1e-142
TR:A0A1S3C7Z7	A0A1S3C7Z7_CUCME	LOW QUALITY PROTEIN: transcription ...	437	3e-142
TR:A0A1R3I888	A0A1R3I888_COCAP	Uncharacterized protein OS=Corchoru...	435	6e-142
TR:A0A9E7HLQ6	A0A9E7HLQ6_9LILI	Anthocyanin regulatory Lc protein O...	442	2e-141
TR:A0AAD6P3A9	A0AAD6P3A9_9ROSI	BHLH domain-containing protein OS=S...	429	4e-141
TR:A0A2P5B2K7	A0A2P5B2K7_PARAD	Basic helix-loop-helix transcriptio...	433	9e-141
TR:A0AAV8Q5J4	A0AAV8Q5J4_ENSVE	Uncharacterized protein OS=Ensete v...	436	2e-140
TR:A0A8B8JZL1	A0A8B8JZL1_ABRPR	Transcription factor EGL1-like OS=A...	430	7e-140
TR:A0AAE1JD53	A0AAE1JD53_9FABA	BHLH domain-containing protein OS=A...	430	8e-140
TR:W1NGV1	W1NGV1_AMBTC	BHLH domain-containing protein OS=Amborella...	431	9e-140
TR:A0A7N0VAP0	A0A7N0VAP0_KALFE	Transcription factor MYC/MYB N-term...	429	1e-139
TR:I1KIA8	I1KIA8_SOYBN	Uncharacterized protein OS=Glycine max OX=3...	430	2e-139
TR:A0A445ABX2	A0A445ABX2_ARAHY	Uncharacterized protein OS=Arachis ...	430	3e-139
TR:A0A9D4VHE8	A0A9D4VHE8_PEA	BHLH domain-containing protein OS=Pis...	422	3e-139
TR:A0A6P4BWJ1	A0A6P4BWJ1_ARADU	Transcription factor EGL1 isoform X...	429	6e-139
TR:A0A6P4BT03	A0A6P4BT03_ARADU	Transcription factor EGL1 isoform X...	429	6e-139
TR:A0A445EIY0	A0A445EIY0_ARAHY	Uncharacterized protein OS=Arachis ...	429	6e-139
TR:A0A445JTI9	A0A445JTI9_GLYSO	Transcription factor EGL1 OS=Glycin...	428	9e-139
TR:A0A9E7HU58	A0A9E7HU58_9LILI	Anthocyanin regulatory Lc protein O...	442	1e-138
TR:A0A9E7HKZ2	A0A9E7HKZ2_9LILI	Anthocyanin regulatory Lc protein O...	442	2e-138
TR:A0A9E7HJG2	A0A9E7HJG2_9LILI	Anthocyanin regulatory Lc protein O...	442	6e-138
TR:A0A3Q7YG33	A0A3Q7YG33_CICAR	Transcription factor GLABRA 3 isofo...	421	7e-138
TR:V4TD94	V4TD94_CITCL	BHLH domain-containing protein OS=Citrus cl...	423	1e-137
TR:A0A9E7HSG1	A0A9E7HSG1_9LILI	Anthocyanin regulatory Lc protein O...	442	2e-137
TR:A0A9Q0HXP5	A0A9Q0HXP5_9POAL	BHLH domain-containing protein OS=R...	431	5e-136
TR:A0A804L3X0	A0A804L3X0_MUSAM	(wild Malaysian banana) hypotetica...	420	1e-135
TR:A0AAD6P404	A0AAD6P404_9ROSI	BHLH domain-containing protein OS=S...	415	1e-135
TR:A0A2H9ZY71	A0A2H9ZY71_9ASPA	Transcription factor EGL1 OS=Aposta...	419	1e-135
TR:A0A8S0R7W2	A0A8S0R7W2_OLLEU	Transcription factor GLABRA 3-like ...	412	7e-135
TR:A0A833RJE6	A0A833RJE6_9POAL	Transcription factor GLABRA 3 OS=Ca...	417	7e-135
TR:A0A445L6M0	A0A445L6M0_GLYSO	Transcription factor GLABRA 3 OS=Gl...	417	1e-134
TR:I1JK86	I1JK86_SOYBN	Uncharacterized protein OS=Glycine max OX=3...	417	1e-134
TR:A0AAE1JDQ1	A0AAE1JDQ1_9FABA	BHLH domain-containing protein OS=A...	416	2e-134
TR:A0A0R0J0F5	A0A0R0J0F5_SOYBN	Uncharacterized protein OS=Glycine ...	416	4e-134

TR:A0A445EIY4	A0A445EIY4_ARAHY	Uncharacterized protein OS=Arachis ...	415	2e-133
TR:A0A9R1W2L0	A0A9R1W2L0_LACSA	BHLH domain-containing protein OS=L...	411	2e-133
TR:A0A5J4ZU50	A0A5J4ZU50_9ASTE	Transcription factor MYC/MYB N-term...	412	4e-133
TR:A0A8T3C4L9	A0A8T3C4L9_DENNO	BHLH domain-containing protein OS=D...	413	9e-133
TR:V4VXZ2	V4VXZ2_CITCL	BHLH domain-containing protein OS=Citrus cl...	410	1e-132
TR:A0A2P5XKZ4	A0A2P5XKZ4_GOSBA	Uncharacterized protein OS=Gossypiu...	411	1e-132
TR:A0AA86VG48	A0AA86VG48_9FABA	Uncharacterized protein OS=Sphenost...	411	3e-132
TR:A0A9E7EJR6	A0A9E7EJR6_9LILI	BHLH domain-containing protein OS=M...	411	4e-132
TR:A0A0S3SYH9	A0A0S3SYH9_PHAAN	Uncharacterized protein OS=Vigna an...	408	7e-131
TR:A0AA35YI36	A0AA35YI36_LACSI	BHLH domain-containing protein OS=L...	402	4e-130
TR:V7AR13	V7AR13_PHAVU	Uncharacterized protein OS=Phaseolus vulgar...	406	5e-130
TR:A0A2P5XIM6	A0A2P5XIM6_GOSBA	BHLH domain-containing protein OS=G...	400	7e-130
TR:A0A0R0KDC8	A0A0R0KDC8_SOYBN	Uncharacterized protein OS=Glycine ...	404	1e-129
TR:A0AA38T3Q5	A0AA38T3Q5_9ASTR	BHLH domain-containing protein OS=C...	401	1e-129
TR:A0A0R0K5G4	A0A0R0K5G4_SOYBN	BHLH domain-containing protein OS=G...	399	4e-129
TR:A0A1S3V9R1	A0A1S3V9R1_VIGRR	Transcription factor EGL1 isoform X...	401	2e-128
TR:V7ANB6	V7ANB6_PHAVU	Uncharacterized protein OS=Phaseolus vulgar...	401	3e-128
TR:A0AAV8RX58	A0AAV8RX58_ENSVE	Uncharacterized protein OS=Ensete v...	400	9e-128
TR:A0A2G2XJQ5	A0A2G2XJQ5_CAPBA	BHLH domain-containing protein OS=C...	393	2e-127
TR:A0A8J5KPG5	A0A8J5KPG5_ZINOF	BHLH domain-containing protein OS=Z...	401	3e-127
TR:A0AA39E1U8	A0AA39E1U8_VITRO	BHLH domain-containing protein OS=V...	392	3e-127
TR:A0A1S3VA45	A0A1S3VA45_VIGRR	Transcription factor EGL1 isoform X...	397	1e-126
TR:A0A061GCJ5	A0A061GCJ5_THECC	Basic helix-loop-helix DNA-binding ...	390	2e-126
TR:A0A8S0RAX6	A0A8S0RAX6_OLEEU	Transcription factor GLABRA 3-like ...	389	4e-126
TR:A0A199VG1	A0A199VG1_ANACO	Transcription factor GLABRA 3 (Frag...	394	1e-125
TR:A0A0L9TBS9	A0A0L9TBS9_PHAAN	Uncharacterized protein OS=Phaseolu...	394	2e-125
TR:A0A540KLC3	A0A540KLC3_MALBA	Transcription factor MYC/MYB N-term...	384	3e-125
TR:A0A804L9P0	A0A804L9P0_MUSAM	(wild Malaysian banana) hypothetica...	392	6e-125
TR:A0A1J7FN55	A0A1J7FN55_LUPAN	BHLH domain-containing protein OS=L...	387	7e-124
TR:A0A4S8J5C7	A0A4S8J5C7_MUSBA	BHLH domain-containing protein OS=M...	390	7e-124
TR:A0A5E4GAD2	A0A5E4GAD2_PRUDU	PREDICTED: mRNAion factor OS=Prunus...	382	2e-123
TR:A0A2C9WP16	A0A2C9WP16_MANES	Transcription factor MYC/MYB N-term...	383	2e-122
TR:A0A9D4VGX8	A0A9D4VGX8_PEA	BHLH domain-containing protein OS=Pis...	380	3e-122
TR:A0A251P6A8	A0A251P6A8_PRUPE	BHLH domain-containing protein OS=P...	379	3e-122
TR:A0A8J5GK07	A0A8J5GK07_ZINOF	BHLH domain-containing protein OS=Z...	384	2e-121
TR:A0AA41SF95	A0AA41SF95_PAPNU	BHLH domain-containing protein OS=P...	381	3e-121
TR:A0A8J5L942	A0A8J5L942_ZINOF	BHLH domain-containing protein OS=Z...	385	5e-121
TR:A0A8S0R6W4	A0A8S0R6W4_OLEEU	Transcription factor GLABRA 3-like ...	375	5e-120
TR:A0A9Q1QQA7	A0A9Q1QQA7_9CARY	Uncharacterized protein OS=Carnegie...	374	4e-118
TR:A0A8T2Z761	A0A8T2Z761_POPDE	BHLH domain-containing protein OS=P...	367	2e-117
TR:A0AAD6R978	A0AAD6R978_9ROSI	Transcription factor GLABRA 3-like ...	361	3e-115
TR:A0A9Q0PW10	A0A9Q0PW10_9ROSI	TRANSCRIPTION FACTOR EGL1 OS=Salix ...	359	2e-114
TR:A0A8J5KHU9	A0A8J5KHU9_ZINOF	BHLH domain-containing protein OS=Z...	365	2e-114
TR:A0A072TV58	A0A072TV58_MEDTR	BHLH transcription factor OS=Medica...	357	4e-113
TR:A0A7N0SZ75	A0A7N0SZ75_KALFE	Uncharacterized protein OS=Kalancho...	361	4e-113
TR:A0A0D2P6G7	A0A0D2P6G7_GOSRA	Uncharacterized protein OS=Gossypiu...	351	1e-110
TR:A0AAE1UV76	A0AAE1UV76_9SOLA	BHLH domain-containing protein OS=A...	351	4e-110
TR:A0A2U1LX35	A0A2U1LX35_ARTAN	Basic helix-loop-helix (BHLH) DNA-b...	349	5e-110
TR:A0A9Q0PDF7	A0A9Q0PDF7_SALPP	TRANSCRIPTION FACTOR TT8 OS=Salix p...	348	2e-109
TR:A0A8B8MC40	A0A8B8MC40_ABRPR	Basic helix-loop-helix protein A-li...	351	5e-109
TR:A0A540LHM0	A0A540LHM0_MALBA	BHLH domain-containing protein OS=M...	338	4e-107
TR:A0A0D3C0F3	A0A0D3C0F3_BRAOL	BHLH domain-containing protein OS=B...	341	5e-107
TR:A0A9W3CWX1	A0A9W3CWX1_RAPSA	Transcription factor EGL1-like isof...	335	1e-106
TR:B9RCC5	B9RCC5_RICCO	Transcription factor, putative OS=Ricinus c...	338	9e-106
TR:A0AA38G6A8	A0AA38G6A8_TAXCH	BHLH domain-containing protein (Fra...	343	1e-105

2025-01-03 14:40 EST

The job with ID ncbiblast-R20250103-194143-0666-71816077-plm has been submitted with these raw input values:

program

Unset

blastp

matrix

Unset

BLOSUM62

alignments

Unset

500

scores

Unset

500

exp

Unset

10

filter

Unset

F

gapalign

Unset

true

compstats

Unset

F

align

Unset

0

stype

Unset

protein

sequence

Unset

```
>QGV13005.1 roseal [Plant binary expression vector pDEL.ROS]  
MEKNCRGVRKGTWTKEEDTLLRQCIEEYGEGKWHQVPHRAGLNRCKRSCRLRWLNILRPN  
IKRGFRSRDEVDLIVRLHKLLGNKWSLIAGRIPGRTANDVKNFWNTHVGKNLGEDGERCR  
KNVMNTKTIKLTNIVRPRARTFTGLHVTWPREVSKTDEFSNVRLTTDEIPDCEKQTQFYN  
DVASPDQDEVEDCIQWWSKLETTEDGELGNLFEEAQQIGN
```

database

Unset

uniprotkb_refprotswissprot

BLASTP 2.16.0+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Reference for composition-based statistics: Alejandro A. Schaffer, L. Aravind, Thomas L. Madden, Sergei Shavirin, John L. Spouge, Yuri I. Wolf, Eugene V. Koonin, and Stephen F. Altschul (2001), "Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements", Nucleic Acids Res. 29:2994-3005.

Database: uniprotkb_refprotswissprot
91,211,443 sequences; 36,979,232,029 total letters

Query= QGV13005.1 roseal [Plant binary expression vector pDEL.ROS]

Length=220

Sequences producing significant alignments:	Score (Bits)	E Value
TR:A0AAE1XR31 A0AAE1XR31_9LAMI Transcription factor OS=Sesamum ala...	250	1e-80
TR:A0AAE2CLG5 A0AAE2CLG5_9LAMI Transcription factor OS=Sesamum ala...	249	2e-80
TR:A0A6I9SSW5 A0A6I9SSW5_SESIN Transcription factor MYB114 OS=Sesa...	248	2e-80
TR:A0A6I9UE42 A0A6I9UE42_SESIN Transcription factor MYB1 OS=Sesamu...	248	6e-79
TR:A0A2G9H3L0 A0A2G9H3L0_9LAMI Transcription factor, Myb superfami...	245	8e-79
TR:A0A2G9HVP6 A0A2G9HVP6_9LAMI Transcription factor, Myb superfami...	244	2e-78
TR:A0A2G9H2Q5 A0A2G9H2Q5_9LAMI Transcription factor, Myb superfami...	244	2e-78
TR:A0AAE1XRW8 A0AAE1XRW8_9LAMI Transcription factor OS=Sesamum ala...	242	1e-77
TR:A0A6I9UFG5 A0A6I9UFG5_SESIN Transcription factor MYB114-like OS...	240	4e-77
TR:A0A5A7PDP0 A0A5A7PDP0_STRAF MYB transcription factor OS=Striga ...	239	3e-76
TR:A0A6I9UNX3 A0A6I9UNX3_SESIN Transcription factor MYB90-like OS=...	238	7e-76
TR:A0AAD4ILA8 A0AAD4ILA8_PERFH Myb domain protein 113 OS=Perilla f...	236	3e-75
TR:A0A2G9FY53 A0A2G9FY53_9LAMI Transcription factor, Myb superfami...	234	7e-75
TR:A0A9N7RNY7 A0A9N7RNY7_STRHE Transcription factor MYB113 OS=Stri...	234	1e-74
TR:A0AAE1YA88 A0AAE1YA88_9LAMI Transcription factor OS=Sesamum ala...	230	4e-73
TR:A0A830C5A5 A0A830C5A5_9LAMI Transcription factor myb113 OS=Phth...	232	4e-73
TR:A0AAE1T6N1 A0AAE1T6N1_9LAMI Transcription factor OS=Sesamum ang...	227	6e-72
TR:A0A2G9H3M6 A0A2G9H3M6_9LAMI Transcription factor, Myb superfami...	224	2e-71
TR:A0A6I9SU10 A0A6I9SU10_SESIN Transcription factor MYB114-like OS...	223	2e-70

TR:A0A022RMP0	A0A022RMP0_ERYGU	Uncharacterized protein OS=Erythran...	223	7e-70
TR:A0A8X8YU62	A0A8X8YU62_SALSN	Myb proto-oncogene protein, plant O...	219	2e-69
TR:A0A5A7PCR9	A0A5A7PCR9_STRAF	Myb transcription factor OS=Striga ...	220	4e-69
TR:A0A022RX70	A0A022RX70_ERYGU	Uncharacterized protein OS=Erythran...	221	4e-69
TR:B9T6Z2	B9T6Z2_RICCO	R2r3-myb transcription factor, putative (Fr...	219	1e-68
TR:A0A8S0Q974	A0A8S0Q974_OLEEU	Transcription factor MYB114-like OS...	217	3e-68
TR:A0A8S0SLB7	A0A8S0SLB7_OLEEU	Uncharacterized protein OS=Olea eur...	216	7e-68
TR:A0A8S0SH61	A0A8S0SH61_OLEEU	Anthocyanin 2 OS=Olea europaea subs...	214	3e-67
TR:A0A1S3ZX05	A0A1S3ZX05_TOBAC	Transcription factor MYB114-like OS...	213	9e-67
TR:A0AAE1SWQ3	A0AAE1SWQ3_9SOLA	MYB transcription factor AN2 OS=Ani...	214	1e-66
TR:A0A8S0UQC2	A0A8S0UQC2_OLEEU	Transcription factor MYB113-like OS...	213	1e-66
TR:A0A830B1Q3	A0A830B1Q3_9LAMI	Transcription factor myb113 OS=Phth...	214	1e-66
TR:U5GI57	U5GI57_POPTR	MYB family protein OS=Populus trichocarpa O...	214	1e-66
TR:A0A9Q1MXT0	A0A9Q1MXT0_9SOLA	Uncharacterized protein OS=Anisodus...	214	1e-66
TR:A0AAD2A990	A0AAD2A990_9LAMI	Myb-like domain-containing protein ...	212	3e-66
TR:A0A6P6TM05	A0A6P6TM05_COFAR	Transcription factor MYB113-like OS...	214	4e-66
TR:A0A087G046	A0A087G046_ARAAL	Uncharacterized protein OS=Arabis a...	213	4e-66
TR:A0A0D2ZRG6	A0A0D2ZRG6_BRAOL	Uncharacterized protein OS=Brassica...	212	9e-66
TR:A0A078IZV4	A0A078IZV4_BRANA	(rape) hypothetical protein OS=Bras...	212	9e-66
TR:A0AA38Z5B0	A0AA38Z5B0_VITRO	Uncharacterized protein OS=Vitis ro...	211	1e-65
TR:A0A8K0MH20	A0A8K0MH20_9ROSA	MYB transcription factor OS=Rhamnel...	212	1e-65
TR:A0A1J6HV32	A0A1J6HV32_NICAT	Transcription factor myb113 OS=Nico...	211	2e-65
TR:A0A834LD45	A0A834LD45_RHOSS	Uncharacterized protein OS=Rhododen...	216	2e-65
TR:A0A5J5BYJ6	A0A5J5BYJ6_9ASTE	Uncharacterized protein OS=Nyssa si...	211	2e-65
TR:V4KC52	V4KC52_EUTSA	Uncharacterized protein OS=Eutrema salsugin...	211	3e-65
TR:A0A061ENQ7	A0A061ENQ7_THECC	R2R3 MYB transcription factor splic...	211	3e-65
TR:A0A022RX75	A0A022RX75_ERYGU	Uncharacterized protein (Fragment) ...	210	5e-65
TR:D7KTP0	D7KTP0_ARALL	Uncharacterized protein OS=Arabidopsis lyra...	210	5e-65
TR:A0AAJ6U311	A0AAJ6U311_POPEU	Transcription factor MYB113-like OS...	210	6e-65
TR:A0A1J6JSB4	A0A1J6JSB4_NICAT	Transcription factor myb90 OS=Nicot...	210	7e-65
TR:A0A565B1A5	A0A565B1A5_9BRAS	Uncharacterized protein OS=Arabis n...	209	7e-65
TR:A0AAD2A6N3	A0AAD2A6N3_9LAMI	Uncharacterized protein OS=Fraxinus...	208	7e-65
TR:A0A8K0H462	A0A8K0H462_9ROSA	MYB transcription factor OS=Rhamnel...	210	8e-65
TR:D7KTP5	D7KTP5_ARALL	MYB transcription factor OS=Arabidopsis lyr...	209	1e-64
TR:A0A6P6TG94	A0A6P6TG94_COFAR	Transcription factor MYB90-like OS=...	210	1e-64
TR:A0A4U5NNM9	A0A4U5NNM9_POPAL	Transcription factor MYB90-like OS=...	210	1e-64
TR:A0A6J0K655	A0A6J0K655_RAPSA	Transcription factor MYB114 OS=Raph...	209	1e-64
TR:A0A7N1A4H8	A0A7N1A4H8_KALFE	Uncharacterized protein OS=Kalancho...	210	2e-64
TR:A0A8X7UKK6	A0A8X7UKK6_BRACI	Uncharacterized protein OS=Brassica...	208	2e-64
TR:A0A8X8B572	A0A8X8B572_BRACI	Uncharacterized protein OS=Brassica...	208	2e-64
TR:M4FF91	M4FF91_BRACM	Anthocyanin R2R3-MYB transcription factor O...	208	2e-64
TR:A0A8T0CNB5	A0A8T0CNB5_CORYI	MYB transcription factor OS=Corymbi...	207	2e-64
TR:A0AA39DF60	A0AA39DF60_VITRO	Uncharacterized protein OS=Vitis ro...	208	2e-64
TR:A0AAD6KEL2	A0AAD6KEL2_9ROSI	Uncharacterized protein OS=Salix ud...	208	3e-64
TR:A0A1Q3D3X1	A0A1Q3D3X1_CEPFO	Myb_DNA-binding domain-containing p...	207	4e-64
TR:A0AAD3XH62	A0AAD3XH62_NEPGR	Uncharacterized protein OS=Nepenthe...	210	4e-64
TR:A0A0D3CYD9	A0A0D3CYD9_BRAOL	Uncharacterized protein OS=Brassica...	207	5e-64
TR:M4CCI7	M4CCI7_BRACM	Uncharacterized protein OS=Brassica campest...	207	6e-64
TR:A0A078J656	A0A078J656_BRANA	(rape) hypothetical protein OS=Bras...	207	6e-64
TR:A0A2Z7CPC7	A0A2Z7CPC7_9LAMI	Uncharacterized protein OS=Dorcocer...	207	6e-64
TR:A0A7N0SZR0	A0A7N0SZR0_KALFE	Uncharacterized protein OS=Kalancho...	208	6e-64
TR:A0A6D2IVR5	A0A6D2IVR5_9BRAS	Uncharacterized protein OS=Microthl...	207	6e-64
TR:A0A8X7PKB9	A0A8X7PKB9_BRACI	Uncharacterized protein OS=Brassica...	207	8e-64
TR:A0A6J0M9U3	A0A6J0M9U3_RAPSA	Transcription factor MYB114 OS=Raph...	207	8e-64
TR:A0A3S3MUE8	A0A3S3MUE8_9MAGN	Transcription factor MYB1 OS=Cinnam...	207	9e-64

