Project title:	Selection and testing of putative 'crumbly' fruit molecular markers for both diagnostic and breeding purposes in *Rubus idaeus*
Project number:	SF 167 (extension for marker development)
Project leader:	Julie Graham – The James Hutton Institute
Report:	Final report
Previous report:	PhD thesis Luca Scolari (SF 167)
Key staff:	Luca Scolari
Location of project:	Invergowrie, Dundee, DD2 5DA, Scotland, UK
Industry Representative:	Louise Sutherland Freiston Associates Ltd.
Date project commenced:	December 2019

DISCLAIMER

While the Agriculture and Horticulture Development Board seeks to ensure that the information contained within this document is accurate at the time of printing, no warranty is given in respect thereof and, to the maximum extent permitted by law the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

[Name]

[Position]

[Organisation] Julie Graham Nov 2020

Signature ..... Date .....

[Name]

[Position]

[Organisation]

Signature ..... Date .....

Report authorised by:

[Name]

[Position]

[Organisation]

Signature ..... Date .....

[Name]

[Position]

[Organisation]

Signature ..... Date .....

CONTENTS

GROWER SUMMARY

Headline

The development of the Glen Moy x Latham raspberry linkage map using the Genotype by Sequencing (GbS) technology (Hackett et al. 2018), due to the higher markers density, allowed the identification of a new 'crumbly' QTL on linkage group 3 (LG3) as well as a more accurate and precise location for those two previously identified on LG1 and LG3. All three QTL proved to be robust across different seasons therefore there is potential to identify molecular markers located within these QTLs. Markers strongly associated with 'crumbly' fruit across a wider gene pool would be valuable for molecular markers assisted breeding and for diagnostic purposes.

Background

Genetic markers represent important progress with application in the crop science particularly in plant breeding (Kebriyaee et al. 2012). Genetic markers are genes or DNA sequences mapped on known chromosome positions and associated with specific genes and/or traits. They are like signs/flags for target genes to which are associated. Genetic markers can be grouped in two main categories, classical (i.e. morphological, cytological and biochemical markers) and molecular markers with the most common being: restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), simple sequence repeats (SSRs) and single-nucleotide polymorphism (SNP) (Nadeem et al. 2018). In this work we focused on SNPs and SSRs markers.

Simple sequence repeats (SSRs) also called microsatellite are nucleotide motifs repeated in tandem. Such motifs can be mononucleotide (A), dinucleotide (GT), trinucleotide (ATT), tetranucleotide (ATCG), pentanucleotide (TAATC) and hexanucleotide (TGTGCA). The sequences flanking the SSRs are conserved and are used to design primers; by means of the polymerase chain reaction (PCR) potential polymorphisms can be easily assessed. SSRs are very abundant in the genome and are co-dominant allowing to distinguish between homozygous and heterozygous alleles (Nadeem et al. 2018).

SNPs are single base-pair changes found into the genome. They can be, depending on the nucleotide substitution, of two different types, transition (i.e. C/T or G/A) or transversions (i.e. C/G, A/T, C/A or T/G). SNPs are very abundant in plant genomes with a frequency ranging from 1 SNP in every 100-300 bp, they can be found in coding or non-coding sequence of

genes and are identified by analysing sequence data stored in database (Nadeem et al. 2018).

Summary

Two different kind of loci, for the identification of potential molecular markers, were considered during this study. Single Nucleotide Polymorphisms (SNPs) and Simple Sequence Repeats (SSRs). Eight markers were selected in total. Five were SNPs and were of the different kinds, gene 'tags', located in non-coding regions of the genome, and target genes located in coding regions while the remaining three markers were SSRs.

The markers are all significantly associated with crumbliness in the Latham x Glen Moy population and in order to determine their wider applicability we required a validation population. The numbers in this population were smaller than ideal however due to being able to identify varieties that never show crumbly symptoms. A population of 63 different raspberry genotypes, with ¾ being selected as being prone to 'crumbliness' while the ¼ as never showing 'crumbly' symptoms was developed. DNA was extracted was from each of the 63 different genotypes and by means of the primers specifically designed to amplify the genome regions containing the selected markers.