SP:Q9FNV8 MY114_ARATH Transcription factor MYB114 OS=Arabidopsis t...	202	1e-63
TR:A0A6J0K8L2 A0A6J0K8L2_RAPSA Transcription factor MYB114 OS=Raph...	206	2e-63
TR:B8YPF8 B8YPF8_VITVI R2R3 MYBA6 transcription factor splice vari...	206	2e-63
TR:A0A9Q1MXN1 A0A9Q1MXN1_9SOLA Uncharacterized protein OS=Anisodus...	206	2e-63
TR:A0A1Q3AXP7 A0A1Q3AXP7_CEPFO Myb_DNA-binding domain-containing p...	206	2e-63
TR:A0AAJ6U237 A0AAJ6U237_POPEU Transcription factor MYB90-like OS=...	206	2e-63
SP:Q9ZTC3 MYB90_ARATH Transcription factor MYB90 OS=Arabidopsis th...	206	2e-63
TR:A0A8T0CNC8 A0A8T0CNC8_CORYI Uncharacterized protein OS=Corymbia...	205	4e-63
TR:A0A8S0VMD0 A0A8S0VMD0_OLEEU Uncharacterized protein OS=Olea eur...	204	4e-63
TR:A0A2Z7CW40 A0A2Z7CW40_9LAMI R2r3-myb transcription factor OS=Do...	208	5e-63
TR:A0A8X7VJW5 A0A8X7VJW5_BRACI Uncharacterized protein OS=Brassica...	204	6e-63
SP:Q9FE25 MYB75_ARATH Transcription factor MYB75 OS=Arabidopsis th...	204	6e-63
TR:A0A8X7W3A0 A0A8X7W3A0_BRACI Uncharacterized protein OS=Brassica...	204	8e-63
TR:A0A4S4EEQ0 A0A4S4EEQ0_CAMSI Uncharacterized protein OS=Camellia...	204	8e-63
TR:A0A3P6B886 A0A3P6B886_BRACM Uncharacterized protein OS=Brassica...	204	9e-63
TR:D7KC40 D7KC40_ARALL Uncharacterized protein OS=Arabidopsis lyra...	204	9e-63
TR:A0AAD6KD17 A0AAD6KD17_9ROSI MYB transcription factor OS=Salix u...	205	9e-63
TR:K4D457 K4D457_SOLLC MYB transcriptional factor OS=Solanum lycop...	205	1e-62
TR:A0AAD5NHG2 A0AAD5NHG2_ACENE Uncharacterized protein OS=Acer neg...	205	1e-62
SP:Q9FNV9 MY113_ARATH Transcription factor MYB113 OS=Arabidopsis t...	204	1e-62
TR:A0A7N2KX83 A0A7N2KX83_QUELO R2R3 MYB transcription factor OS=Qu...	204	1e-62
TR:A0A7N2KWR3 A0A7N2KWR3_QUELO Uncharacterized protein OS=Quercus ...	204	1e-62
TR:U5GFP3 U5GFP3_POPTR MYB family protein OS=Populus trichocarpa O...	204	2e-62
TR:A0AAD3Y7Q0 A0AAD3Y7Q0_NEPGR Uncharacterized protein OS=Nepenthe...	202	2e-62
TR:A0A6J5WRS7 A0A6J5WRS7_PRUAR Uncharacterized protein OS=Prunus a...	203	2e-62
TR:A0A8K0H4N7 A0A8K0H4N7_9ROSA MYB transcription factor OS=Rhamnel...	201	2e-62
TR:A0A6P5XJX0 A0A6P5XJX0_DURZI Transcription factor MYB90-like OS=...	204	2e-62
TR:A0A5E4GJD9 A0A5E4GJD9_PRUDU Uncharacterized protein OS=Prunus d...	202	3e-62
TR:A0A3Q7IPC0 A0A3Q7IPC0_SOLLC Uncharacterized protein OS=Solanum ...	204	3e-62
TR:A0A8T1QZP0 A0A8T1QZP0_CARIL Uncharacterized protein OS=Carya il...	203	3e-62
TR:A0A5C7IL71 A0A5C7IL71_9ROSI Uncharacterized protein OS=Acer yan...	204	3e-62
TR:A0A7J7I7M7 A0A7J7I7M7_CAMSI Uncharacterized protein OS=Camellia...	202	3e-62
TR:A0A7N1A7I8 A0A7N1A7I8_KALFE Uncharacterized protein OS=Kalancho...	204	4e-62
TR:A0A5J5C1F1 A0A5J5C1F1_9ASTE Uncharacterized protein OS=Nyssa si...	202	4e-62
TR:A0A7N0SZF5 A0A7N0SZF5_KALFE Uncharacterized protein OS=Kalancho...	204	4e-62
TR:A0AA39SXN7 A0AA39SXN7_ACESA Uncharacterized protein OS=Acer sac...	203	4e-62
TR:A0A2C9WK05 A0A2C9WK05_MANES MYB transcription factor OS=Manihot...	202	5e-62
TR:H6S2W7 H6S2W7_PRUPE R2R3 MYB transcription factor 10 OS=Prunus ...	202	5e-62
TR:A0A8X7Y721 A0A8X7Y721_POPTO Uncharacterized protein OS=Populus ...	202	6e-62
TR:A0A2K1X736 A0A2K1X736_POPTR MYB family protein OS=Populus trich...	203	6e-62
TR:A0A540MY14 A0A540MY14_MALBA Uncharacterized protein OS=Malus ba...	200	6e-62
TR:A0A2P6RR61 A0A2P6RR61_ROSCH Putative transcription factor MYB-H...	201	7e-62
TR:A0A6P5S4N2 A0A6P5S4N2_PRUAV LOW QUALITY PROTEIN: transcription ...	200	8e-62
TR:A0A8X7UX69 A0A8X7UX69_BRACI Uncharacterized protein OS=Brassica...	202	8e-62
TR:A0A8S9PEH9 A0A8S9PEH9_BRACR Uncharacterized protein OS=Brassica...	202	8e-62
TR:A0A9N7R6I6 A0A9N7R6I6_STRHE Transcription factor MYB113 OS=Stri...	199	9e-62
TR:A0A443N944 A0A443N944_9MAGN Transcription factor MYB114 isoform...	201	1e-61
TR:A0A314UTA0 A0A314UTA0_PRUYE R2R3-MYB transcription factor OS=Pr...	200	1e-61
TR:A0A9P0YLC9 A0A9P0YLC9_CUSEU Uncharacterized protein OS=Cuscuta ...	201	1e-61
TR:A0A5D2YIC1 A0A5D2YIC1_GOSMU Uncharacterized protein OS=Gossypiu...	201	1e-61
TR:A0AAD6LI35 A0AAD6LI35_9ROSI Transcription factor MYB1-like OS=P...	201	2e-61
TR:A0AAD8P4X4 A0AAD8P4X4_TARER Uncharacterized protein OS=Tagetes ...	200	2e-61
TR:A0A8X8W752 A0A8X8W752_SALSN Myb proto-oncogene protein, plant O...	199	2e-61
TR:A0AAE0AMI3 A0AAE0AMI3_9ROSI Uncharacterized protein OS=Dipteron...	200	2e-61
TR:A0A6D2J0V6 A0A6D2J0V6_9BRAS Uncharacterized protein OS=Microthl...	200	2e-61

TR:A0AAD2E824	A0AAD2E824_9LAMI	Uncharacterized protein OS=Fraxinus...	200	3e-61
TR:A0AA38Z5Y2	A0AA38Z5Y2_VITRO	Uncharacterized protein OS=Vitis ro...	200	3e-61
TR:A0A8X7VKS1	A0A8X7VKS1_BRACI	Uncharacterized protein OS=Brassica...	200	3e-61
TR:F6HSK4	F6HSK4_VITVI	Uncharacterized protein OS=Vitis vinifera O...	200	4e-61
TR:A0A314YUH7	A0A314YUH7_PRUYE	Myb10 V1-3 OS=Prunus yedoensis var....	197	4e-61
TR:A0A8B8NT24	A0A8B8NT24_9MYRT	Transcription factor MYB75-like OS=...	199	4e-61
TR:A0AAD2DTN1	A0AAD2DTN1_9LAMI	Uncharacterized protein OS=Fraxinus...	197	4e-61
TR:A0A2R6R5C6	A0A2R6R5C6_ACTCC	Transcription factor like OS=Actini...	200	4e-61
TR:A0A9D3WM64	A0A9D3WM64_9ROSI	Uncharacterized protein OS=Gossypiu...	200	4e-61
TR:A0A3Q7INJ7	A0A3Q7INJ7_SOLLC	Uncharacterized protein OS=Solanum ...	200	4e-61
TR:A0A2I4EAZ8	A0A2I4EAZ8_JUGRE	Transcription factor MYB1-like OS=J...	200	4e-61
TR:A0AA38Z5M5	A0AA38Z5M5_VITRO	Uncharacterized protein OS=Vitis ro...	199	5e-61
TR:A0AAJ6U2G7	A0AAJ6U2G7_POPEU	Transcription factor MYB113-like OS...	200	6e-61
TR:A0A8J6CXL5	A0A8J6CXL5_9ROSI	Uncharacterized protein OS=Gossypiu...	199	6e-61
TR:A0A7N0VIE3	A0A7N0VIE3_KALFE	MYB transcription factor OS=Kalanch...	201	6e-61
TR:A0AAD3S4I8	A0AAD3S4I8_NEPGR	Uncharacterized protein OS=Nepenthe...	199	7e-61
TR:A0A8T0CSU4	A0A8T0CSU4_CORYI	MYB transcription factor OS=Corymbi...	199	7e-61
TR:A0A251V1K3	A0A251V1K3_HELAN	Putative homeodomain-like protein O...	199	8e-61
TR:A0A443N926	A0A443N926_9MAGN	Transcription factor MYB1 OS=Cinnam...	199	1e-60
SP:A0A2R6S148	MYB1_ACTCC	Transcription factor MYB1 OS=Actinidia ch...	198	1e-60
TR:A0A6P5S864	A0A6P5S864_PRUAV	Transcription factor MYB75-like OS=...	199	1e-60
TR:A0A7N2KX93	A0A7N2KX93_QUELO	Uncharacterized protein OS=Quercus ...	199	1e-60
TR:A0A6J0P3K6	A0A6J0P3K6_RAPSA	Transcription factor MYB90 OS=Rapha...	200	1e-60
TR:A0A5D2YIN4	A0A5D2YIN4_GOSMU	Uncharacterized protein OS=Gossypiu...	199	1e-60
TR:A0A1U8MHH3	A0A1U8MHH3_GOSHI	Transcription factor MYB90-like OS=...	199	1e-60
TR:A0A1U8P3C3	A0A1U8P3C3_GOSHI	Transcription factor MYB114-like is...	196	1e-60
TR:A0AAD2DWX0	A0AAD2DWX0_9LAMI	Uncharacterized protein OS=Fraxinus...	198	2e-60
TR:A0A1U8K0Z0	A0A1U8K0Z0_GOSHI	PAP1D-T OS=Gossypium hirsutum OX=36...	198	2e-60
TR:A0A2P5XR24	A0A2P5XR24_GOSBA	Uncharacterized protein OS=Gossypiu...	198	2e-60
TR:A0A5D2U8N6	A0A5D2U8N6_GOSMU	Uncharacterized protein OS=Gossypiu...	198	2e-60
TR:A0A8X7TTX7	A0A8X7TTX7_BRACI	Uncharacterized protein OS=Brassica...	198	2e-60
TR:D7KTP3	D7KTP3_ARALL	Uncharacterized protein OS=Arabidopsis lyra...	194	2e-60
TR:W8QH67	W8QH67_PRUPE	Putative MYB105 OS=Prunus persica OX=3760 G...	197	3e-60
TR:A0A6P3Z2B6	A0A6P3Z2B6_ZIZJJ	Transcription factor MYB75-like OS=...	197	3e-60
TR:A0A8T1QZF2	A0A8T1QZF2_CARIL	Uncharacterized protein OS=Carya il...	197	3e-60
TR:A0A7N2KY08	A0A7N2KY08_QUELO	Uncharacterized protein OS=Quercus ...	197	3e-60
TR:R0GID0	R0GID0_9BRAS	Uncharacterized protein OS=Capsella rubella...	197	3e-60
TR:D7KTP4	D7KTP4_ARALL	Uncharacterized protein OS=Arabidopsis lyra...	197	3e-60
TR:A0AAJ6T8Z8	A0AAJ6T8Z8_POPEU	Transcription factor MYB113-like OS...	197	4e-60
TR:A0A0D3CYD2	A0A0D3CYD2_BRAOL	Uncharacterized protein OS=Brassica...	197	5e-60
TR:A0A5N5GE74	A0A5N5GE74_9ROSA	R2R3 MYB transcription factor 10 OS...	197	5e-60
TR:G7K1W9	G7K1W9_MEDTR	Putative transcription factor MYB-HB-like f...	198	5e-60
TR:A0A072TKL4	A0A072TKL4_MEDTR	R2R3-myb transcription factor OS=Me...	197	5e-60
TR:A0A9D4YBH4	A0A9D4YBH4_PEA	Uncharacterized protein OS=Pisum sati...	197	6e-60
TR:A0AAJ6U2C9	A0AAJ6U2C9_POPEU	Transcription factor MYB90-like OS=...	197	6e-60
TR:A0A0D2PP18	A0A0D2PP18_GOSRA	PAP1D OS=Gossypium raimondii OX=297...	197	6e-60
TR:A0A445J565	A0A445J565_GLYSO	Transcription factor MYB113 OS=Glyc...	196	7e-60
TR:K7LFP2	K7LFP2_SOYBN	MYB transcription factor OS=Glycine max OX=...	196	7e-60
TR:A0AAD5D4I9	A0AAD5D4I9_AMBAR	Uncharacterized protein OS=Ambrosia...	197	7e-60
TR:A0A200PYD2	A0A200PYD2_MACCD	SANT/Myb domain OS=Macleania cordata...	198	7e-60
TR:C7A830	C7A830_MEDTR	MYB transcription factor LAP1 OS=Medicago t...	196	8e-60
TR:A0A8B8PTU5	A0A8B8PTU5_9MYRT	Transcription factor MYB113-like OS...	196	8e-60
TR:A0A1U7ZHH6	A0A1U7ZHH6_NELNU	Transcription factor MYB114-like OS...	196	9e-60
TR:A0AA86S181	A0AA86S181_9FABA	Uncharacterized protein OS=Sphenost...	196	9e-60
TR:A0A0D2PWL9	A0A0D2PWL9_GOSRA	PAP2D OS=Gossypium raimondii OX=297...	196	1e-59

TR:A0A078IFF6	A0A078IFF6_BRANA	BnaCnng17820D protein OS=Brassica n...	196	1e-59
TR:A0A9Q0ZZL5	A0A9Q0ZZL5_9ROSI	TRANSCRIPTION FACTOR MYB8-RELATED-R...	197	1e-59
TR:A0A8T0CUX0	A0A8T0CUX0_CORYI	Uncharacterized protein OS=Corymbia...	195	1e-59
TR:A0A5B6UXF9	A0A5B6UXF9_9ROSI	Transcription factor MYB114-like pr...	196	1e-59
TR:A0A5D2U5K5	A0A5D2U5K5_GOSMU	Uncharacterized protein OS=Gossypiu...	196	1e-59
TR:A0A1U8P0V3	A0A1U8P0V3_GOSHI	PAP2D-T OS=Gossypium hirsutum OX=36...	196	1e-59
TR:A0A8J5Z410	A0A8J5Z410_9ROSI	Uncharacterized protein OS=Gossypiu...	196	1e-59
TR:A0AA39AJ70	A0AA39AJ70_VITRO	Myb-related transcription factor A1...	196	1e-59
TR:V4ST16	V4ST16_CITCL	Uncharacterized protein OS=Citrus clementin...	196	1e-59
TR:A0A6P5SDR8	A0A6P5SDR8_PRUAV	Transcription factor MYB75-like OS=...	196	1e-59
TR:A0A6P5SES1	A0A6P5SES1_PRUAV	Transcription factor MYB75-like OS=...	196	2e-59
TR:A0A9W7HWA1	A0A9W7HWA1_HIBTR	Uncharacterized protein OS=Hibiscus...	196	2e-59
TR:M5WUC9	M5WUC9_PRUPE	R2R3 MYB transcription factor 10 OS=Prunus ...	195	2e-59
TR:A0A9W7HY46	A0A9W7HY46_HIBTR	Uncharacterized protein OS=Hibiscus...	196	2e-59
TR:A0A2R6R5A3	A0A2R6R5A3_ACTCC	Transcription factor like OS=Actini...	195	2e-59
TR:A0A834YCA2	A0A834YCA2_TETSI	Uncharacterized protein (Fragment) ...	191	2e-59
TR:A0A6A1VEK5	A0A6A1VEK5_9ROSI	Uncharacterized protein OS=Morella ...	196	3e-59
TR:A0AAD5D4E3	A0AAD5D4E3_AMBAR	Uncharacterized protein OS=Ambrosia...	195	4e-59
TR:A0A5N5GDD6	A0A5N5GDD6_9ROSA	R2R3 MYB transcription factor 10 OS...	194	4e-59
TR:A0A314YR75	A0A314YR75_PRUYE	Myb10 V1-2 OS=Prunus yedoensis var....	194	4e-59
TR:A0AA35YMG7	A0AA35YMG7_LACSI	Uncharacterized protein OS=Lactuca ...	195	5e-59
TR:A0A1Q3D4A2	A0A1Q3D4A2_CEPFO	Myb_DNA-binding domain-containing p...	194	5e-59
TR:A0A2K1X734	A0A2K1X734_POPTR	MYB family protein OS=Populus trich...	195	5e-59
TR:A0A8X7Y6J4	A0A8X7Y6J4_POPTO	Uncharacterized protein OS=Populus ...	196	5e-59
TR:A0A2U1NZK7	A0A2U1NZK7_ARTAN	MYB6 OS=Artemisia annua OX=35608 GN...	194	6e-59
TR:A0A9D3WKJ1	A0A9D3WKJ1_9ROSI	Uncharacterized protein OS=Gossypiu...	194	6e-59
TR:A0A2I0JFE6	A0A2I0JFE6_PUNGR	Uncharacterized protein OS=Punica g...	190	6e-59
TR:A0AAD6LH85	A0AAD6LH85_9ROSI	Uncharacterized protein OS=Populus ...	204	6e-59
TR:W8QQ69	W8QQ69_PRUPE	MYB104 OS=Prunus persica OX=3760 GN=PRUPE_6...	194	7e-59
TR:A0A1S3TTY9	A0A1S3TTY9_VIGRR	Transcription factor MYB90 OS=Vigna...	194	7e-59
TR:A0A251T4M8	A0A251T4M8_HELAN	Putative production of anthocyanin ...	194	7e-59
TR:A0A151S3A8	A0A151S3A8_CAJCA	Transcription factor MYB113 family ...	192	7e-59
TR:A0A3N7G536	A0A3N7G536_POPTR	MYB family protein OS=Populus trich...	194	7e-59
TR:A0A9W7HZN5	A0A9W7HZN5_HIBTR	Uncharacterized protein OS=Hibiscus...	195	7e-59
TR:A0A9W7HY50	A0A9W7HY50_HIBTR	Uncharacterized protein OS=Hibiscus...	195	8e-59
TR:A0AAD3TJ96	A0AAD3TJ96_NEPGR	Uncharacterized protein OS=Nepenthe...	195	9e-59
TR:A0A7J6I693	A0A7J6I693_CANSA	Uncharacterized protein OS=Cannabis...	196	1e-58
TR:A0A6A5N558	A0A6A5N558_LUPAL	Putative transcription factor MYB-H...	193	1e-58
TR:A0A2I4EBF8	A0A2I4EBF8_JUGRE	Transcription factor MYB114-like OS...	193	1e-58
TR:A0AAD4X9M2	A0AAD4X9M2_9MAGN	Uncharacterized protein OS=Papaver ...	191	1e-58
TR:A0A6P4BEZ5	A0A6P4BEZ5_ZIZJJ	Transcription factor MYB113-like OS...	194	1e-58
TR:A0AAD6LHM1	A0AAD6LHM1_9ROSI	Transcription factor MYB1-like OS=P...	194	1e-58
TR:A0A6A3C2P3	A0A6A3C2P3_HIBSY	Transcription factor MYB1 OS=Hibisc...	193	1e-58
TR:A0AAD6KF29	A0AAD6KF29_9ROSI	Uncharacterized protein OS=Salix ud...	194	1e-58
TR:G7K1X2	G7K1X2_MEDTR	R2R3-myb transcription factor OS=Medicago t...	194	1e-58
TR:G7K1U3	G7K1U3_MEDTR	R2R3-myb transcription factor OS=Medicago t...	194	1e-58
TR:A0A314YS29	A0A314YS29_PRUYE	Myb10 V2 OS=Prunus yedoensis var. n...	189	1e-58
TR:B9H7E6	B9H7E6_POPTR	MYB family protein OS=Populus trichocarpa O...	194	1e-58
TR:A0A5B6UPR1	A0A5B6UPR1_9ROSI	Transcription factor MYB114-like pr...	193	1e-58
TR:A0A8B8L551	A0A8B8L551_ABRPR	Transcription factor MYB90-like OS=...	193	1e-58
TR:A0AA35US64	A0AA35US64_LACSI	Uncharacterized protein OS=Lactuca ...	193	2e-58
TR:A0A5E4GJB3	A0A5E4GJB3_PRUDU	PREDICTED: mRNAion factor OS=Prunus...	191	2e-58
TR:A0A5E4FLP2	A0A5E4FLP2_PRUDU	PREDICTED: mRNAion OS=Prunus dulcis...	193	2e-58
TR:A0A059VKQ9	A0A059VKQ9_PUNGR	MYB transcription factor OS=Punica ...	193	2e-58
TR:A0A1S2XN23	A0A1S2XN23_CICAR	Transcription factor MYB114-like OS...	192	2e-58

TR:A0A9D5B984	A0A9D5B984_PEA	Uncharacterized protein OS=Pisum sati...	193	2e-58
TR:A0A9D4YC55	A0A9D4YC55_PEA	Uncharacterized protein (Fragment) OS...	193	2e-58
TR:A0A7J6UYF1	A0A7J6UYF1_THATH	Transcription factor OS=Thalictrum ...	192	3e-58
TR:A0A8B8N268	A0A8B8N268_9MYRT	Transcription factor MYB75-like OS=...	192	3e-58
TR:A0AAF1ARW0	A0AAF1ARW0_DAUCS	Uncharacterized protein OS=Daucus c...	194	3e-58
TR:A0A2U1KG48	A0A2U1KG48_ARTAN	MYB6 OS=Artemisia annua OX=35608 GN...	192	3e-58
TR:A0A067KGQ2	A0A067KGQ2_JATCU	MYB family protein OS=Jatropha curc...	192	4e-58
TR:A0A5E4FLT6	A0A5E4FLT6_PRUDU	PREDICTED: mRNAion OS=Prunus dulcis...	192	4e-58
TR:A0AA39E825	A0AA39E825_VITRO	Uncharacterized protein OS=Vitis ro...	195	4e-58
TR:A0A2G5DSV2	A0A2G5DSV2_AQUCA	Uncharacterized protein OS=Aquilegi...	191	5e-58
TR:A0A9Q0ZZT2	A0A9Q0ZZT2_9ROSI	TRANSCRIPTION FACTOR MYB8-RELATED-R...	191	5e-58
TR:A0A6P5U6P2	A0A6P5U6P2_PRUAV	Transcription factor MYB90-like OS=...	192	5e-58
TR:A0AAD8MNG4	A0AAD8MNG4_9APIA	Transcription factor MYB6 OS=Heracl...	193	6e-58
TR:A0A498J9W0	A0A498J9W0_MALDO	Uncharacterized protein OS=Malus do...	190	7e-58
TR:A0A445FBD4	A0A445FBD4_GLYSO	Transcription factor MYB113 OS=Glyc...	192	7e-58
TR:K7MW48	K7MW48_SOYBN	MYB transcription factor OS=Glycine max OX=...	192	7e-58
TR:Q6L9M8	Q6L9M8_VITVI	Myb-related transcription factor VvMYBA2 OS...	192	7e-58
TR:A0A445J5B5	A0A445J5B5_GLYSO	Transcription factor MYB113 OS=Glyc...	191	8e-58
TR:K7LFP3	K7LFP3_SOYBN	Uncharacterized protein OS=Glycine max OX=3...	191	8e-58
TR:A0AAD8LAZ0	A0AAD8LAZ0_TARER	Uncharacterized protein OS=Tagetes ...	191	8e-58
TR:A0A9Q0VPV6	A0A9Q0VPV6_9ROSI	TRANSCRIPTION FACTOR MYB113-RELATED...	192	9e-58
TR:A0A2Z6PIX0	A0A2Z6PIX0_TRISU	Uncharacterized protein OS=Trifolium...	191	9e-58
TR:A0AAF0X376	A0AAF0X376_DAUCS	Uncharacterized protein OS=Daucus c...	189	1e-57
TR:A0A2G5CKQ6	A0A2G5CKQ6_AQUCA	Uncharacterized protein OS=Aquilegi...	191	1e-57
TR:A0A5P1FDN0	A0A5P1FDN0_ASPOF	Uncharacterized protein OS=Asparagu...	193	1e-57
TR:A0A2I0IWM3	A0A2I0IWM3_PUNGR	Uncharacterized protein OS=Punica g...	187	1e-57
TR:A0AA39AI16	A0AA39AI16_VITRO	Uncharacterized protein OS=Vitis ro...	190	1e-57
TR:A0A9R1W8H0	A0A9R1W8H0_LACSA	R2R3-MYB transcriptional factor 1 O...	191	1e-57
TR:Q6L9M7	Q6L9M7_VITVI	Myb-related transcription factor OS=Vitis v...	188	1e-57
TR:A0AA39AH01	A0AA39AH01_VITRO	Myb-related transcription factor A1...	191	1e-57
TR:Q6L973	Q6L973_VITVI	Myb-related transcription factor OS=Vitis v...	191	1e-57
TR:A0A067KGP8	A0A067KGP8_JATCU	MYB family protein OS=Jatropha curc...	190	1e-57
TR:A0A444X8B2	A0A444X8B2_ARAHY	Uncharacterized protein OS=Arachis ...	192	1e-57
TR:A0A6A2YCQ7	A0A6A2YCQ7_HIBSY	Transcription factor MYB90 OS=Hibis...	191	1e-57
TR:A0A6A5MSP3	A0A6A5MSP3_LUPAL	Putative transcription factor MYB-H...	191	1e-57
TR:G7K1V2	G7K1V2_MEDTR	R2R3-myb transcription factor OS=Medicago t...	192	1e-57
TR:A0A151S360	A0A151S360_CAJCA	Transcription factor MYB113 family ...	191	1e-57
TR:S8CSC4	S8CSC4_9LAMI	Anthocyanin 2 (Fragment) OS=Genlisea aurea ...	186	1e-57
TR:A0A1J7IAQ8	A0A1J7IAQ8_LUPAN	Uncharacterized protein OS=Lupinus ...	190	2e-57
TR:A0AAD6KDE7	A0AAD6KDE7_9ROSI	Uncharacterized protein OS=Salix ud...	191	2e-57
TR:A0A2U1M1Y0	A0A2U1M1Y0_ARTAN	MYB6 OS=Artemisia annua OX=35608 GN...	191	2e-57
TR:A0A151TCC8	A0A151TCC8_CAJCA	Transcription factor MYB113 family ...	190	2e-57
TR:A0A5E4FN61	A0A5E4FN61_PRUDU	PREDICTED: mRNAion OS=Prunus dulcis...	190	2e-57
TR:A0AA86RMY8	A0AA86RMY8_9FABA	Uncharacterized protein OS=Sphenost...	189	2e-57
TR:A0A978VAU8	A0A978VAU8_ZIZJJ	Transcription factor MYB114-like OS...	191	3e-57
TR:A0A9W7HZ11	A0A9W7HZ11_HIBTR	Uncharacterized protein OS=Hibiscus...	190	3e-57
TR:A0A8X7XX58	A0A8X7XX58_POPTO	Uncharacterized protein OS=Populus ...	191	3e-57
TR:A0A2G5DS31	A0A2G5DS31_AQUCA	Uncharacterized protein OS=Aquilegi...	189	3e-57
TR:A0A5D2YLM1	A0A5D2YLM1_GOSMU	Uncharacterized protein OS=Gossypiu...	190	3e-57
TR:A0A067K5D7	A0A067K5D7_JATCU	MYB family protein OS=Jatropha curc...	190	4e-57
TR:A0A2I4GH24	A0A2I4GH24_JUGRE	Transcription factor MYB113-like OS...	191	4e-57
TR:V7B128	V7B128_PHAVU	Uncharacterized protein OS=Phaseolus vulgar...	189	5e-57
TR:A0A175YA44	A0A175YA44_DAUCS	Uncharacterized protein OS=Daucus c...	191	5e-57
TR:A0A8B8L9K5	A0A8B8L9K5_ABRPR	Transcription factor MYB90-like OS=...	189	5e-57
TR:A0A7I8KH99	A0A7I8KH99_SPIIN	Uncharacterized protein OS=Spirodel...	190	6e-57

TR:A0A9R1W208	A0A9R1W208_LACSA	Uncharacterized protein OS=Lactuca ...	190	6e-57
TR:A0AAE1KDB3	A0AAE1KDB3_9FABA	Uncharacterized protein OS=Acacia c...	190	6e-57
TR:A0A9Q0K5E1	A0A9Q0K5E1_9MAGN	Uncharacterized protein OS=Protea c...	190	6e-57
TR:A0A6P4D3L5	A0A6P4D3L5_ARADU	Transcription factor MYB113-like OS...	189	6e-57
TR:D7KTN8	D7KTN8_ARALL	Uncharacterized protein OS=Arabidopsis lyra...	189	6e-57
TR:A0AAE1KKX4	A0AAE1KKX4_9FABA	Uncharacterized protein OS=Acacia c...	190	6e-57
TR:A0A5N5GFU2	A0A5N5GFU2_9ROSA	Transcription factor MYB90-like OS=...	189	7e-57
TR:A0A0R0IMD6	A0A0R0IMD6_SOYBN	MYB transcription factor OS=Glycine...	189	7e-57
TR:A0AAD5D5D6	A0AAD5D5D6_AMBAR	Uncharacterized protein OS=Ambrosia...	189	7e-57
TR:V7B1T9	V7B1T9_PHAVU	Uncharacterized protein OS=Phaseolus vulgar...	188	8e-57
TR:A0A2Z7C5R6	A0A2Z7C5R6_9LAMI	R2r3-myb transcription factor OS=Do...	187	9e-57
TR:A0A9W7LI59	A0A9W7LI59_HIBTR	Uncharacterized protein OS=Hibiscus...	188	1e-56
TR:A0AAF0WQL4	A0AAF0WQL4_DAUCS	Uncharacterized protein OS=Daucus c...	190	1e-56
TR:A0A540LX25	A0A540LX25_MALBA	Uncharacterized protein OS=Malus ba...	186	1e-56
TR:A0A2I4GH38	A0A2I4GH38_JUGRE	Transcription factor MYB113-like OS...	190	1e-56
TR:A0AA35UU89	A0AA35UU89_LACSI	Uncharacterized protein OS=Lactuca ...	189	1e-56
TR:A0A0D2TM73	A0A0D2TM73_GOSRA	Uncharacterized protein OS=Gossypiu...	188	1e-56
TR:A0A8B8LSP1	A0A8B8LSP1_ABRPR	Transcription factor MYB90-like OS=...	188	1e-56
TR:A0A5D2XBL5	A0A5D2XBL5_GOSMU	Uncharacterized protein OS=Gossypiu...	188	1e-56
TR:M9TFA0	M9TFA0_GOSHI R2R3	MYB transcription factor OS=Gossypium ...	188	1e-56
TR:A0A1U8PZG7	A0A1U8PZG7_NELNU	Transcription factor MYB114-like OS...	184	2e-56
TR:A0A6P4C734	A0A6P4C734_ARADU	Transcription factor MYB1-like OS=A...	189	2e-56
TR:A0A445B358	A0A445B358_ARAHY	Uncharacterized protein OS=Arachis ...	189	2e-56
TR:G7K1V6	G7K1V6_MEDTR	Myb-like DNA-binding domain protein OS=Medi...	189	2e-56
TR:A0A0L9U222	A0A0L9U222_PHAAN	Transcription factor MYB114 Myb-rel...	188	2e-56
TR:A0A9Q0L510	A0A9Q0L510_9MAGN	Uncharacterized protein OS=Protea c...	187	2e-56
TR:A0AAF1AUP5	A0AAF1AUP5_DAUCS	Uncharacterized protein OS=Daucus c...	189	2e-56
TR:A0A6A4NHU2	A0A6A4NHU2_LUPAL	Putative transcription factor MYB-H...	188	2e-56
TR:A0AA41VGF8	A0AA41VGF8_PAPNU	Uncharacterized protein OS=Papaver ...	188	2e-56
TR:A0AAD6PUE0	A0AAD6PUE0_9ROSI	Uncharacterized protein OS=Populus ...	197	2e-56
TR:I1N4F9	I1N4F9_SOYBN	MYB transcription factor OS=Glycine max OX=...	187	2e-56
TR:A0A444ZNZ2	A0A444ZNZ2_ARAHY	Uncharacterized protein OS=Arachis ...	187	2e-56
TR:G7K1U9	G7K1U9_MEDTR R2R3-myb	transcription factor OS=Medicago t...	189	3e-56
TR:A0A1U8PTF7	A0A1U8PTF7_GOSHI	Transcription factor MYB113-like OS...	187	3e-56
TR:A0A8K0MNB8	A0A8K0MNB8_9ROSA	Uncharacterized protein OS=Rhamnelli...	186	3e-56
TR:A0AA39RDB3	A0AA39RDB3_ACESA	Uncharacterized protein OS=Acer sac...	189	3e-56
TR:A0A6P4AS78	A0A6P4AS78_ZIZJJ	Transcription factor MYB113-like OS...	186	3e-56
TR:A0A5N5GNP5	A0A5N5GNP5_9ROSA	Transcription factor MYB75-like OS=...	187	4e-56
TR:A0A9D4Y9P3	A0A9D4Y9P3_PEA	Uncharacterized protein OS=Pisum sati...	187	4e-56
TR:A0A5N6ML76	A0A5N6ML76_9ASTR	Uncharacterized protein OS=Mikania ...	186	4e-56
TR:A0A5D2SWX3	A0A5D2SWX3_GOSMU	Uncharacterized protein OS=Gossypiu...	187	4e-56
TR:A0A8J5XYN7	A0A8J5XYN7_9ROSI	Uncharacterized protein OS=Gossypiu...	187	4e-56
TR:A0A9D3VH59	A0A9D3VH59_9ROSI	Uncharacterized protein OS=Gossypiu...	187	4e-56
TR:A0AAD7QBG1	A0AAD7QBG1_QUISA	Myb transcription factor OS=Quillaj...	186	4e-56
TR:A0A5C7HYC9	A0A5C7HYC9_9ROSI	MYB transcription factor OS=Acer ya...	188	5e-56
TR:A0AAE1MY52	A0AAE1MY52_9FABA	Uncharacterized protein OS=Acacia c...	187	5e-56
TR:A0A6A6KIS2	A0A6A6KIS2_HEVBR	Uncharacterized protein OS=Hevea br...	188	6e-56
TR:A0A9D5D2F6	A0A9D5D2F6_9LILI	Uncharacterized protein OS=Dioscore...	185	7e-56
TR:A0AAD8LC49	A0AAD8LC49_TARER	Uncharacterized protein OS=Tagetes ...	187	7e-56
TR:A0A445J566	A0A445J566_GLYSO	Transcription factor MYB113 OS=Glyc...	189	7e-56
TR:A0AAD8MWA9	A0AAD8MWA9_9APIA	Transcription factor MYB6 OS=Heracl...	188	7e-56
TR:A0A8J6D1Z2	A0A8J6D1Z2_9ROSI	Uncharacterized protein OS=Gossypiu...	187	8e-56
TR:A0A1U8AFK2	A0A1U8AFK2_NELNU	Transcription factor MYB114-like OS...	185	8e-56
TR:A0A9Q0GTT1	A0A9Q0GTT1_9MAGN	Uncharacterized protein OS=Protea c...	187	8e-56
TR:A0AAD6PTK2	A0AAD6PTK2_9ROSI	Uncharacterized protein OS=Populus ...	197	9e-56