Analysis of the allelic polymorphisms was conducted for each marker and potentially significant associations between alleles and 'crumbly' or not 'crumbly' were assessed; the goal was to identify at least one marker significantly linked with one of the two phenotypes through a Genome Wise Association Study (GWAS) using the selected 63 genotypes population as sample. Though not ideal this allows us to examine the allele status of the markers from the LxM population in a wider gene pool.

Financial Benefits

The James Hutton Institute (JHI) is the sole source of *Rubus* plant material for entry into the UK Plant Health Certification Scheme. Provision of disease-free true-to-type nuclear stock of certified varieties represents a statutory requirement and the propagation of high-health planting material is crucial to the establishing of commercial production throughout the UK. The current test of fruiting mother stocks and releasing only those that have passed, may not guarantee control of the condition and thus no foolproof testing method exists.

The identification of 'crumbly' molecular markers could pave the way for the development of faster diagnostic test to replace the current fruiting test; the same molecular knowledge could

be useful to the Raspberry Breeding programme based at James Hutton Limited allowing the selection of new varieties crumbly free or at least more resistant to this condition.

Action Points

Rubus 256e can be considered as a marker for crumbly fruit. Other regions have been identified but need further validation.

SCIENCE SECTION

Introduction

In this work, data is presented on the genetic basis of the ‘crumbly’ phenotype using the GbS map of Hackett et al. (2018) to re-analyse the previous data from Graham et al. (2015). The idea was to identify potentially interesting genes related to ‘crumbly’ fruit whose relevance would have been strengthen by their location inside a genome region already tightly associated to the ‘crumbly’ fruit phenotype such as the ‘crumbly’ QTLs. The aim was to try to find putative gene markers and in fact the new GbS linkage map allowed the identification of new and more significant ‘crumbly’ markers associated with the ‘crumbly’ QTLs. Eight of these markers were selected and subjected to a Genome Wide Association Study (GWAS) to try identifying one or more marker/s strongly associated with ‘crumbly’ fruit across a wider gene pool that could be used for molecular markers assisted breeding and for diagnostic purposes.

Materials and methods

Molecular protocols

Nucleic acid extraction (DNA isolation)

Genomic DNA was isolated from buds of 63 different genotypes (see Table A1 in appendix for the full list) of raspberry (*Rubus idaeus*). Minimum 70 to maximum 90 mg of plant material were weighed and then fine ground with mortar and pestle using liquid nitrogen. The extraction/purification was performed using the DNeasy Plant Mini Kit (Qiagen, Germany), following the manufacturer instructions

Analysis of nucleic acid

Gel electrophoresis

Agarose gels were prepared by mixing 0.75 g of agarose with 50 mL of 1X Tris-borate/EDTA (TBE) buffer. The mixture was heated in a microwave on medium power for 1 minute to allow the agarose gel to dissolve. The mixture was then cooled to ca. 50-60 °C before adding 1 drop of ethidium bromide. The gel was cast in a tank with the required size comb and allowed to set under a fume hood for 1 hour. Once the gel set, the comb was removed and sufficient TBE buffer was added to ensure the gel was fully submerged. Samples were then loaded onto the gel, in the wells formed by the comb, and separated for 50 minutes by electrophoresis at 40 Volts. Imaging of the gel was under UV light, using the UVITech transilluminator (UVITech, Cambridge, UK).

Enzymatic manipulation of nucleic acids

Design of primers

Gene specific primers for PCR were designed using the online software Primer3web version 4.1.0. (primer3.ut.ee). Primer sequences for Sanger sequencing were purchased from Eurofins Genomics (Ebersberg, Germany) while those primers for Fragment analysis (genotyping) were purchased from Sigma-Aldrich (USA) and of these, the forward primers were HEX (hexachloro-fluorescein) labelled.

Polymerase Chain Reaction (PCR)

PCR reactions were carried out in 0.2 mL non-skirted 96-well PCR plates (ThermoScientific, UK) using a final volume of 25 µL containing 0.2 µL (5 U/µL) of Taq DNA Polymerase and 2.5 µL of PCR buffer (F. Hoffmann-La Roche, Switzerland), gene specific forward and reverse primers (0.1 µL each primer) at a final concentration of 0.2 µM, 2.5 µL of dNTPs (Invitrogen™ Corporation, USA) at a final concentration of 0.2 mM and about 50 ng of DNA template (10 µL of DNA stock solution 1:10 dilution). The final volume of 25 µL was reached by adding 9.6 µL of sterile distilled water (SDW). Thermal cycling conditions were as follows: 5 minutes denaturation at 95 °C followed by 35 cycles of 94 °C for 1 minute, 57 °C (melting temperature for the primer pair) for 1 minute at 72 °C for the extension of the expected fragments.