TR:A0A0R0F4D4	A0A0R0F4D4_SOYBN	MYB transcription factor OS=Glycine...	184	1e-55
TR:A0AA86S9D7	A0AA86S9D7_9FABA	Uncharacterized protein OS=Sphenost...	185	1e-55
TR:A0AA39AGV5	A0AA39AGV5_VITRO	Uncharacterized protein OS=Vitis ro...	185	1e-55
TR:A0AAD5GKV0	A0AAD5GKV0_AMBAR	Uncharacterized protein OS=Ambrosia...	186	1e-55
TR:A0A9R1VPL4	A0A9R1VPL4_LACSA	Uncharacterized protein OS=Lactuca ...	186	1e-55
TR:W8QG90	W8QG90_PRUPE	Putative MYB106 OS=Prunus persica OX=3760 G...	186	1e-55
TR:A0A6I9RUC5	A0A6I9RUC5_ELAGV	Transcription factor MYB113-like OS...	185	2e-55
TR:F6I062	F6I062_VITVI	Uncharacterized protein OS=Vitis vinifera O...	185	2e-55
TR:A0A068FIV3	A0A068FIV3_ELAGV	Transcription factor MYB1 isoform X...	185	2e-55
TR:A0A7J7M7H0	A0A7J7M7H0_9MAGN	Uncharacterized protein OS=Kingdoni...	186	2e-55
TR:A0A9Q0K4R4	A0A9Q0K4R4_9MAGN	Uncharacterized protein OS=Protea c...	186	2e-55
TR:A0A2P6RGL3	A0A2P6RGL3_ROSCH	Putative transcription factor MYB-H...	186	2e-55
TR:A0A5N5JFU3	A0A5N5JFU3_9ROSI	Uncharacterized protein OS=Salix br...	186	2e-55
TR:A0AA39E3U6	A0AA39E3U6_VITRO	Uncharacterized protein OS=Vitis ro...	186	2e-55
TR:A0A151TCA1	A0A151TCA1_CAJCA	Transcription factor MYB113 family ...	184	2e-55
TR:A0AAF0WJP4	A0AAF0WJP4_DAUCS	Uncharacterized protein OS=Daucus c...	187	2e-55
TR:A0A8X7Y0Z6	A0A8X7Y0Z6_POPTO	Uncharacterized protein OS=Populus ...	184	2e-55
TR:A0A6A6KI32	A0A6A6KI32_HEVBR	Uncharacterized protein OS=Hevea br...	186	2e-55
TR:A0A072TH52	A0A072TH52_MEDTR	Myb transcription factor (Fragment)...	183	2e-55
TR:A0A9D5B6V5	A0A9D5B6V5_PEA	Uncharacterized protein (Fragment) OS...	185	2e-55
TR:A0A1S2XJ75	A0A1S2XJ75_CICAR	Transcription factor MYB90-like iso...	185	2e-55
TR:A0A1U8Q1K8	A0A1U8Q1K8_NELNU	Transcription factor MYB114-like OS...	184	2e-55
TR:A0A3Q7XN18	A0A3Q7XN18_CICAR	Transcription factor MYB90-like OS=...	184	3e-55
TR:A0AAD4S5T0	A0AAD4S5T0_9MAGN	Uncharacterized protein OS=Papaver ...	188	3e-55
TR:A0A9D5D270	A0A9D5D270_9LILI	Uncharacterized protein OS=Dioscore...	185	3e-55
TR:A0AA35YYK2	A0AA35YYK2_LACSI	Uncharacterized protein OS=Lactuca ...	185	3e-55
TR:A0A445FXI7	A0A445FXI7_GLYSO	Transcription factor MYB113 OS=Glyc...	184	3e-55
TR:A0A9Q0KT42	A0A9Q0KT42_9MAGN	Uncharacterized protein OS=Protea c...	184	3e-55
TR:A0A1B5Z7F8	A0A1B5Z7F8_TRISU	R2R3-MYB transcription factor OS=Tr...	184	4e-55
TR:A0A5E4GJA7	A0A5E4GJA7_PRUDU	PREDICTED: mRNAion OS=Prunus dulcis...	184	4e-55
TR:A0A2I0IGA3	A0A2I0IGA3_PUNGR	Transcription factor MYB1-like OS=P...	184	4e-55
TR:A0A1S3TFY7	A0A1S3TFY7_VIGRR	Transcription factor MYB113-like OS...	184	4e-55
TR:A0A061F261	A0A061F261_THECC	Myb domain protein 113, putative OS...	184	5e-55
TR:A0A072TGV7	A0A072TGV7_MEDTR	Myb transcription factor OS=Medicag...	184	5e-55
TR:A0A7I8KVV3	A0A7I8KVV3_SPIIN	Uncharacterized protein OS=Spirodel...	185	6e-55
TR:A0A1J7HSN0	A0A1J7HSN0_LUPAN	Uncharacterized protein OS=Lupinus ...	183	6e-55
TR:M5WV61	M5WV61_PRUPE	Uncharacterized protein OS=Prunus persica O...	184	6e-55
TR:I1N2Q1	I1N2Q1_SOYBN	Uncharacterized protein OS=Glycine max OX=3...	183	7e-55
TR:A0A0B2SF14	A0A0B2SF14_GLYSO	Transcription factor MYB113 OS=Glyc...	183	7e-55
TR:A0A8B8Q099	A0A8B8Q099_9MYRT	Transcription repressor MYB6-like O...	183	8e-55
TR:A0A6A5PNG4	A0A6A5PNG4_LUPAL	Putative transcription factor MYB-H...	183	8e-55
TR:A0AAD8LY89	A0AAD8LY89_9APIA	Transcription factor MYB6 OS=Heracl...	185	8e-55
TR:A0AAD8IC59	A0AAD8IC59_9APIA	Transcription factor MYB6 OS=Heracl...	185	8e-55
TR:A0A2P5WUF4	A0A2P5WUF4_GOSBA	Uncharacterized protein OS=Gossypiu...	183	9e-55
TR:A0AA41UY17	A0AA41UY17_PAPNU	Uncharacterized protein OS=Papaver ...	186	1e-54
TR:A0A2I4GH28	A0A2I4GH28_JUGRE	Transcription factor MYB1-like OS=J...	184	1e-54
TR:A2Q618	A2Q618_MEDTR	Homeodomain-related OS=Medicago truncatula ...	184	1e-54
TR:V7B0V2	V7B0V2_PHAVU	Uncharacterized protein OS=Phaseolus vulgar...	182	1e-54
TR:A0A8X7Y444	A0A8X7Y444_POPTO	Uncharacterized protein OS=Populus ...	184	1e-54
TR:V7B4D5	V7B4D5_PHAVU	Uncharacterized protein OS=Phaseolus vulgar...	182	1e-54
TR:D7U3S9	D7U3S9_VITVI	Uncharacterized protein OS=Vitis vinifera O...	182	1e-54
TR:A0AA36E493	A0AA36E493_LACSI	Uncharacterized protein OS=Lactuca ...	182	1e-54
TR:A0A443N903	A0A443N903_9MAGN	Transcription factor MYB1 OS=Cinnam...	184	1e-54
TR:A0A9D4Y042	A0A9D4Y042_PEA	Uncharacterized protein OS=Pisum sati...	183	1e-54
TR:A0A7I8KWR4	A0A7I8KWR4_SPIIN	Uncharacterized protein OS=Spirodel...	184	1e-54

TR:B9RF16	B9RF16_RICCO	R2r3-myb transcription factor, putative OS=...	180	1e-54
TR:A0AA38TQG4	A0AA38TQG4_9ASTR	Uncharacterized protein OS=Centaure...	184	1e-54
TR:A0A2U1MPG2	A0A2U1MPG2_ARTAN	Transcription factor GbMYB1 OS=Arte...	183	2e-54
TR:A0A7I8KQY8	A0A7I8KQY8_SPIIN	Uncharacterized protein OS=Spirodel...	184	2e-54
TR:A0A0S3RVG8	A0A0S3RVG8_PHAAN	Uncharacterized protein OS=Vigna an...	182	2e-54
TR:A0A0L9U5Q7	A0A0L9U5Q7_PHAAN	Transcription factor MYB114 Myb-rel...	182	2e-54
TR:A0A5N5F5C0	A0A5N5F5C0_9ROSA	Transcription factor MYB90-like OS=...	184	2e-54
TR:A0A7J8ZWR8	A0A7J8ZWR8_9ROSI	Uncharacterized protein OS=Gossypiu...	182	2e-54
TR:A0A540LX00	A0A540LX00_MALBA	Uncharacterized protein OS=Malus ba...	182	2e-54
TR:A0AAD8JDK7	A0AAD8JDK7_9APIA	Transcription factor MYB6 OS=Heracl...	184	2e-54
TR:A0A151UE37	A0A151UE37_CAJCA	Transcription factor MYB113 family ...	182	2e-54
TR:A0A067K5D1	A0A067K5D1_JATCU	MYB family protein OS=Jatropha curc...	183	3e-54
TR:A0SQ35	A0SQ35_MALDO	MYB1 transcription factor OS=Malus domestic...	182	3e-54
TR:A0A7N0T069	A0A7N0T069_KALFE	Uncharacterized protein OS=Kalancho...	183	3e-54
TR:A0A6P5RFZ6	A0A6P5RFZ6_PRUAV	Transcription factor WER-like OS=Pr...	182	3e-54
TR:A0A9D5HNG0	A0A9D5HNG0_9LILI	Uncharacterized protein OS=Dioscore...	183	3e-54
TR:M5W0N9	M5W0N9_PRUPE	R2R3-MYB transcription factor OS=Prunus per...	182	4e-54
TR:A0AAD6LHY2	A0AAD6LHY2_9ROSI	Transcription factor MYB1-like OS=P...	182	4e-54
TR:A0A540MU14	A0A540MU14_MALBA	Uncharacterized protein OS=Malus ba...	183	5e-54
TR:A0AA35VPU3	A0AA35VPU3_LACSI	Uncharacterized protein OS=Lactuca ...	181	5e-54
TR:A0A4P1R424	A0A4P1R424_LUPAN	Uncharacterized protein OS=Lupinus ...	181	5e-54
TR:A0A5E4EB89	A0A5E4EB89_PRUDU	PREDICTED: mRNAion OS=Prunus dulcis...	182	6e-54
TR:A0A8X7Y0Y2	A0A8X7Y0Y2_POPTO	Uncharacterized protein OS=Populus ...	182	6e-54
TR:R0I0F5	R0I0F5_9BRAS	Uncharacterized protein OS=Capsella rubella...	181	7e-54
TR:A0AAD6LHP1	A0AAD6LHP1_9ROSI	Transcription factor MYB1-like OS=P...	181	7e-54
TR:A0AAE1SXF5	A0AAE1SXF5_9SOLA	Uncharacterized protein OS=Anisodus...	181	1e-53
TR:A0AA39AGZ2	A0AA39AGZ2_VITRO	Uncharacterized protein OS=Vitis ro...	180	1e-53
TR:A0A1U8B239	A0A1U8B239_NELNU	Myb-related protein 330-like OS=Nel...	184	1e-53
TR:A0A4Y7KPM5	A0A4Y7KPM5_PAPSO	Uncharacterized protein OS=Papaver ...	178	1e-53
TR:A0A2P6R5A7	A0A2P6R5A7_ROSCH	Putative transcription factor MYB-H...	182	1e-53
TR:A0A6A4NSR2	A0A6A4NSR2_LUPAL	Putative transcription factor MYB-H...	180	1e-53
TR:A0A8B8J198	A0A8B8J198_PHODC	Transcription factor MYB114-like OS...	181	1e-53
TR:A0AA36E6X6	A0AA36E6X6_LACSI	Uncharacterized protein OS=Lactuca ...	181	1e-53
TR:A0A9D5HNQ8	A0A9D5HNQ8_9LILI	Uncharacterized protein OS=Dioscore...	177	1e-53
TR:I1KK72	I1KK72_SOYBN	Uncharacterized protein OS=Glycine max OX=3...	180	2e-53
TR:A0A0B2PC47	A0A0B2PC47_GLYSO	Transcription factor MYB113 OS=Glyc...	180	2e-53
TR:A0A6J5XUL3	A0A6J5XUL3_PRUAR	Uncharacterized protein OS=Prunus a...	181	2e-53
TR:A0A6P5YVY9	A0A6P5YVY9_DURZI	Transcription factor MYB113-like OS...	180	2e-53
TR:A0A2K3NSC7	A0A2K3NSC7_TRIPR	Transcription factor MYB 113-like p...	180	2e-53
TR:A0A8X7XYP4	A0A8X7XYP4_POPTO	Uncharacterized protein OS=Populus ...	181	2e-53
TR:A0AA38TNV2	A0AA38TNV2_9ASTR	Uncharacterized protein OS=Centaure...	181	2e-53
TR:A0A9D5D3C9	A0A9D5D3C9_9LILI	Uncharacterized protein OS=Dioscore...	181	2e-53
TR:A0A6A2YXA4	A0A6A2YXA4_HIBSY	Transcription factor MYB75 OS=Hibis...	180	2e-53
TR:A0A067JIU0	A0A067JIU0_JATCU	MYB family protein OS=Jatropha curc...	179	3e-53
TR:A0A4Y7K9Q6	A0A4Y7K9Q6_PAPSO	Uncharacterized protein OS=Papaver ...	181	3e-53
TR:A0A1J7H8E2	A0A1J7H8E2_LUPAN	Uncharacterized protein OS=Lupinus ...	179	3e-53
TR:A0A2G2YLG5	A0A2G2YLG5_CAPAN	Transcription factor OS=Capsicum an...	180	3e-53
TR:A0A9W7M087	A0A9W7M087_HIBTR	Uncharacterized protein OS=Hibiscus...	179	3e-53
TR:A0A6J1A668	A0A6J1A668_9ROSI	Transcription factor MYB114-like is...	179	3e-53
TR:A0A9D5B8I9	A0A9D5B8I9_PEA	Uncharacterized protein OS=Pisum satl...	179	4e-53
TR:A0A6A2Y061	A0A6A2Y061_HIBSY	Transcription factor MYB1 OS=Hibisc...	178	5e-53
TR:A0AAD4SPB1	A0AAD4SPB1_9MAGN	Uncharacterized protein OS=Papaver ...	183	5e-53
TR:A0A1S3B4A4	A0A1S3B4A4_CUCME	Transcription factor MYB90-like OS=...	178	8e-53
TR:A0A1S2XHF2	A0A1S2XHF2_CICAR	Transcription factor MYB90-like iso...	179	1e-52
TR:A0AAJ6U147	A0AAJ6U147_POPEU	Transcription factor MYB90-like iso...	178	1e-52

TR:A0A6J1KE93	A0A6J1KE93_CUCMA	Transcription factor MYB1-like OS=C...	177	1e-52
TR:A0A6D2J0U5	A0A6D2J0U5_9BRAS	Uncharacterized protein OS=Microthl...	178	1e-52
TR:A0A8B8JA01	A0A8B8JA01_PHODC	Transcription factor MYB114 OS=Phoe...	175	2e-52
TR:A0A6J1CTG0	A0A6J1CTG0_MOMCH	Transcription factor WER-like OS=Mo...	177	2e-52
TR:A0A8X8C5A3	A0A8X8C5A3_POPTO	Uncharacterized protein OS=Populus ...	178	3e-52
TR:G7K1W1	G7K1W1_MEDTR	Myb-like DNA-binding domain protein OS=Medi...	179	3e-52
TR:A0A6I9RUD0	A0A6I9RUD0_ELAGV	Transcription factor MYB1-like OS=E...	177	3e-52
TR:A0A0K9NQ42	A0A0K9NQ42_ZOSMR	Myb-related transcription factor OS...	177	3e-52
TR:A0AAD6LJ69	A0AAD6LJ69_9ROSI	Transcription factor MYB1-like OS=P...	177	4e-52
TR:G7K1V4	G7K1V4_MEDTR	Myb-like DNA-binding domain protein OS=Medi...	176	6e-52
TR:A0A2P6RR54	A0A2P6RR54_ROSCH	Putative transcription factor MYB-H...	175	8e-52
TR:A0A0D2SCI4	A0A0D2SCI4_GOSRA	Uncharacterized protein OS=Gossypiu...	174	8e-52
TR:A0A1U8KRM9	A0A1U8KRM9_GOSHI	Transcription factor MYB114-like OS...	174	8e-52
TR:A0A5D2S157	A0A5D2S157_GOSMU	Uncharacterized protein OS=Gossypiu...	174	8e-52
TR:A0A7N0UKU5	A0A7N0UKU5_KALFE	Uncharacterized protein OS=Kalancho...	176	8e-52
TR:A0A5N5H0D8	A0A5N5H0D8_9ROSA	Myb-related protein 308-like OS=Pyr...	175	8e-52
TR:A0A5D2WGV0	A0A5D2WGV0_GOSMU	Uncharacterized protein OS=Gossypiu...	174	8e-52
TR:A0A1U8IEZ4	A0A1U8IEZ4_GOSHI	Transcription factor MYB114-like OS...	174	8e-52
TR:A0A9D3ZG26	A0A9D3ZG26_9ROSI	Uncharacterized protein OS=Gossypiu...	174	8e-52
TR:A0A8T1QQA3	A0A8T1QQA3_CARIL	Uncharacterized protein OS=Carya il...	179	9e-52
TR:A0A6A2XX59	A0A6A2XX59_HIBSY	Transcription factor MYB1 OS=Hibisc...	175	1e-51
TR:A0AAE1KMK3	A0AAE1KMK3_9FABA	Uncharacterized protein OS=Acacia c...	176	1e-51
TR:A0A251V046	A0A251V046_HELAN	Putative homeodomain-like, Myb-like...	175	1e-51
TR:A0A498K558	A0A498K558_MALDO	MYB domain class transcription fact...	176	2e-51
TR:A0AAD6LKC7	A0AAD6LKC7_9ROSI	Uncharacterized protein OS=Populus ...	182	2e-51
TR:A0A540LQG1	A0A540LQG1_MALBA	Uncharacterized protein OS=Malus ba...	176	2e-51
TR:A0A2G5DS33	A0A2G5DS33_AQUCA	Uncharacterized protein OS=Aquilegi...	174	2e-51
TR:A0A9D4Y818	A0A9D4Y818_PEA	Uncharacterized protein OS=Pisum satl...	175	2e-51
TR:A0A8J5YAK6	A0A8J5YAK6_9ROSI	Uncharacterized protein OS=Gossypiu...	173	2e-51
TR:A0A9K3H3M2	A0A9K3H3M2_HELAN	Transcription factor MYB family OS=...	174	5e-51
TR:A0A6J1A5R1	A0A6J1A5R1_9ROSI	Transcription factor MYB114-like is...	174	5e-51
TR:A0A8S0UK24	A0A8S0UK24_OLEEU	Transcription factor MYB114-like (F...	175	5e-51
TR:A0A022RNF2	A0A022RNF2_ERYGU	Uncharacterized protein OS=Erythran...	176	6e-51
TR:A0A314ZA25	A0A314ZA25_PRUYE	Uncharacterized protein OS=Prunus y...	184	6e-51
TR:A0A540NFM9	A0A540NFM9_MALBA	Uncharacterized protein OS=Malus ba...	175	7e-51
TR:A0A2G5E7N3	A0A2G5E7N3_AQUCA	Uncharacterized protein OS=Aquilegi...	172	9e-51
TR:A0A5J5B1P4	A0A5J5B1P4_9ASTE	Uncharacterized protein OS=Nyssa si...	172	1e-50
TR:A0A2G5DS32	A0A2G5DS32_AQUCA	Uncharacterized protein OS=Aquilegi...	172	1e-50
TR:A0A4Y7K3W8	A0A4Y7K3W8_PAPSO	Uncharacterized protein OS=Papaver ...	177	2e-50
TR:A0A7N0TZP9	A0A7N0TZP9_KALFE	Uncharacterized protein OS=Kalancho...	172	2e-50
TR:A0A9D5BSW7	A0A9D5BSW7_9LILI	Uncharacterized protein OS=Dioscore...	174	2e-50

Appendix 9: *In silico* analyses and protease digestion studies to assess the likely allergenicity of Delila and Rosea 1 proteins

The DEL and ROS1 protein sequences were searched against the allergenonline.org database. The results appear below.

1/3/25, 1:46 PM allergenonline.org/search36.php

AllergenOnline Search Results

Note: As of August 2015 we have included gid: groupid in the fasta results that provides detailed information on the allergenicity references for the group, type of allergen, other sequences belonging to the same group and more.