PCR reagents	**volume (µL)**
dNTPs mix[a]	250
reaction buffer	250
Taq polymerase	10
forward primer	20
reverse primer	20
sterile distilled water (sdw)	960

[a]mix, in equal amount, of the four deoxynucleotides (i.e. dATP, dGTP, dCTP and dTTP)

Table 1. Polymerase Chain Reaction (PCR) solution mix for 100 reactions. Volume of the different reagents for the final of 1.5 mL suitable for 100 PCR reactions.

Enzymatic clean-up protocol for Sanger sequencing

5 µL of PCR reaction for each DNA sample were transferred to a new 0.2 mL non-skirted 96-well PCR plates. To each well, containing the amplified DNA, were added 1 µL of Shrimp Alkaline Phosphatase (rSAP) and 1 µL of Exonuclease I both from New England BioLabs Inc. (USA). The PCR plate was then moved to the thermocycler, for the second and last step of

the clean-up protocol, with the following settings, 37 °C for 15 minutes and 80 °C for another 15 minutes; the latter temperature for the complete and irreversible inactivation of the enzymes.

Protocol for preparation of samples for fragment analysis (genotyping)

For 100 reactions, 10 µL of Gene Scan™ 500 ROX™ and 990 µL of highly deionized (Hi-Di) formamide both Applied Biosystems S.A. (USA) were mixed together. Two µL of SDW diluted PCR reaction were mixed with 8 µL of ROX mix reaction buffer and transferred to a new 0.2 mL non-skirted 96-well PCR plates (ThermoScientific, UK) for the fragment analysis (genotyping).

Sanger sequencing of DNA

The fluorescent Sanger sequencing was carried out by the Genome Technology lab at The James Hutton Institute (Dundee, Scotland, UK). DNA was extracted (see section 2.2.2) from all the sixty-three different raspberry cultivars and selections (see Table A.1 appendix). A 0.2 mL non-skirted 96-well PCR plates (ThermoScientific, UK) was prepared containing 0.2 mL of DNA stock solution (1:10 dilution) for each of the sixty-three samples. A Polymerase Chain Reaction (PCR) to amplify the region to be sequenced was performed as followed, 10 µL of DNA stock solution (1:10) were mixed with 15 µL of PCR reaction mix. The mix prepared for 100 reactions contained 250 µL of dNTPs, Deoxynucleoside triphosphates (Invitrogen™ Corporation, USA), 250 µL of buffer mix and 20 µL of Taq DNA polymerase both (10 µM) and both from (F. Hoffmann-La Roche, Switzerland), 10 µL forward primer and 10 µL reverse primer Eurofins Genomics (Ebersberg, Germany) and 960 µL sterile distilled water. The plate with the PCR reactions was then loaded in a thermocycler, Alfa Thermo Cycler (PCR max, UK) with the following reaction settings (see Table 2).

After amplification, the DNA samples were cleaned-up to remove any residual dNTPs that could affect the sequencing reaction; 5 µL of PCR reaction for each DNA sample were transferred to a new 0.2 mL non-skirted 96-well PCR plates and the enzymatic clean-up protocol was performed by adding, 1 µL of Shrimp Alkaline Phosphatase (rSAP) and 1 µL of Exonuclease I both from New England BioLabs Inc (USA) to each sample.

Temperature °C	Time (min)	no. of cycles
95	5	1
94		
57	1	35
72		
72	8	1

Table 2. PCR settings for the amplification of the genomic regions to be sequenced for markers detection.

The PCR plate was then moved in the thermocycler, for the second and last step of the clean-up protocol, with the following settings, 37 °C for 15 minutes and 80 °C for another 15 minutes; the latter temperature to completely and irreversibly inactivation of the enzymes. Once the clean-up reaction was complete, the samples were sent to the Genome Technology lab at The James Hutton Institute for the Sanger sequencing; the ideal amount of DNA for the sequencing for amplified fragments of size between 200 and 500 bp would be 3-5 ng; an estimation of the amount of DNA was done by running few random samples on a 1.5% agarose gel and then comparing the DNA band intensity with those of known amount of DNA used as reference.

For the sequencing reaction, the 5 µL of samples were further processed by adding 1 µL of (10 µM) of forward primer, 1 µL of Big Dye Terminator (version 3.1) reaction mix Applied Biosystems S.A. (USA), 1.5 µL of 5X dilution buffer Applied Biosystems S.A. (USA) and distilled water to bring final reaction volume to 10 µL. The plate containing the samples was then transferred to TETRAD thermal cycler, Applied Biosystems S.A. (USA), using the following programme, hold at 96 °C for 1 minute and then a cycle to be repeated 25 times at 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes. After amplification, the samples were cleaned up by adding for each 10 µL of reaction 2.5 µL of EDTA (125 mM; ph 8.0) and after a brief spinning 30µl of 95% ethanol. The sample were vortexed and spun briefly, incubated at room temperature for 15 minutes and then spun for 30 min at 3000 revolutions per minute (rpm) by keeping the sample at 4 °C. After that, the samples were spun upside down at 100 g for 10 sec, a further cleaning up step was performed by adding 150 µL of 70% ethanol. The sample were briefly vortexed and spun for 10 minutes at 3000 rpm at 4 °C, again this step was followed by spinning the samples upside down at 100 g for 10 seconds; this whole cleaning step was then repeated and after that the samples were left drying at room temperature. Once dried the samples were re-suspended in 10 µL of highly deionized (Hi-Di) formamide, Applied biosystems S.A. (USA) and then analysed on the capillary sequencer ABI3730 DNA analyser Applied biosystem S.A. (USA) with a 36 cm capillary array and Applied Biosystem's POP7 polymer. The samples were run using Applied Biosystems 3730 Data collection software version 4.0 and the sequencing Data coming off the machine were analysed using Applied Biosystem's Sequence Analysis version 6.0; all software Applied Biosystem S.A. (USA).

Fragment analysis for genotyping

The DNA SSR genotyping was carried out by the Genome Technology Lab at The James Hutton Institute (Dundee, Scotland, UK). The DNA samples, used for this analysis, were those

utilised for the sequencing so the same DNA extraction procedure was applied. The PCR amplification, of the genome regions containing the markers, was performed in the same way described above; the only difference was the use of fluorescently labelled forward primers. After amplification, 5 µL of the PCR reactions were run on an electrophoretic gel (1.5 % agarose) to estimate the amount of DNA and then, accordingly, to calculate the dilution rate; usually 280 µL of sterile distilled water (SDW) per PCR reaction were added. To prepare the samples for the analysis, a reaction mix containing 10 µL of Gene Scan™ 500 ROX™ and 990 µL of highly deionized (Hi-Di) formamide both Applied Biosystems S.A. (USA); the 1 mL volume (ROX + Hi-Di) reaction mix refers to 100 reactions. Two µL of SDW diluted PCR reaction were mixed with 8 µL of ROX mix reaction buffer and transferred to a new 0.2 mL non-skirted 96-well PCR plates (ThermoScientific, UK) for the fragment analysis (genotyping). The samples were processed with the Genetic Analyzer 3730 Applied Biosystems S.A. (USA), the data were collected by means of the Genetic Analyzer 3730 data collection software version 4.0 Applied Biosystems S.A. (USA). The final data analysis was performed with GeneMapper software version 5.0 Applied Biosystems™ (USA).

Selection of a population of raspberry genotypes for a Genome Wise Association study (GWAS)

For GWAS, a list of 63 different genotypes (see Table A.4.1 in appendix) was produced to allow us to examine variability in the allele status and the identification of any associations with the phenotype and the potential 'crumbly' markers that could be of economic importance for molecular breeding programs and for diagnostic purposes (i.e. plant health certification). The population of genotypes was carefully selected following consultation with Nikki Jennings (raspberry breeder at The James Hutton Limited), however this selection was not straight forward as some conflicting information was available on the status of the genotypes depending on location and season and this was a noted limitation in the GWAS performed here. Forty-five genotypes, a mix of cultivars and selections, both primocane and floricane fruiting forms, were selected as plants that had been reported as showing 'crumbly' symptoms. The remaining eighteen genotypes, circa 25% of the total, were again a mix, of cultivars and selections both floricane and primocane but these were selected because they were never reported to have displayed 'crumbly' symptoms anywhere in any season and this was the reason for a low number of non 'crumbly' genotypes being included. Where possible, the plant pedigree was recorded because one of the criteria for the analysis to identify potential robust 'crumbly' markers, was to test them on a population of unrelated genotypes. Such approach would permit the selection of markers with a broad applicability.