%_id 1 = 100% identity, alen=alignment length

AllergenOnline Database v22 (May 25, 2023)

NOTE Addition of Allergenicity* column on the Browse Database page with classification based on Group references was added on 10 May 2018. Please review the "allergenicity" of any matches you find here with the Browse page and look at Group References (gid) if you want to further evaluate relevance of alignments.

fasta36.exe -q -B -m 9i -w 80 -E 1 -d 20 C:\Windows\Temp\all7102.tmp version2236.fasta

User Query #1 >QGV13005.1 rosea1 [Plant binary expression vector pDEL.ROS]

User Query #1

>QGV13005.1 rosea1 [Plant binary expression vector pDEL.ROS]
MEKNCRGVRK GTWTKEDTL LRQCIEEYGE GKWHQVPHRA GLNRCRKSCR LRWLNLYLRPN IKRGRFSRDE VDLIVRLHKL
LGNKWSLIAG RIPGRTANDV KNFWNTHVGK NLGEDGERCR KNVMNTKTIK LTNIVRPRAR TFTGLHVTWP REVGKTDEFS
NVRLLTDEIP DCEKQTQFYN DVASPDQVE DCIQWWSKLL ETEDGELGN LFEEAQQIGN

fasta36.exe -q -B -m 9i -w 80 -E 1 -d 20 C:\Windows\Temp\all7102.tmp version2236.fasta

FASTA searches a protein or DNA sequence data bank

version 36.3.8g Oct, 2018

Please cite:

W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

Query: C:\Windows\Temp\all7102.tmp

1>>>QGV13005.1 rosea1 [Plant binary expression vector pDEL.ROS] - 220 aa

Library: version2236.fasta

558897 residues in 2291 sequences

Statistics: Altschul/Gish params: n0: 220 Lambda: 0.158 K: 0.019 H: 0.100

statistics sampled from 328 (328) to 328 sequences

Algorithm: FASTA (3.8 Nov 2011) [optimized]

Parameters: BL50 matrix (15:-5), open/ext: -10/-2

ktp: 2, E-join: 1 (0.434), E-opt: 0.2 (0.143), width: 16

Scan time: 0.000

!! No sequences with E() < 1

>>>///

220 residues in 1 query sequences

558897 residues in 2291 library sequences

Scomplib [36.3.8g Oct, 2018]

start: Fri Jan 03 12:46:29 2025 done: Fri Jan 03 12:46:29 2025

Scan time: 0.000 Display time: 0.000

Function used was FASTA [36.3.8g Oct, 2018]

www.allergenonline.org/search36.php 1/11/3/25, 1:52 PM 80mer Sliding Window Search Results

80mer Sliding Window Search Results

Database AllergenOnline Database v22 (May 25, 2023)

Input Query >QGV13004.1 Delila [Plant binary expression vector pDEL.ROS]

MATGIQNQKIVPENLRKQLAIAVRSIQWSYAIFWSNSVAQPGVLEWGDGFYNGDIKTRKT
VQSVELNPDQLGLQRSDQLRELYESLSLGETNTQAKRPTAALSPEDLTDAEWFFLVCMMSF
IFNIGQGLPGRTLARNQAVVWLCNAHRADTKVFSRLLAKSASIQTVVCFPYAEGVVELGA
TELVPEDLNLIQHIKTSFLDSPATVPKIPNYVSNSITNNNDLICEALEHANIPENDLDQL
LNCPTNICSPDNSLDDFADNLLIDESNLAEGINGEVPQTQSWPFMDDAISNCLNSSMNS
SDCISQTHENLESFAPLSDGKGAPETNNCMHSTQECNQQIENTGVQGDEVHYQRVLSNLL

KSSHQLVLGPYFRNGNRESSFVSWNKDGSSGTHVPRSGTSQRFLKKVLFVARMHENSRL
DAGKQKGNSDCLAKPTADEIDRNHVLSEKRRERKINERFMILASLVPSGGKVDKVSILDH
TIDYLRGLERKVDLESNKMKVGRGRESTTTKTLHDAIERTSDNYGTTTRTSNVKKPLTNK
RKASDDTKIGAVNSRGRKLDLTDNITVNITNKDVLIVVSCSSKEFVLLVMEAVRRLSL
DSETVQSSNRDGIISITIKACKGLKVASASVIKQALQKVTMKS

Length 644

Number of 80 mers 565

Number of Sequences with hits 0

No Matches of Greater than 35% Identity Found

AllergenOnline Database v22 (May 25, 2023)

www.allergenonline.org/search8036.cgi 1/11/3/25, 1:49 PM Search Results

>QGV13004.1 Delila [Plant binary expression vector pDEL.ROS]

number of 8mer = 637

Number of Sequences with at least one 8mer match = 0

www.allergenonline.org/shortmer.php 1/11/3/25, 1:49 PM allergenonline.org/search36.php

AllergenOnline Search Results

Note: As of August 2015 we have included gid: groupid in the fasta results that provides detailed information on the allergenicity references for the group, type of allergen, other sequences belonging to the same group and more.

%_id 1 = 100% identity, alen=alignment length

AllergenOnline Database v22 (May 25, 2023)

NOTE Addition of Allergenicity* column on the Browse Database page with classification based on Group references was added on 10 May 2018. Please review the "allergenicity" of any matches you find here with the Browse page and look at Group References (gid) if you want to further evaluate relevance of alignments.

fasta36.exe -q -B -m 9i -w 80 -E 1 -d 20 C:\Windows\Temp\all654.tmp version2236.fasta

User Query #1 >QGV13004.1 Delila [Plant binary expression vector pDEL.ROS]

User Query #1

>QGV13004.1 Delila [Plant binary expression vector pDEL.ROS]

MATGIQNQKI VPENLRKQLA IAVRSIQWSY AIFWSNSVAQ PGVLEWGDGF YNGDIKTRKT VQSVELNPDQ LGLQRSDDLRL
ELYESLSLGE TNOQAKRPTA ALSPEDLTD EWFVLVCMSE IFNIGQGLPG RTLARNQAVW LCNARADTK VFSRSLAKS
ASIQTVVCFP YAEGVVELGA TELVPEDLNL IQHIKTSFLD SPATVPKIPN YVSNISITNN DLICEALEHA NIPENDLDQL
LNCPDNTNCS PDNSLDDFAD NLLIDESNLA EGINGEVPQT QSWPFMDDAI SNCLNSSMNS SOCISQTHEN LESFAPLSDG
KGAPETNNCM HSTQECCNQI ENTGVQGDV HYQRVLSNLL KSSHQLVLGP YFRNGNRESS FVSWNKDGSS GTHVPRSGTS
QRFLKKVLFV VARMHENSRL DAGKQKGNSD CLAKPTADEI DRNHVLSEK RREKINERFM ILASLVPSGG KVDKVSILDH
TIDYLRGLER KVDLESNKM VKGRGRESTT KTLHDAIER TSDNYGTTTRT SNVKKPLTNK RKASDDTKIG AVNSRGRKLD
SLTDNITVNI TNKDVLIIVS CSSKEFVLE VMEAVRRLSL DSETVQSSNR DGIISITIKA KCKGLKVASA SVIKQALQKV TMKS

fasta36.exe -q -B -m 9i -w 80 -E 1 -d 20 C:\Windows\Temp\all654.tmp version2236.fasta

FASTA searches a protein or DNA sequence data bank

version 36.3.8g Oct, 2018

Please cite:

W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448

Query: C:\Windows\Temp\all654.tmp

1>>>QGV13004.1 Delila [Plant binary expression vector pDEL.ROS] - 644 aa

Library: version2236.fasta

558897 residues in 2291 sequences

Statistics: Altschul/Gish params: n0: 644 Lambda: 0.158 K: 0.019 H: 0.100

statistics sampled from 318 (318) to 318 sequences

Algorithm: FASTA (3.8 Nov 2011) [optimized]

Parameters: BL50 matrix (15:-5), open/ext: -10/-2

ktup: 2, E-join: 1 (0.421), E-opt: 0.2 (0.139), width: 16

Scan time: 0.000

!! No sequences with E() < 1

>>>///

644 residues in 1 query sequences

558897 residues in 2291 library sequences

Scomplib [36.3.8g Oct, 2018]

start: Fri Jan 03 12:49:19 2025 done: Fri Jan 03 12:49:19 2025

Scan time: 0.000 Display time: 0.000

Function used was FASTA [36.3.8g Oct, 2018]

www.allergenonline.org/search36.php 1/11/3/25, 1:48 PM 80mer Sliding Window Search Results

80mer Sliding Window Search Results

Database AllergenOnline Database v22 (May 25, 2023)

Input Query >QGV13005.1 rosea1 [Plant binary expression vector pDEL.ROS]

MEKNCRGVRKGTWTKEEDTLRQCIEEYGEKWHQVPHRAGLNRCRKSCRLRWLNLYLRPN

IKRGRFSRDEVDLIVRLHKLLGNKWSLIAGRIPGRTANDVKNFWNTHVGKNLGEDGERCR
KNVMNTKTIKLTNIVRPRARTFTGLHVTWPREVVGKTDEFSNVRLTTDEIPDCEKQTQFYN
DVASPDQDEVEDCIQWWSKLLETTEDEGELGNLFEEAQQIGN
Length 220
Number of 80 mers 141
Number of Sequences with hits 0

No Matches of Greater than 35% Identity Found

AllergenOnline Database v22 (May 25, 2023)

www.allergenonline.org/search8036.cgi 1/1 1/3/25, 1:47 PM Search Results

>QGV13005.1 rosea1 [Plant binary expression vector pDEL.ROS]

number of 8mer = 213

Number of Sequences with at least one 8mer match = 0

www.allergenonline.org/shortmer.php 1/1

Conclusion

No sequences of known allergens match DEL or ROS1, when searched for identity >35% over the whole sequence or as 80mers, no to any 8mers from either sequence match any known allergic epitope.

Proteolytic digestion

The *Delila* (*Del*) and *Rosea1* (*Ros1*) genes from snapdragon are the transgenes present in the *Del/Ros1*-N transformation event and encode two transcription factors which are expressed, post-breaker (ripening), in transformed purple tomatoes.

The DEL and ROS1 proteins were also assessed for their susceptibility to simulated gastrointestinal digestion. An analysis was conducted to test the susceptibility of a protein to digestion is to use a standardized digestion protocol and to monitor the disappearance of the full-length protein over time, and potentially the appearance of smaller peptides that represent visible intermediates in the digestion process (Thomas et al., 2004).

PROCEDURES

Expression and clean-up of Delila and Rosea1 recombinant proteins

Delila and Rosea1 proteins were expressed in *E.coli* (strain BL21) following cloning of the Delila and Rosea 1 cDNAs into a His-tag expression vector.

Transformation (prepare cells for expression fresh every 2 weeks)

1. Transform each plasmid (528-Ros, 528TvDel) into BL21-Codon plus Cells.
2. Spread on LB+Carbenicillin+Chloramphenicol.
3. Incubate overnight at 37°C.

Pre-culture

1. Pick a single colony.
2. Inoculate each colony in 2x10 mL LB+G with 50 µg/mL Carbenicillin + 30 µg/mL Chloramphenicol.
3. Shake overnight at 37°C.

Sub-culture

1. Inoculate each 5 mL pre-culture in 2X 500 mL LB+G with 50 µg/mL Carbenicillin + 30 µg/mL Chloramphenicol.
2. Shake at 180 rpm at 37°C 2~3 hours until OD=0.4~0.6
3. Collect 1mL of non-induced culture in 1.5 mL tube for the SDS-PAGE check.

Induction

1. Add 250 µL of 1M IPTG to each 500 mL sub-culture.
2. Shake 16 hours at 150 rpm at 16°C.

Collection

1. Collect 500 μ L of induced culture in 1.5 mL tube for the SDS-PAGE check.
2. Pellet cells in 250 mL tube (for each 500 mL culture) by spinning 15 min at 8000 rpm at 4°C.
3. Remove LB media completely. (pellet can be washed with 10 mL of buffer)
4. Freeze cells with Liquid Nitrogen.
5. Keep frozen cells in -80°C freezer.

Preparation of sample solution—All the procedures to be performed on ice

1. Thaw the frozen pellets on ice (2 to 4 tubes at one time).
2. Add
 - 5 (or 7.5) mL of “Start and Wash Buffer”
 - 500 μ L of 7x Complete Mini solution
 - 10 mg Lysozyme (-20°C)
 to each pellet from 500 mL culture.
3. Suspend by vortex or pipetting and incubate on ice for one hour.
4. Transfer sample solution to smaller plastic bottle (Mix two samples in one bottle)
5. Sonicate for 10 min. (Repeat 10 sec. pulse /10 sec. interval at level 10)
6. Transfer sample to 40 mL centrifuge tube.
7. Centrifuge at 15000 rpm at 4°C for 20 min.
8. Transfer supernatant to 20 mL glass Universal Bottle (Protein crude sample solution)
 Note: Avoid taking the sticky bubbly layer, which is caused by insufficient sonication or high concentration of *E. coli*.
9. Filter the sample solution with a 0.45 μ m filter and a 10 mL Syringe.

Purification of Delila—All the processes were performed on ice or at 4°C

Preparation of sample solution (See above “Induction and Collection, Sample preparation”)

Binding

1. Add sample solution to the 20 mL glass Universal Bottle containing NiNTA agarose.
2. Shake gently on the orbital shaker at 4°C for one hour.

Wash

3. Set the empty PD-10 column on the top of the 200 or 400 mL glass bottle (for the collection of waste) in the fridge or cold room.
4. Apply the sample solution mixed with NiNTA Agarose to the column.
5. Wash with 10 ml ice-cold “Start and Wash buffer”.
6. Wash with 30 ml ice-cold “Wash Buffer with 25% glycerol”.
7. Wash with 60 ml ice-cold “Start and Wash buffer”.

Elution

8. Elute with 2 mL of “Elution Buffer, 125 mM imidazole”.
 Collect 500 μ L fractions in 1.5 mL tube.
9. Elute with 2 mL of “Elution Buffer, 250 mM imidazole”.
 Collect 500 μ L fractions in 1.5 mL tube.

Check protein

10. Add 5 μ L 4x loading buffer to 15 μ L of each elution for SDS-PAGE checking.
11. Boil the samples for 5 min and load 15 μ L of each sample on SDS-PAGE gel.
12. Run SDS-PAGE at 200 V for 30 min and stain.

Dialysis

1. Choose the eluted fractions that have single band (80 kDa).
2. Wash dialysis tube with distilled water and tie one of the edges.

3. Apply 0.5~2 mL sample to dialysis tube. (Mix some of fractions according to the result of SDS-PAGE checking).
4. Close the other side of the dialysis tube with a clip.
5. Attach the stir bar to the clip with a rubber band.
6. Put tube in 500 mL final buffer and stir for two hours at 4°C.
7. Change the buffer and stir overnight at 4°C.
8. Collect dialyzed sample from the tube.
9. Add 5 μ L 4x loading buffer to 15 μ L of each fraction for SDS-PAGE checking.
10. Boil the samples for 5 min. and load 15 μ L on SDS-PAGE gel.
11. Run SDS-PAGE at 200 V for 30 min and stain.
12. Measure the concentration of proteins by Bradford method.

Storage of purified protein

His-tagged Delila kept binding activity during storage on ice in the cold room for approximately two weeks but lost activity or degraded over longer storage times.

Purification of Rosea1—All procedures were performed on ice

Preparation of sample solution

Binding

1. Stick the prepared HisTrap column into the ice.
2. Slowly (one or two drops per 1 sec.) apply the sample solution to the column with 1 mL Syringe.

Wash

3. Wash with 10 ml ice-cold "Start and Wash buffer".
4. Wash with 10 ml ice-cold "Wash Buffer with 25% glycerol".
5. Wash with 40 ml ice-cold "Start and Wash buffer".

Elution

6. Elute with 3 mL of "Elution Buffer, 250 mM imidazole".
Collect 500 μ L fractions in 1.5 mL tube.
7. Elute with 6 mL of "Elution Buffer, 500 mM imidazole".
Collect 500 μ L fractions in 1.5 mL tubes.

Check protein

8. Add 5 μ L 4x loading buffer to 15 μ L of each elution for SDS-PAGE checking.
9. Boil the samples for 5 min and load 15 μ L of each sample on SDS-PAGE gel.
10. Run SDS-PAGE at 200 V for 30 min and stain.

HiTrap Desalting

1. Choose the eluted fractions that have two bands (The size of His-Rosea1 and the lower band).
2. Apply 1.5 mL sample to HiTrap Desalting Column.
3. Add 5 mL of final buffer and collect each 500 μ L fraction in 1.5 mL tube.
4. Add 5 μ L 4x loading buffer to 15 μ L of each fraction for SDS-PAGE checking.
5. Boil the samples for 5 min and load 15 μ L on SDS-PAGE gel.
6. Run SDS-PAGE at 200 V for 30 min and stain.
7. Measure the concentration of proteins by Bradford method.

Storage of purified protein

His-tagged Rosea1 keep its binding activity during storage on ice in the cold room for approximately two weeks but lost activity or degraded over longer storage periods.

Digestion of Delila and Rosea1 recombinant proteins

For pepsin digestion of the snapdragon transcription factor proteins that are expressed in the transformed purple tomatoes, a method similar to that of Thomas *et al* (2004) was used. In brief, a partially purified sample of transcript at a concentration of 400µg/ml of protein was mixed 1:1 with 2x simulated gastric fluid (SGF; 70mM NaCl, pH 0.9) containing 4000 U/ml of porcine pepsin pre-warmed to 37°C, resulting in a pepsin: protein ratio of 10:1 at a pH of 1.2. The pepsin-protein mixture was incubated at 37°C and samples taken at 0.5, 2, 5, 10, 20, 30 and 60 min for the zero time point samples were prepared by quenching pepsin in SGF before the addition of the protein sample. Each sample was quenched by the addition of 200mM sodium bicarbonate and kept on ice until analysis by SDS-PAGE or stored at -80°C.

SDS-PAGE

Samples for SDS-PAGE were mixed with 4x LDS-sample buffer + 10xDTT, heated at 70°C for 10min and then loaded onto a 12% Bis-Tris gel (Life Technologies) alongside a pre-stained molecular weight marker and electrophoresed for 55min at 200V constant. Following electrophoresis the gel was washed 3x with demineralized water and stained with GelCode Blue Stain Reagent (Life Technologies) for 1h at room temperature with shaking. Gel was de-stained with 2x washes in demineralized water for 1hour

Digestion of Rosea 1 monitored using mass spectrometry.

1. Bands were cut out from gels using razor blades based on the expected ROSEA protein band size just above the pepsin band size. These were then cut up to small ~1mm squares and put in 0.2ml PCR strips in water.
2. Used 400mM ammonium bicarbonate prepared for Digest (0.791g in 25mL MilliQ Water (Sigma A6141 - Batch No 030M0144)).
3. Used remaining Solution B (200mM Ammonium Bicarbonate in 50% Acetonitrile) for Digest.
4. Wash gel plugs in 200µL Solution B (to equilibrate to pH 8ish and to remove the coomassie blue) and then leave for 15 mins at room temperature
5. Remove solution with Pastette then repeat Step 4 then leave for 15min
6. Extra wash step required to remove coomassie for some gel pieces. Remove solution with Pastette then repeat Step 4 then leave for 15min
7. Remove solution with Pastette and wash with 200µL Acetonitrile (Fisher A/0626/BP17 Batch No 1349248) (vortex briefly) to remove excess aqueous solutions.
8. Remove solution with Pastette and shrink plug by washing in 200µL Acetonitrile (Fisher A/0626/BP17 Batch No 1349248) for 10 mins.
9. Remove solution with Pastette and air dry for 10mins to remove all of the acetonitrile with lids open.
10. Prepared 50 mM Ammonium Bicarbonate (7mL 400mM Ammonium Bicarbonate in 49mL MilliQ Water.
11. Prepared 10mM DTT (37.6mg DTT (Sigma D9163 Batch No 128K10933) into 24.4mL 50mM Ammonium Bicarbonate
12. Add 200µL 10mM DTT in 50mM Ammonium Bicarbonate: Incubate for 30 min @ 60°C PCR machine 3 (E418)
13. Prepared 100mM Iodoacetamide in 50mM Ammonium bicarbonate (0.130g Sigma I1140 Batch No 057K53011) into 7.027mL 50mM Ammonium Bicarbonate.
14. Remove liquid with pastette (do not shrink gel) and add 200µL 100mM Iodoacetamide in 50mM Ammonium bicarbonate soln: Incubate for 30 min @ room temp in dark (in ProPick).
15. Remove liquid with pastette. Wash gel plugs in 200µL solution B then leave for 15min (to equilibrate to pH 8ish)
16. Remove liquid with pastette. Wash gel plugs in 200µL solution B - then leave for 15min (to equilibrate to pH 8ish)

17. Remove solution with Pastette and wash with 200 μ L Acetonitrile (Fisher A/0626/BP17 Batch No 1349248) to remove aqueous solutions.
18. Remove solution with Pastette and shrink plug by washing in 200 μ L Acetonitrile (Fisher A/0626/BP17 Batch No 1349248) for 10 mins
19. Remove solution with Pastette and air dry for 10 mins to remove all of the acetonitrile with lids open.
20. Prepared 10mM Ammonium Bicarbonate (1 mL 400mM Ammonium Bicarbonate added to 39mL MilliQ water).
21. 10mM Ammonium Bicarbonate (1 x 0.495mL) added to 1 x 5 μ L (5 μ g) Trypsin Gold (Promega V528A in 50mM Acetic Acid (prepared 15-Mar-13). i.e. 10ng/ μ L.
22. Add 10 μ L Trypsin to make gel piece turn clear from white. Used fresh tip for each addition. A further 10 μ L of 10mM ABC was added to ensure gel pieces remained in liquid.
23. Samples incubated at 37°C using PCR Machine 1 (W311) for 3 hrs.
24. Prepared 1% Formic acid (0.40mL Formic Acid (Sigma F0507 - batch No 026K0707) to 40 mL with MilliQ water).
25. Add 20 μ L of 1% formic acid (same volume as the total Trypsin + Buffer, agitate by tapping tube, and stand for 10 min before freezing in dry ice until Orbi Prepping.
26. **Samples prepared for Orbitrap analysis.** Samples thawed and the digest solution removed from tube using a pipette into a labelled 1.5mL Eppendorf tube. Used fresh tip for each sample.
27. The gel pieces were then washed with 40 μ L of 50% Acetonitrile (Fisher A/0626/BP17 Batch No 1221667) (volume= Trypsin + Buffer + 1% Formic Acid) to recover more digest peptides from the gel.
28. After short vortex the tubes were left to stand for 10mins at room temp before being removed and put into the labelled Eppendorf tube (used in Step 26). Used fresh tip for each sample.
29. The extracted digest samples were then dried down at the Low Drying setting (No heat) on a Speed Vac SC110 (Savant) fitted with a Refrigerated Condensation Trap and a Vac V-500 (Buchi). The samples were then frozen on dry ice and then stored at -80°C until ready for Orbitrap analysis.

Orbitrap analysis

Time points, 0 mins, 10 mins and 20-60 mins (pooled) were selected to investigate the presence of ROSEA protein in the digest at those time points.

The samples were dissolved in 70 μ L 0.5% Formic Acid, vortexing for approx 10 seconds then pulse centrifuging liquid to the bottom of the tube. The samples were then sonicated in a sonicating water bath (Kerry Ultrasonic bath (W306)) for 10 mins before a further pulse centrifuge. 22 μ L of the sample was added to the appropriate well on a Thermo-Fast 96 well skirted low profile plate used for samples in the NanoAquity autosampler. After additions the plate was sealed using Eppendorf self-adhesive PCR film. To limit issues of carryover the timepoint expected to have the least intact ROSEA protein was run first and a blank was run between each sample.

The analysis was performed using a LTQ-Orbitrap mass spectrometer (Thermo) and a nanoflow-HPLC system (nanoACQUITY: Waters). Peptides were trapped on line to a Symmetry C18 Trap (5 μ m, 180 μ m x 20 mm) which was then switched in-line to a UPLC BEH C18 Column, (1.7 μ m, 75 μ m x 250 mm) held at 45°C. Peptides were eluted by a gradient of acetonitrile in 0.1% formic acid over 72 min at a flow rate of 250 nL min⁻¹.

The mass spectrometer was operated in positive ion mode with a nano-spray source at a capillary temperature of 200°C. The Orbitrap was run with a resolution of 60,000 over the mass range m/z 300–1,800 and an MS target of 10^6 and 1 sec maximum scan time. The MS/MS was triggered by a minimal signal of 2000 with an Automatic Gain Control target of 30,000 ions and maximum scan time of 150 ms. For MS/MS events selection of 2+ and 3+ charge states were used. Dynamic exclusion was set to 1 count and 30 sec exclusion time with an exclusion mass window of ± 20 ppm.

Proteins were identified by searching the Thermo RAW files converted to Mascot generic format by Proteome Discover 1.1 (Thermo) and proteins were identified by interrogating the Uniprot (taxonomy *Viridiplantae*) using the MASCOT v2.4.1 search engine.

For this analysis Individual ions scores **>38** indicate **identity** or **extensive homology** ($p < 0.05$) (Ions score is $-10\log(P)$, where P is the probability that the observed match is a random event).

The Mascot data was collated into Scaffold v4.1.1 (Proteome Software) to allow comparison between samples

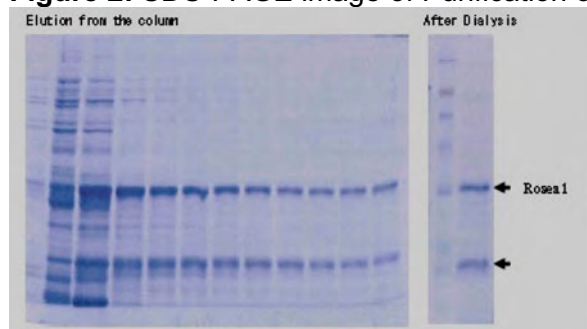
RESULTS

Purification of Delila and Rosea1

Figure 1: SDS-PAGE image of Purification of His-tag Delila



Figure 2: SDS-PAGE image of Purification of His-tag Rosea1



Digestion of Delila and Rosea1 recombinant proteins

Figure 3 . Delila digestion time course

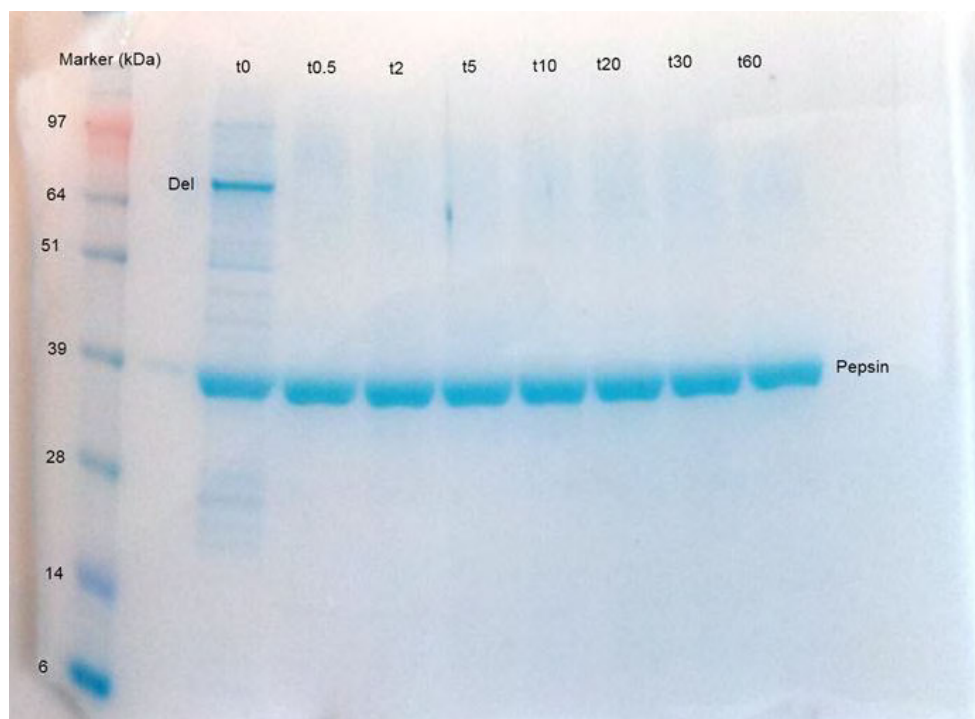
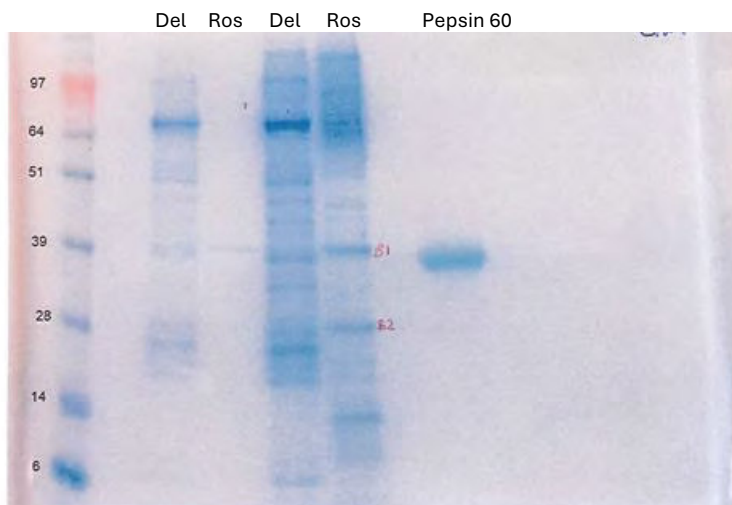


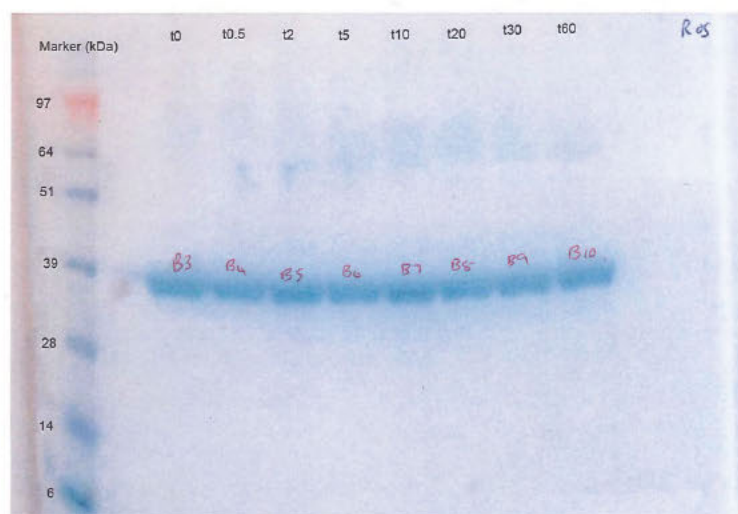
Figure 3 shows a gel of Delila (Del) protein digested with pepsin in SGF indicating that Del is rapidly digested with no full length protein detectable after 0.5 min.

Figure 4: Rosea Protein Gel 1



Gel assays did not allow us to monitor easily the degradation of the Rosea1 protein following protease digestion, due to pepsin running very close to the Rosea 1 protein (Band B1), so digestion of Rosea 1 was followed using Orbitrap mass spectrometry.

Figure 5: Rosea Digest Gel 2 – different digestion times (in min)



Bands B1 and B2 from Gel 1 (Figure 4) and B3-10 from Gel 2 (Figure 5) were cut out. Time points, 0 mins (B3), 10 mins (B7) and 20-60 mins pooled (B8-10) were selected to investigate the ROSEA protein in the digest at those time points. Samples B1 and B2 were also analyzed to confirm identity. To limit issues of carryover the timepoint (B10) expected to have the least intact ROSEA protein was run first (see Table 1 below).

Sample Name	Orbitrap Run time	Vol 0.5% Formic Acid samples dissolved	Vol loaded onto well	Orbi plate locale	Orbi injection volume	Data file name
Blank (20% Acetonitrile)	55		20µl	A01	3µl	FM130409_61
B10	01:22:00	70µl	22µl	A02	20µl	FM130409_62
Blank (20% Acetonitrile)	55		20µl	A03	3µl	FM130409_63
B7	01:22:00	70µl	22µl	A04	20µl	FM130409_64
Blank (20% Acetonitrile)	55		20µl	A05	3µl	FM130409_65
B3	01:22:00	70µl	22µl	A06	20µl	FM130409_66
Blank (20% Acetonitrile)	55		20µl	A07	3µl	FM130409_67
B1	01:22:00	70µl	22µl	A10	20µl	FM130409_70
Blank (20% Acetonitrile)	55		20µl	A11	3µl	FM130409_71

Table1: Orbitrap sample details

Mass spectroscopy analysis of digested B1 band (Not digested)

All proteins are identified using peptides whose Ions scores >38

Protein Hits (24 Proteins) – Rosea1 (highlighted in yellow) is the most abundant peptide

Protein View: Q2VAZ7

SubName: Full=ROSEA1;

Database: Uniprot
Score: 255
Nominal mass (M_r): 26122
Calculated pI: 8.97
Taxonomy: Antirrhinum majus

Search parameters

MS data file: \\Nbi-cfs1\ifr_proteomics\130904_SILAC\MarkW\MGFfiles\MarkW_B01_FM130904_70.mgf
Enzyme: Trypsin: cuts C-terminal side of K or R unless next residue is P
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Deamidated (NQ), Oxidation (M)

Protein sequence coverage: 42%

Matched peptides in mass spectroscopy analysis shown in **bold red**.

```
1  MEKNCRGVRK GTWTKEEDTL LRQCIEEYGE GKWHQVPHRA GLNRCRKSCR
51 LRWLNYLRPN IKRGRFSRDE VDLIVRLHKL LGNKWSLIAG RIPGRTANDV
101 KNFWNTHVGK NLGEDGERCR KNVMNTKTIK LTNIIVRPRAR TFTGLHVTWP
151 REVGKTDEFS NVRLTTDEIP DCEKQTQFYN DVASPDQDEVE DCIQWWSKLL
201 ETTEDEGELGN LFEEAQQIGN
```

Start	–	End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Peptide
10	–	22	526.2814	1575.823	1575.826	-2.02	2	46	0.0023	R.KGTWTKEDTLLR.Q
23	–	32	606.7631	1211.512	1211.513	-0.95	0	62	3.00E-05	R.QCIEEYGEGK.W
53	–	62	658.8776	1315.741	1315.74	0.39	0	40	0.0039	R.WLNYLRPNIK.R
66	–	76	674.8635	1347.713	1347.715	-1.6	1	47	0.0016	R.FSRDEVDLIVR.L
69	–	76	479.764	957.5134	957.5131	0.31	0	63	0.00033	R.DEVDLIVR.L
69	–	76	479.7641	957.5137	957.5131	0.63	0	52	0.0042	R.DEVDLIVR.L
85	–	91	401.7321	801.4496	801.4497	-0.099	0	38	0.13	K.WSLIAGR.I
102	–	110	551.7742	1101.534	1101.536	-1.52	0	39	0.0031	K.NFWNTHVGK.N
141	–	151	657.8513	1313.688	1313.688	-0.085	0	57	0.00015	R.TFTGLHVTWPR.E
152	–	163	690.8389	1379.663	1379.668	-3.56	1	49	0.00049	R.EVGKTDEFSNVR.L
152	–	163	690.8392	1379.664	1379.668	-3.04	1	44	0.0024	R.EVGKTDEFSNVR.L
164	–	174	660.8029	1319.591	1319.592	-0.25	0	74	3.70E-06	R.LTTDEIPDCEK.Q

Results of mass spectroscopy analysis of digested B3 band (Digestion t=0 mins)

All proteins are identified using peptides whose Ions scores >38

Protein Hits (12 Proteins) – Rosea1 (highlighted in yellow) is the most abundant peptide

Family	Member	Database	Accession	Score	Mass	Num. of matches	Num. of significant matches	Num. of sequences	Num. of significant sequences	emPAI	Description
1	1	Uniprot	Q2VAZ7	122	26122	5	5	5	5	0.83	SubName: Full=ROSEA1;
2	1	Uniprot	M4E270	90	71990	1	1	1	1	0.05	SubName: Full=Uncharacterized protein;
3	1	Uniprot	D8RIL3	48	37809	1	1	1	1	0.09	SubName: Full=Putative uncharacterized protein;
4	1	Uniprot	P85925	46	1393	1	1	1	1	3.07	RecName: Full=Unknown protein 18; Flags: Fragment;
5	1	Uniprot	K7U658	46	85140	2	2	1	1	0.04	SubName: Full=Uncharacterized protein;
6	1	Uniprot	I0Z6Z4	43	38739	1	1	1	1	0.09	SubName: Full=SGNH hydrolase;
7	1	Uniprot	M1C004	43	50594	1	1	1	1	0.07	RecName: Full=Phosphoglycerate kinase; EC=2.7.2.3;
8	1	Uniprot	C0PD32	43	53122	1	1	1	1	0.06	SubName: Full=Uncharacterized protein;
9	1	Uniprot	A4S3F3	41	43718	3	3	1	1	0.08	SubName: Full=Predicted protein;
10	1	Uniprot	D7L6M7	40	117265	1	1	1	1	0.03	SubName: Full=Transcription factor jumonji domain-containing protein;
11	1	Uniprot	E1UHX7	40	103778	1	1	1	1	0.03	SubName: Full=Putative auxin response factor 6/8;
12	1	Uniprot	R0H2U3	40	79601	1	1	1	1	0.04	SubName: Full=Uncharacterized protein;

Protein View: Q2VAZ7 SubName: Full=ROSEA1;

Database: Uniprot
Score: 122
Nominal mass (M_r): 26122
Calculated pI: 8.97
Taxonomy: Antirrhinum majus

Search parameters

MS data file: \\Nbi-cfs1\ifr_proteomics\130904 SILAC\Mark W\MGFfiles\MarkW_B03_FM130904_66.mgf
Enzyme: Trypsin: cuts C-terminal side of K or R unless next residue is P
Fixed modifications: Carbamidomethyl (C)
Variable modifications: Deamidated (NQ), Oxidation (M)

Protein sequence coverage: 23%

Matched peptides in mass spectroscopy analysis shown in **bold red**.

```

1 MEKNCRGVRK GTWTKEEDTL LRQCIIEEYGE GKWHQVPHRA GLNRCRKSCR
51 LRWLNLYLRPN IKRGRFSRDE VDLIVRLHLK LGNKWSLIAG RIPGRTANDV
101 KNFWNTHVGK NLGEDGERCR KNVMNTKTIK LTNIVRPRAR TFTGLHVTWP
151 REVGKTDEFS NVRLTTDEIP DCEKQTQFYN DVASPDQDEVE DCIQWWSKLL
201 ETTEDGELGN LFEEAQQIGN
  
```

Start	–	End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Peptide
10	–	22	526.2824	1575.825	1575.826	-0.16	2	51	0.0003	R.KGTWTKEEDTLR.Q
23	–	32	606.7634	1211.512	1211.513	-0.53	0	44	0.0063	R.QCIEEYGEGK.W
102	–	110	551.7756	1101.537	1101.536	0.91	0	38	0.01	K.NFWNTHVGK.N
156	–	163	484.2274	966.4403	966.4407	-0.4	0	60	0.00029	K.TDEFSNVR.L
164	–	174	660.8033	1319.592	1319.592	0.48	0	60	0.00012	R.LTTDEIPDCEK.Q

Results of mass spectroscopy analysis of digested B7 band (Digestion t=10mins)
 Protein Hits (10 proteins) – Rosea1 (highlighted in yellow) is the 7th most abundant peptide

Family	Member	Database	Accession	Score	Mass	Num. of matches	Num. of significant matches	Num. of sequences	Num. of significant sequences	emPAI	Description
1	1	Uniprot	P85925	109	1393	3	3	2	2	66.37	RecName: Full=Unknown protein 18; Flags: Fragment;
2	1	Uniprot	A4S3F3	58	43718	5	5	1	1	0.08	SubName: Full=Predicted protein;
3	1	Uniprot	I1J5E7	46	27583	1	1	1	1	0.12	SubName: Full=Uncharacterized protein;
4	1	Uniprot	D8RIL3	46	37809	2	2	1	1	0.09	SubName: Full=Putative uncharacterized protein;
5	1	Uniprot	K7U658	45	85140	3	3	1	1	0.04	SubName: Full=Uncharacterized protein;
6	1	Uniprot	M4EDF4	44	15909	1	1	1	1	0.21	SubName: Full=Uncharacterized protein;
7	1	Uniprot	Q2VAZ7	43	26122	2	1	2	1	0.13	SubName: Full=ROSEA1;
8	1	Uniprot	M1C004	43	50594	1	1	1	1	0.07	RecName: Full=Phosphoglycerate kinase; EC=2.7.2.3;
9	1	Uniprot	I0Z6Z4	40	38739	1	1	1	1	0.09	SubName: Full=SGNH hydrolase;
10	1	Uniprot	E1UHX7	38	103778	1	1	1	1	0.03	SubName: Full=Putative auxin response factor 6/8;

Protein View: Q2VAZ7

SubName: Full=ROSEA1;

Database: Uniprot
Score: 43
Nominal mass (M_r): 26122
Calculated pI: 8.97
Taxonomy: Antirrhinum majus

Search parameters

MS data file: \\Nbi-cfs1\ifr_proteomics\130904 SILAC\Mark W\MGFfiles\MarkW_B07_FM130904_64.mgf
Enzyme: Trypsin: cuts C-terminal side of K or R unless next residue is P
Fixed modifications: Carbamidomethyl (C)

Variable modifications: Deamidated (NQ), Oxidation (M)

Protein sequence coverage: 8%

Matched peptides in mass spectroscopy analysis shown in **bold red**.

```
1  MEKNCRGVRK  GTWTKEEDTL  LRQCIEEYGE GKWHQVPHRA  GLNRCRKSCR
51 LRWLNLYLRPN  IKRGRFSRDE VDLIVRLHKL  LGNKWSLIAG  RIPGRTANDV
101 KNFWNTHVGK  NLGEDGERCR  KNVMNTKTIK  LTNIVRPRAR  TFTGLHVTWP
151 REVGKTDEFS  NVRLTTDEIP  DCEKQTQFYN  DVASPQDEVE  DCIQWWSKLL
201 ETTEDEGELGN  LFEEAQQIGN
```

Notes:

There was clear evidence of digestion at 10 minutes because the ROSEA protein was not the top protein identified from the green plant database.

Start	–	End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Peptide
23	–	32	606.7635	1211.512	1211.513	-0.33	0	43	0.0022	R.QCIEEYGEGK.W
69	–	76	479.7632	957.5119	957.5131	-1.23	0	39	0.072	R.DEVDLIVR.L

Results of mass spectroscopy analysis of digested B8-10 band (Digestion t=20-60mins)
Protein Hits (15 proteins) – Rosea1 cannot be detected

Notes

Family	Member	Database	Accession	Score	Mass	Num. of matches	Num. of significant matches	Num. of sequences	Num. of significant sequences	emPAI	Description
1	1	Uniprot	P85925	80	1393	2	2	2	2	15.56	RecName: Full=Unknown protein 18; Flags: Fragment;
2	1	Uniprot	M4E270	63	71990	1	1	1	1	0.05	SubName: Full=Uncharacterized protein;
3	1	Uniprot	D8RIL3	52	37809	2	2	1	1	0.09	SubName: Full=Putative uncharacterized protein;
4	1	Uniprot	I1J5E7	46	27583	1	1	1	1	0.12	SubName: Full=Uncharacterized protein;
5	1	Uniprot	B4YPM8	45	28267	1	1	1	1	0.12	SubName: Full=HcrVf2-like protein; Flags: Fragment;
6	1	Uniprot	M0T8I2	44	56599	1	1	1	1	0.06	SubName: Full=Uncharacterized protein;
7	1	Uniprot	B0LOY2	44	28570	1	1	1	1	0.12	SubName: Full=Chloroplast chlorophyll a/b binding protein;
8	1	Uniprot	I1N7C5	44	7336	1	1	1	1	0.48	SubName: Full=Uncharacterized protein;
9	1	Uniprot	B9IEU3	44	59752	1	1	1	1	0.06	SubName: Full=Predicted protein;
10	1	Uniprot	K7U658	42	85140	1	1	1	1	0.04	SubName: Full=Uncharacterized protein;
11	1	Uniprot	M1C004	42	50594	1	1	1	1	0.07	RecName: Full=Phosphoglycerate kinase; EC=2.7.2.3;
12	1	Uniprot	R0H2U3	40	79601	1	1	1	1	0.04	SubName: Full=Uncharacterized protein;
13	1	Uniprot	I1L239	40	49876	1	1	1	1	0.07	SubName: Full=Uncharacterized protein;
14	1	Uniprot	A4S3F3	38	43718	1	1	1	1	0.08	SubName: Full=Predicted protein;
15	1	Uniprot	R0HCI6	38	49751	1	1	1	1	0.07	SubName: Full=Uncharacterized protein;

No significant ROSEA1 peptides could be identified after 20-60 min digestion with pepsin. Rosea 1 protein was substantially lost after 10 min and completely lost after 20-60 min digestion with pepsin. These data indicated that pepsin digestion was completely effective.

CONCLUSION

Delila (Del) and Rosea 1 (Ros1) proteins, produced in the *Del/Ros1*-N transformation event are both rapidly degraded in standard digestion tests.

This fact, and the lack of similarity to any known allergens, and the widespread distribution of proteins of the same classes in edible plants (see Appendix 8) provide ample evidence that neither the DEL nor ROS1 proteins are an allergenicity risk in the purple tomato.

Reference

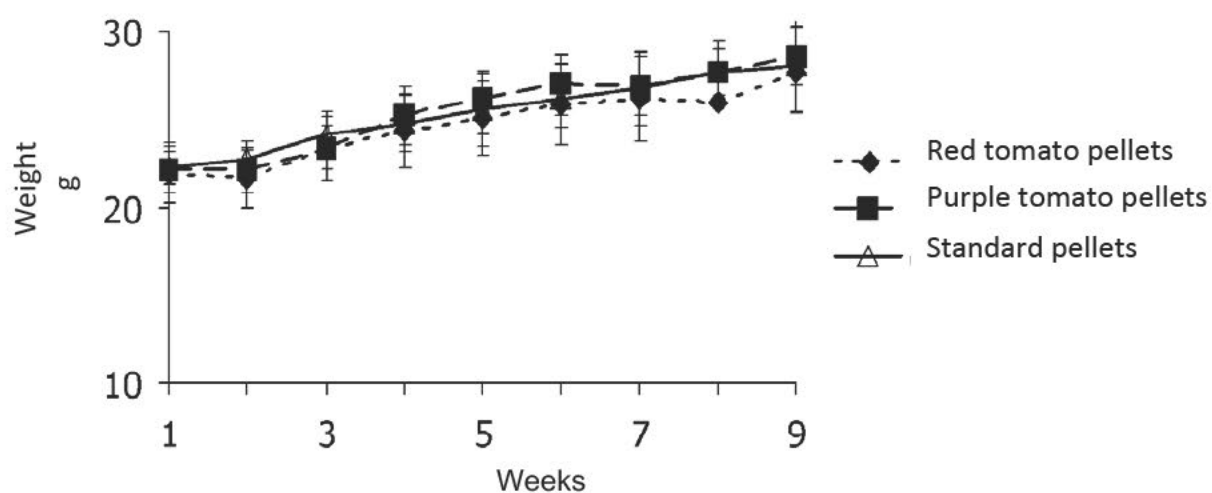
Thomas, K; Aalbers, M. Bannon, G.A. Bartels, M. Dearman, R.J. Esdaile, D.J. Fu, T.J. Glatt, C.M. Hadfield, N. & Hatzos, C. (2004) A multi-laboratory evaluation of a common in vitro pepsin digestion assay protocol used in assessing the safety of novel proteins. Regul. Toxicol. Pharm., 39: 87–98. <https://doi.org/10.1016/j.yrtph.2003.11.003>

Appendix 10: Feeding study to assess effect of purple tomatoes in diet on laboratory mice

To investigate whether the levels of anthocyanins achieved in *Del/Ros1-N* purple tomatoes impacted growth, development or behavior in a dietary context in laboratory mice, 'wild type' C57B16 mice were fed with complete feed 'custom formula' for rodents, either without supplement, or supplemented with 10% powder from freeze-dried red (control) tomatoes or purple (*Del/Ros1N*) tomatoes. The anthocyanin and carotenoid contents of the diet are detailed in Table 1.

Nutritional studies were undertaken at the European Institute of Oncology, Milano, Italy with wild type (C57/B16) laboratory mice fed diets supplemented with control (red) tomatoes or purple tomatoes (10% w/w) and compared to those fed the standard diet alone. No differences in food intake or weight gain compared to the standard diet were observed over 11 weeks (Figure 1; Butelli et al., 2008).

a



b

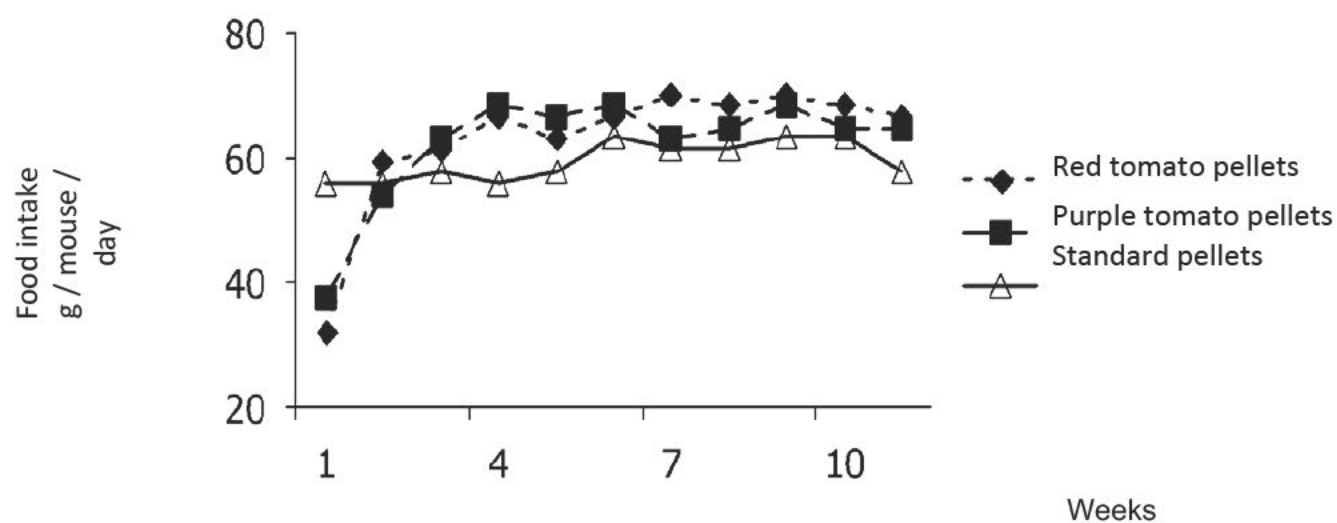


Figure 1: Body weight (a) and energy intake (b) of WT C57B16 mice fed with pellets of the standard diet, pellets supplemented with 10% red tomato powder and pellets supplemented with 10% purple tomato powder.

Table 1. Anthocyanins and carotenoid content of mouse food pellets, used in the experiment shown in Figure 1, supplemented with 10% wild type or *Del/Ros1-N* tomato powder. Results are expressed as µg/g of food, nd: not detected.

	WT tomato	<i>Del/Ros1 N</i> tomato
ANTHOCYANINS		
delphinidin 3(caffeoyl)-rutinoside-5-glucoside	n.d	17.16
delphinidin 3(cis-coumaroyl) rutinoside-5-glucoside	n.d	12.81
delphinidin 3(trans-coumaroyl) rutinoside-5-glucoside	n.d	208.15
delphinidin 3(feruloyl) rutinoside-5-glucoside	n.d	21.51
petunidin 3-(p-coumaroyl) rutinoside-5-glucoside	n.d	257.2
petunidin 3-(feruloyl) rutinoside-5-glucoside	n.d	25.07
Total anthocyanins	n.d	541.91
CAROTENOIDS		
α-carotene		
β-carotene	0.31 ± 0.07	0.80 ± 0.01
lycopene	0.20 ± 0.06	< 0.1
lutein	1.87 ± 0.20	2.18 ± 0.06
zeaxanthin	0.40 ± 0.19	0.42 ± 0.07
Total carotenoids	2.78 ± 0.52 µg/g	3.4 ± 0.14

For further information see **Butelli et al., (2008)**.

Appendix 12: Additional information on analyses of nutrient composition of tomatoes derived from event *Del/Ros1-N*

Stock *Del/Ros1-N* purple tomatoes in the “MoneyMaker” genetic background were compared to non-transformed (‘wild type’) control, MoneyMaker tomatoes. Representative composite samples of both types of tomatoes were analyzed for their content of a number of nutrients and non-nutrients, following the guidelines in: *OECD ENV/JM/MONO(2008)26; Consensus Document on Compositional Considerations for New Varieties of Tomato: Key food and feed nutrients, toxicants and allergens*.

Tomatoes were analyzed for:

Proximates – Protein, fat, total carbohydrates, fiber, ash

Minerals – Magnesium and potassium

Carotenoids – β -carotene and lycopene

Vitamins – C, K and folate (B9)

The samples were analyzed as follows:

Ten samples of tomato fruit were collected when “red ripe” from 5 red tomato plants and 5 purple tomato plants. For each sample, approx. 1 kg (10 fruits) were collected from individual plants (from each of the 5 red fruit plants and 5 purple fruit plants). Results were received individually (10 independent analyses) and samples were not combined. All ripe fruits were at the same stage of ripening and collected at the same time, from the same glasshouse at the John Innes Centre, Norwich UK, in 2018. No chemical pesticides, only biological control agents, are used in the greenhouses at the John Innes Centre, where the plants were grown and maintained for all analyses.

The fruit were washed in distilled water and then sent fresh to Eurofins for analysis. Composite samples were obtained by combining 10 fruits from different plants into 5 independent samples for *Del/Ros1-N* fruit (in MoneyMaker genetic background) and MoneyMaker red, control fruit. These 5 independent samples were used for analysis of all the nutrients.

The analytical methods utilized for compositional analysis by Eurofins (<https://www.eurofins.com/food-and-feed-testing/food-testing-services/nutritional-testing/>) are summarized in Table 1. All methods were performed to GLP and accredited through ISO17025. Technical summaries of each method are appended below for reference.

Table 1: Methodology utilized for compositional analysis by Eurofins

Analyte	Test code	SOP Number*	Method	Comments
Moisture	UD006	Q/005 H/007	Gravimetry	
Crude Protein	UD001	Z/001	N x 6.25	Dumas Method Chang, S.K. and Zhang, Y., (2017). Protein analysis. <i>Food analysis</i> , pp.315-331.
Ash	UD007	Q/001	Gravimetry	
Carbohydrate (available)	UD017	Q/035	Calculation	
Fructose	UD08W	CHROM/344	IC-PAD	Ion Chromatography- Pulse Amperometric Detection (Dionex)
Galactose	UD08W	CHROM/344	IC-PAD	
Glucose	UD08W	CHROM/344	IC-PAD	
Lactose	UD08W	CHROM/344	IC-PAD	
Maltose	UD08W	CHROM/344	IC-PAD	
Sucrose	UD08W	CHROM/344	IC-PAD	
Total sugars	UD08W	CHROM/344	IC-PAD	
Total dietary fibre (AOAC)	B7039	H/085	Enzymatic- gravimetry AOAC method 985.29	Neil H. Mermelstein Food Technology Magazine July 1st 2011 Analyzing for Dietary Fiber
Energy (kcal)	UD771	Q/035	Calculation for Humans	
Energy (kJ)	UD771	Q/035	Calculation for Humans	
Salt	UD815		via sodium x 2.5	
Monounsatur. FAs	UDFB1	CHROM/215	GC-FID	Gas Chromatography -Flame Ionization Detection
Polyunsatur. FAs	UDFB1	CHROM/215	GC-FID	
Satur. FAs	UDFB1	CHROM/215	GC-FID	
Trans FAs	UDFB1	CHROM/215	GC-FID	
Mg	UD028	ICP/003	ICP-OES	Inductively coupled Plasma - Optical Emission Spectrometry
K	UD029	ICP/003	ICP-OES	
Na	UD015	ICP/003	ICP-OES	
Beta-carotene	†A7270		Rp-HPLC-VIS/DAD	Reverse Phase High Pressure Liquid Chromatography - Visible/Diode Array Detection
Folate (vitamin B9)	†A7280		Microbiological assay	
Ascorbic acid (vitamin C)	†A7291		Rp-HPLC—UV/DAD	Reverse Phase High Pressure Liquid Chromatography- Ultra Violet/Diode Array Detection
Phylloquinone (vitamin K1)	†A7293		Rp-HPLC-FLD	Reverse Phase High Pressure Liquid Chromatography- Fluorescence Detection
Lycopene	†A7534		Spectrophotometry	
Total fat	†AAMG0		Acid hydrolysis	

*SOP GLP approved

Table 2: Statistical analysis of nutritional composition of red and purple tomatoes. Microsoft 365 Excel was used for statistical analysis. Data were analysed using an unpaired Student's t-test (using 5 independent samples of 1 kg fruit. Symbol Meaning: ns = P > 0.05; * = P ≤ 0.05; ** = P ≤ 0.01; *** = P ≤ 0.001; **** = P ≤ 0.0001

Analyte	units	RED	PURPLE	T-test
moisture	g/100g	94.72 ± 0.12	95.1 ± 0.12	ns
Crude protein	g/100g	0.64 ± 0.05	0.7 ± 0.09	ns
Ash	g/100g	0.32 ± 0.02	0.48 ± 0.02	***
CHO	g/100g	3.06 ± 0.29	3.26 ± 0.27	ns
Fructose	g/100g	1.46 ± 0.05	1.22 ± 0.09	ns

Galactose	g/100g	<0.1 ± 0	0.1 ± 0	ns
Glucose	g/100g	1.26 ± 0.02	1.04 ± 0.08	*
Lactose	g/100g	<0.1 ± 0	0.1 ± 0	ns
Maltose	g/100g	<0.1 ± 0	0.1 ± 0	ns
Sucrose	g/100g	<0.1 ± 0	0.1 ± 0	ns
Total sugar	g/100g	2.72 ± 0.06	2.24 ± 0.17	*
Total fibre AOAC	g/100g	<1.1 ± 0.19	0.66 ± 0.07	ns
Energy kcal	kcal/100g	18.4 ± 0.83	16.8 ± 0.77	ns
Energy kj	kJ/100g	81.8 ± 1.78	71 ± 2.95	*
Total fat	g/100g	<0.3 ± 0.04	0.3 ± 0	ns
Salt	g/100g	<0.025 ± 0	0.025 ± 0.003	ns
Mono unsat FAs	g/100g	<0.1 ± 0	0.1 ± 0	ns
Poly unsat FAs	g/100g	<0.1 ± 0.03	0.1 ± 0	ns
Sat Fas	g/100g	<0.1 ± 0	0.1 ± 0	ns
Trans FAs	g/100g	<0.1 ± 0	0.1 ± 0	ns
Mg	g/100g	0.006 ± 0.0004	0.008 ± 0.0003	*
K	g/100g	0.16 ± 0.0038	0.23 ± 0.0113	***
Na	g/100g	<0.01 ± 0	0.01 ± 0.0013	ns
beta carotene	ug/100g	451.6 ± 42.43	661.8 ± 64.79	*
Folate B9	ug/100g	8.92 ± 0.28	14.22 ± 0.31	****
Ascorbate C	mg/100g	6.86 ± 0.18	8.1 ± 0.86	ns
Phylloquinone K1	ug/100g	2.69 ± 0.14	1.99 ± 0.1	**
Lycopene	mg / kg	n/a	20.9	na

In the case where a sample was <X and other samples had a measurable number, the <X values were recorded as zero for statistical purposes and analysis of variance was applied to the samples. Regardless, none of the analytes in this category are different from values seen in wild-type tomatoes, so we regard the Purple Tomato as no different from any other tomato with respect to these components.

Table 3: Chlorogenic acid amounts in common foods (from <http://phenol-explorer.eu/contents/polyphenol/475>)

Phenolic acids - Hydroxycinnamic acids - 5-Caffeoylquinic acid	
Sunflower seed, meal	454.48 mg/100 g FW
Globe artichoke, heads, raw	202.23 mg/100 g FW
Chicory [Red], raw	168.15 mg/100 g FW
Highbush blueberry, raw	131.18 mg/100 g FW
Burdock root, raw	126.76 mg/100 g FW
Coffee beverage [Filter], decaffeinated	116.85 mg/100 ml
Chicory [Green], raw	100.88 mg/100 g FW
Lowbush blueberry, raw	86.99 mg/100 g FW
Robusta coffee beverage [Filter]	75.80 mg/100 ml
Coffee beverage [Filter]	70.03 mg/100 ml
Loquat	54.75 mg/100 g FW
Rowanberry, pure juice	54.40 mg/100 ml
Apple [Cider], peeled	49.33 mg/100 g FW
Arabica coffee beverage [Filter]	43.09 mg/100 ml
Plum, prune	38.79 mg/100 g FW
Plum, prune, juice from concentrate	36.40 mg/100 ml
Spearmint, dried	31.00 mg/100 g FW
Oregano, dried (wild marjoram)	28.70 mg/100 g FW
Potato, raw	26.73 mg/100 g FW
Apple [Cider], pure juice	21.45 mg/100 ml
Plum, prune, pure juice	20.35 mg/100 ml
Common sage, dried	19.85 mg/100 g FW
Apple [Dessert], peeled	18.22 mg/100 g FW
Peach, whole	15.55 mg/100 g FW
Pear, pure juice	15.03 mg/100 ml
Apple [Dessert], whole, raw	13.37 mg/100 g FW
Pear, whole	11.27 mg/100 g FW
Carrot, raw	8.88 mg/100 g FW

Food Testing

Legend

(#)	The recognition of this test depends on the type of sample.
CAS	The CAS Registry Number is a unique identifier assigned by the Chemical Abstracts Service to chemical substances.
LOQ	Limit of quantification.
RL	Reporting limit.
LOD	Limit of detection.
MLOQ	Maximum limit of quantification.
MU	Measurement uncertainty.
MSPT	Maximum Sample Preservation Time

A7534-3
Lycopene

DO NOT USE THIS TEST

Aim Lycopene is a carotenoid responsible for the red colour of the tomato.
Tomato peel is 5 times more concentrated in lycopene than the pulp.

Technique Spectrophotometry

Method Extraction and quantification by spectrophotometry. In order to declare conformity to existing regulations or customer specifications, the uncertainty associated with the result will be added or removed in such a way that the result can be interpreted in any case regarding the specifications or regulations in force. It will not be taken into account in case of standards which already incorporate the measurement uncertainties.

Method reference Internal

Applied on Tomato based products
Other matrices: please contact us
Note : only applicable to naturally content lycopene

Laboratory Eurofins Analytics France (Nantes) Not accredited

Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Lycopene	502-65-8	2 mg/l	2 mg/l			18.7	

Sampling / shipping requirements	Type of sample	Amount of sample optimal/minimal	Container	MSPT
	All applicable samples	50 g / 25 g	Suitable container	-

A7293-1 (HI)
Vitamin K1

Aim 'Determines the content of phyloquinone. Results are expressed in µg phyloquinone /100 g.
Vitamin K1 (phyloquinone) is formed in the chloroplasts of green plants and found in green plants and their fruits.

Technique LC-FLD

Method 'Vitamin K1 is enzymatically treated to remove fat and extracted by n-hexane. The determination is carried out

08/01/2025

using rp-HPLC with post-column reduction and fluorimetric detection (Ex.: 243 nm, Em.: 430 nm)

Method reference	EN 14148:2003 mod.						
Applied on	Food, feed and nutritional supplements						
Laboratory	Eurofins Vitamin Testing Denmark			(HI) DS EN ISO/IEC 17025 DANAK 581			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility
	Phylloquinone (vitamin K1)	84-80-0	0.80 µg/100 g	0.80 µg/100 g		600 µg/100 g	10.0
Sampling / shipping requirements	Type of sample	Amount of sample optimal/minimal		Container		MSPT	
	All applicable samples	50 g / 20 g		Suitable container		-	

A7280-1 (DF)
Vitamin B6

DO NOT USE THIS TEST

Technique	Microbiology						
Method	Microbiology						
Method reference	LST AB 269.1						
Applied on	Food, feeds						
Laboratory	Eurofins Vitamin Testing Denmark			(DF) DS EN ISO/IEC 17025 DANAK 168			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility
	Vitamin B6 (pyridoxine)	65-23-6	0.1 mg/100 g	0.1 mg/100 g			
Sampling / shipping requirements	Type of sample	Amount of sample optimal/minimal		Container		MSPT	
	All applicable samples	200 g / 50 g		Suitable container		-	

AAMG0-1 (R)(UC)
Total fat (microwave method)

Aim	Determination of total fat						
Technique	Gravimetry [Microwave method]						
Method	Hydrolysis and extraction of fat by sulfuric acid and cyclohexane. Total fat is determined by gravimetry after evaporation of cyclohexane. In order to declare conformity to existing regulations or customer specifications, the uncertainty associated with the result will be added or removed in such a way that the result can be interpreted in any case regarding the specifications or regulations in force. It will not be taken into account in case of standards which already incorporate the measurement uncertainties.						
Method reference	Internal			All applicable types of sample			

08/01/2025

Food Testing

Applied on	Food, Feed except: dairy products, surimi, additives, yeast, algae products, spices, coffee/tea/herbal tea, alcoholic beverages (except syrup, Baileys'milk and pina colada), non alcoholic beverage (except soy milk and coconut milk), premix for animals., vegan dishes, dry mushroom, nougat, oilseeds,animal meal as flour, vinegar and derivatives, mustard and derivatives, nuts and confectionary with nut inclusions, butter or vegetable fat with more than 80%fat							
Laboratory	Eurofins Analyses Nutritionnelles France (FR)				Recognition depends on the type of sample			
Scope of recognitions	Type of sample		Recognition					
	Dairy product		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Matrices: is feed, COFRAC feed matrix family is Complete or complementary composed feed		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Matrices: is feed, COFRAC feed matrix family is Raw material		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Matrices: is food, COFRAC food matrix family is Cereal products		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Matrices: is food, COFRAC food matrix family is Composed food		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Matrices: is food, COFRAC food matrix family is Egg products		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Matrices: is food, COFRAC food matrix family is Fat products		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Matrices: is food, COFRAC food matrix family is Fishing products		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Matrices: is food, COFRAC food matrix family is Fruits and vegetables		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Matrices: is food, COFRAC food matrix family is Meat products		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Matrices: is food, COFRAC food matrix family is Sweet and sweetened products		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Nut, nut product		Not accredited					
	Nut, nut product: physical state is liquid or viscous/oily		(UC) COFRAC TESTING (scope on www.cofrac.fr) 1-7085					
	Seeds, oilseeds and seed milling product: seed or grain		Not accredited					
	All other applicable types of sample		Not accredited					
	Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility
Total fat		57-10-3	0.3 g/100 g	0.3 g/100 g			R = 0.28 + racine (6% * teneur)	U = 0.2 + racine (3% * teneur)
Sampling / shipping requirements	Type of sample		Amount of sample optimal/minimal		Container		MSPT	
	Fish product, seafood product		200 g / 100 g		Suitable container		-	
	Meat, meat product		200 g / 100 g		Suitable container		-	
	Terrestrial snail, terrestrial snail derived product		200 g / 100 g		Suitable container		-	
	All other applicable samples		100 g / 50 g		Suitable container		-	
Related / preparation	For preparation	AAPCA-1		Meat and fish products preparation	Grinding			

08/01/2025

tests (see appendix)	Internal calculation test	Y5IBL-1	Internal control sum of nutriments	Calculation
-------------------------	------------------------------	---------	---------------------------------------	-------------

08/01/2025

Appendix

AAPCA-1		Meat and fish products preparation						
Aim	Preparation according to NF V04-416							
Technique	Grinding							
Method	Meat product preparation standard NF V04-416 specifies that 200 g of product should be ground for analysis, to guarantee homogeneity. If this is not the case, the AAPCA test appears, followed by the words "insufficient quantity". This ensures accredited results even if we deviate from the preparation standard.							
Method reference	Internal				All applicable types of sample			
Applied on	Meat and Fish products							
Laboratory	Eurofins Analyses Nutritionnelles France (FR)				Not accredited			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Sample quantity control as per NF V 04-416							
	Weight of the sample analyzed							

Y5IBL-1		Internal control sum of nutriments	
Technique	Calculation		
Applied on	Food		
Laboratory	Eurofins Analyses Nutritionnelles France (FR)		Not accredited

08/01/2025

Food Testing

Legend

(§)	The recognition of this test depends on the parameters.
CAS	The CAS Registry Number is a unique identifier assigned by the Chemical Abstracts Service to chemical substances.
LOQ	Limit of quantification.
RL	Reporting limit.
LOD	Limit of detection.
MLOQ	Maximum limit of quantification.
MU	Measurement uncertainty.
MSPT	Maximum Sample Preservation Time

A7270-1 (§)(HI)		Beta-carotene (trans-), (Pro-vitamin A), juice and vegetables.							
Aim	Beta-carotene have a vitamin A activity. 1 µg retinol (RE) = 6 µg trans-beta-carotene. The vitamin A is present as beta-carotene in plants and as retinols in animal origin products. This test is recommended for products origin from the plant kingdom. For some meat or dairy products or ready complete meals it can be necessary to mesure both beta-carotene and retinols (test A7272). Synonyms: Pro-vitamin A.								
Technique	LC-DAD								
Method	The sample is pretreated with a proteolytic enzyme and extracted three times with a solution of ethanol. The separation is performed by RP HPLC on a C30 column and detected by a diode array detector (DAD) at 450 nm. All-trans β-carotene is quantified from a 3-point calibration curve.								
Method reference	EN 12823-2:2000								
Applied on	Fruit, fruit juices, vegetables and samples low in fat (< 3%).								
Laboratory	Eurofins Vitamin Testing Denmark (HI) (§) DS EN ISO/IEC 17025 DANAK 581								
Parameters	Accred.	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	-	beta-Carotene (all-trans)	7235-40-7	5 µg/100 g	5 µg/100 g			14.0	
	-	cis-beta-Carotene	7235-40-7	5.00 µg/100 g	5.00 µg/100 g				
	-	Total beta-Carotene		5.00 µg/100 g	5.00 µg/100 g				
Sampling / shipping requirements	Type of sample		Amount of sample optimal/minimal		Container		MSPT		
	All applicable samples		100 g / 25 g		Suitable container		-		

06/06/2024

Legend

CAS	The CAS Registry Number is a unique identifier assigned by the Chemical Abstracts Service to every chemical substance.
LOQ	Limit of quantification.
RL	Reporting limit.
LOD	Limit of detection.
MLOQ	Maximum limit of quantification.
MU	Measurement uncertainty.

A7291-1 (HI) Vitamin C (Ascorbic acid + dehydroascorbic acid)

Technique	LC-DAD ()						
Method	The sample is extracted in an aqueous solution containing trichloroacetic acid (TCA) and the antioxidant tris(2-carboxyethyl)phosphine (TCEP). Vitamin C is degraded in a basic environment, which the presence of TCA prevents. TCA contributes additionally to trap dissolved proteins. TCEP reduces dehydroascorbic acid (DHAA) to ascorbic acid (AA), so DHAA is also determined and also prevents the degradation of vitamin C. The final extract is analysed by RP-HPLC with UV detection at 265 nm. The Iso-ascorbic acid will be seen as a peak after the ascorbic acid peak.						
Method reference	Food Chemistry, 94 (2006) 626-631						
Applied on	Food and feed.						
Laboratory	Eurofins Vitamin Testing Denmark			(HI) DS EN ISO/IEC 17025 DANAK 581			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility MU
	Ascorbic acid (vitamin C)	50-81-7	0.50 mg/100 g	0.50 mg/100 g			5.0
Test instructions	Send samples in the original closed packaging for this test.						

16/02/2021

Food Testing

Legend

CAS	The CAS Registry Number is a unique identifier assigned by the Chemical Abstracts Service to chemical substances.
LOQ	Limit of quantification.
RL	Reporting limit.
LOD	Limit of detection.
MLOQ	Maximum limit of quantification.
MU	Measurement uncertainty.

A7293-1 (HI)		Vitamin K1					
Aim	'Determines the content of phyloquinone. Results are expressed in µg phyloquinone /100 g. Vitamin K1 (phyloquinone) is formed in the chloroplasts of green plants and found in green plants and their fruits.						
Technique	LC-FLD						
Method	'Vitamin K1 is enzymatically treated to remove fat and extracted by n-hexane. The determination is carried out using rp-HPLC with post-column reduction and fluorimetric detection (Ex.: 243 nm, Em.: 430 nm)						
Method reference	EN 14148:2003 mod.						
Applied on	Food, feed and nutritional supplements						
Laboratory	Eurofins Vitamin Testing Denmark			(HI) DS EN ISO/IEC 17025 DANAK 581			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility MU
	Phyloquinone (vitamin K1)	84-80-0	0.80 µg/100 g	0.80 µg/100 g		600 µg/100 g	10.0
Sampling / shipping requirements	Optimal quantity	50 g	Minimal quantity	20 g	Container	Suitable container (ORPA)	

11/8/2021

Food Testing

Legend

CAS	The CAS Registry Number is a unique identifier assigned by the Chemical Abstracts Service to chemical substances.
LOQ	Limit of quantification.
RL	Reporting limit.
LOD	Limit of detection.
MLOQ	Maximum limit of quantification.
MU	Measurement uncertainty.

PUD8B-1		Basic Group 2	
Applied on	Food (not Dairy)		
Content (see appendix)	B7039-5 (YS)	Total Dietary Fibre (AOAC 991.43)	Enzymatic-gravimetry
	UD001-1 (YS)	Crude Protein (Nx6.25) (Dumas)	Dumas (TCD)
	UD003-1 (YS)	Total Fat	Gravimetry
	UD006-3 (YS)	Moisture	Gravimetry
	UD007-1 (YS)	Ash	Gravimetry
	UD015-1 (YS)	Sodium	ICP-OES
	UD017-3 (YS)	Available Carbohydrate	Calculation
	UD08W-2 (YS)	Sugar Profile	IC-PAD
	UD771-1 (YS)	Energy (Human)	Calculation
	UD815-4 (YS)	Salt (calculated via sodium x 2.5)	Calculation
	UDFA1-1 (YS)	Fatty Acids (Sats/Monos/Polys/Total Trans) (normalised)	GC-FID
	UDFB1-1 (YS)	Fatty Acids (Sats/Monos/Polys/Total Trans) (mass basis)	Calculation
	UDRT1-1	ICP-OES Run Test (Food/Feed)	ICP-OES

4/1/2021

Appendix

B7039-5 (YS) Total Dietary Fibre (AOAC 991.43)								
Aim	Enzymatic-Gravimetric Method							
Method	Samples are analysed singularly. The first stage for all samples is that fat extraction and sugar extraction is performed prior to undergoing sequential enzymatic digestion by heat stable alpha-amylase, protease and amyloglycosidase to remove starch and protein. For total dietary fibre (TDF), enzyme digestate is treated with alcohol to precipitate soluble dietary fibre before filtering, and weighed. TDF residue values are corrected for protein, ash, and blank values.							
Applied on	Food and food products							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)			(YS) BS EN ISO/IEC 17025:2017 UKAS 0342				
Parameters	The information given below may depend on the type of sample.							
	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Total Dietary Fibre (AOAC 991.43)		0.5 g/100 g	0.5 g/100 g				

UD001-1 (YS) Crude Protein (Nx6.25) (Dumas)								
Applied on	Food / Feed							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)				(YS) BS EN ISO/IEC 17025:2017 UKAS 0342			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Crude Protein (Nx6.25) (Dumas)		0.1 g/100 g	0.1 g/100 g				

UD003-1 (YS)		Total Fat						
Method	Acid hydrolysis. The sample is digested with hydrochloric acid,filtered and dried then extracted with petroleum ether.							
Applied on	Food, petfood (not suitable for butter or butter type samples)							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)				(YS) BS EN ISO/IEC 17025:2017 UKAS 0342			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibili ty	MU
	Total fat		0.5 g/100 g	0.5 g/100 g	0.1 g/100 g			

UD006-3 (YS) Moisture	
Method	The sample is dried in a fan oven at 102±2 °C for a minimum of six hours but no longer than 16hrs
Applied on	Food / Feed (with sugar content of <20%)
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton) (YS) BS EN ISO/IEC 17025:2017 UKAS 0342

4/1/2021

Food Testing

Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Moisture		0.1 g/100 g	0.1 g/100 g				

UD007-1 (YS)		Ash						
Method	The sample is ashed in a muffle furnace at 550 degrees celcius for 16 hours.							
Applied on	Food / Feed							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)				(YS) BS EN ISO/IEC 17025:2017 UKAS 0342			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Ash		0.1 g/100 g	0.1 g/100 g				

UD015-1 (YS)		Sodium						
Method	The sample is microwave digested in acid and determined by ICP-OES.							
Applied on	Food / Feed							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)			(YS) BS EN ISO/IEC 17025:2017 UKAS 0342				
Parameters	The information given below may depend on the type of sample.							
	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Sodium	7440-23-5	0.01 g/100 g	0.01 g/100 g	0.002 g/100 g			

UD017-3 (YS) Available Carbohydrate								
Applied on	Food							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)				(YS) BS EN ISO/IEC 17025:2017 UKAS 0342			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Carbohydrates (available)		1 g/100 g	1 g/100 g				

UD08W-2 (YS)		Sugar Profile						
Method	A suitable test portion is subjected to aqueous extraction, clarification and filtration. The extract test solution is analysed by High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection							
Applied on	all sample types - not suitable for raw flour							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)				(YS) BS EN ISO/IEC 17025:2017 UKAS 0342			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Fructose	57-48-7	0.1 g/100 g	0.1 g/100 g	0.1 g/100 g			
	Galactose	59-23-4	0.1 g/100 g	0.1 g/100 g	0.1 g/100 g			

4/1/2021

Food Testing

Glucose	50-99-7	0.1 g/100 g	0.1 g/100 g	0.1 g/100 g
Lactose	63-42-3	0.1 g/100 g	0.1 g/100 g	0.1 g/100 g
Maltose	69-79-4	0.1 g/100 g	0.1 g/100 g	0.1 g/100 g
Sucrose	57-50-1	0.1 g/100 g	0.1 g/100 g	0.1 g/100 g
Total sugars		0.1 g/100 g	0.1 g/100 g	0.1 g/100 g

UD771-1 (YS) Energy (Human)

Applied on	Food						
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton) (YS) BS EN ISO/IEC 17025:2017 UKAS 0342						
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility MU
	Energy value (kcal)		0.1 kcal/100 g	0.1 kcal/100 g			
	Energy value (kJ)		0.1 kJ/100 g	0.1 kJ/100 g			

UD815-4 (YS) Salt (calculated via sodium x 2.5)

Method	Salt calculated via sodium. Sodium analysis performed by ICP-OES						
Applied on	Food, Feed						
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton) (YS) BS EN ISO/IEC 17025:2017 UKAS 0342						
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility MU
	Salt (via sodium x 2.5)		0.025 g/100 g	0.025 g/100 g	0.005 g/100 g		

UDFA1-1 (YS) Fatty Acids (Sats/Monos/Polys/Total Trans) (normalised)

Applied on	Food (mass basis)						
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton) (YS) BS EN ISO/IEC 17025:2017 UKAS 0342						
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility MU
	Monounsaturated fatty acids		0.02 g/100 g fat	0.02 g/100 g fat			
	Polysaturated fatty acids		0.02 g/100 g fat	0.02 g/100 g fat			
	Saturated fatty acids		0.02 g/100 g fat	0.02 g/100 g fat			
	Trans fatty acids		0.02 g/100 g fat	0.02 g/100 g fat			

UDFB1-1 (YS) Fatty Acids (Sats/Monos/Polys/Total Trans) (mass basis)

Applied on	Food (mass basis)						
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton) (YS) BS EN ISO/IEC 17025:2017 UKAS 0342						

4/1/2021

Food Testing

Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Monounsaturated fatty acids		0.1 g/100 g	0.1 g/100 g				
	Polyunsaturated fatty acids		0.1 g/100 g	0.1 g/100 g				
	Saturated fatty acids		0.1 g/100 g	0.1 g/100 g				
	Trans Fatty Acids		0.1 g/100 g	0.1 g/100 g				

UDRT1-1 ICP-OES Run Test (Food/Feed)								
Applied on	Food / Feed							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)				Not accredited			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Analytical run test							

4/1/2021

Legend

(#)	The recognition of this test depends on the type of sample.
CAS	The CAS Registry Number is a unique identifier assigned by the Chemical Abstracts Service to chemical substances.
LOQ	Limit of quantification.
RL	Reporting limit.
LOD	Limit of detection.
MLOQ	Maximum limit of quantification.
MU	Measurement uncertainty.
MSPT	Maximum Sample Preservation Time

UD028-2 (v)(YS)		Magnesium						
Technique	ICP-OES							
Method reference	Internal Method ICP/003							
Applied on	Food / Feed (test unsuitable for essential oils)							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)				Recognition depends on the type of sample			
Scope of recognitions	Not accredited				Essential oils			
	(YS) BS EN ISO/IEC 17025:2017 UKAS 0342				Food / Feed			
	Not accredited				All other applicable types of sample			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Magnesium	7439-95-4	0.0005 g/100 g	0.0005 g/100 g				
Sampling / shipping requirements	Type of sample		Amount of sample optimal/minimal		Container		MSPT	
	All applicable samples		100 g / 50 g		Suitable container		-	
Related / preparation tests (see appendix)	All samples	UDRT1-1	ICP-OES Run Test (Food/Feed)		ICP injection			

04/10/2024

Appendix

UDRT1-1		ICP-OES Run Test (Food/Feed)						
Technique	ICP injection							
Applied on	Food / Feed							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)				Not accredited			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Analytical run test							

04/10/2024

Food Testing

Legend

CAS	The CAS Registry Number is a unique identifier assigned by the Chemical Abstracts Service to chemical substances.
LOQ	Limit of quantification.
RL	Reporting limit.
LOD	Limit of detection.
MLOQ	Maximum limit of quantification.
MU	Measurement uncertainty.

UD029-1 (YS)		Potassium						
Technique	ICP-OES							
Method reference	Internal							
Applied on	Food / Feed (test unsuitable for essential oils)							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)				(YS)	BS EN ISO/IEC 17025:2017 UKAS 0342		
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Potassium	7440-09-7	0.005 g/100 g	0.005 g/100 g				
Sampling / shipping requirements	Type of sample	Amount of sample			Container			
	All applicable samples	100 g / 50 g			Suitable container			
Related / preparation tests (see appendix)	All samples	UDRT1-1		ICP-OES Run Test (Food/Feed)		ICP injection		

6/17/2022

Appendix

UDRT1-1		ICP-OES Run Test (Food/Feed)						
Technique	ICP injection							
Applied on	Food / Feed							
Laboratory	Eurofins Food Testing UK Ltd (Wolverhampton)				Not accredited			
Parameters	Parameter	CAS	LOQ	RL	LOD	MLOQ	Reproducibility	MU
	Analytical run test							

6/17/2022