Statistical analysis

Statistical analyses were performed in GenStat 18th edition (VSN International, UK). The chi-square analysis and the permutation test were performed to identify statistically significative association between the selected 'crumbly' markers and the population of 63 genotypes loosely related to each other with about ¾ of them labelled as showing 'crumbly' symptoms while the remaining ¼ being never 'crumbly'.

Results

Steps towards the selection of 'crumbly' genetic markers for breeding assisted and diagnostic purposes

In total, eight potential 'crumbly' fruit loci were selected for investigation for marker development. Five of these loci were mapped as Single Nucleotide polymorphisms (SNPs), while the other three were Simple Sequence Repeats (SSRs). The eight gene markers selected were of two different kinds; the so called gene 'tags', located in non-coding regions of the genome, and the target genes, on the other hand, located in coding ones (Collard et al. 2005). **MOY_34151, MOY_35728** and **MOY_36258** were the only genes targets all the rest were genes 'tag'. These three genes were located on LG1 (**MOY_34151**) while the other two were both on LG3 one in the QTL identified here (**MOY_35728**) and the other in the original **'**crumbly' QTL (**MOY_36258**).

Two SNPs markers, **s3407_p12510_R23** and **s182_p91185_R6,** the two new most significant 'crumbly' markers for, respectively, the QTL on LG1 and the QTL on LG3, were identified during this work by re-analysing the 'crumbly' score (Graham et al. 2015) with the new GbS linkage map (Hackett et al. 2018).

The last three genes were all SSRs, **Rub2a1** was the most significant marker for the 'crumbly' QTL identified during this work while **ERubLR_SQ05.3_D11AOC** and **Rub256e** were the original 'crumbly' markers, respectively for the QTL on LG3 and the one on LG1; identified by Graham et al. (2015) before the development of the new GbS linkage map (Hackett et al. 2018). These two SSR markers were also considered for the Genome Wide Association Study (GWAS) because they were inside their corresponding 'crumbly' QTL and they were tightly linked to the new most significative markers. Therefore, the principle followed was, the more markers selected and tested the more chances to find at least one 'crumbly' marker being representative of a wide population of individuals showing 'crumbly' symptoms.

The design of the primers for the marker (**s3407_p12510_R23)** failed due to missing sequences in the **scaffold3407** where it is mapped. A new SNP marker (**s353_p21288_R19**), tightly linked with (**s3407_p12510_R23)** was then used.

The primers for all the eight markers were designed as described in material and methods section. They were used for a Genome Wide Association Study (GWAS) whose aim was to identify a 'crumbly' marker either always or never linked to the 'crumbly' population analysed. Ideally the marker/s should be shared by all the genotypes labelled as 'crumbly' (45 genotypes in this case study) while being absent in all those that never displayed 'crumbly' symptoms (18 genotypes in this case study) or vice versa (see Table A.4.1 for full list of genotypes).

The identification of the five SNPs markers, the two gene tags and the three gene targets, was performed by sequencing the amplified regions containing the markers across the genotypes. The procedure accomplished was described in material and methods of this report by using the primers designed as described always in the material and methods section. The results of the first sequencing analysis, performed only with the forward primers, did not identify any SNP marker that was always (or never) found in all the 45 genotypes, having 'crumbly' phenotype, out of the 63 genotypes (see full list of genotypes in appendix Table A.4.1) tested. Fourteen, out of the sixty-three, genotypes selected for the GWAS did not produce any amplification fragments; suggesting that their genome sequence differed in the primer design regions (see 3 for further details) therefore these were not informative in these genotypes.

The amplified region sequenced, with the procedure described in section 4.2.4, for the five different markers varied slightly between the genotypes. The contig, the set of overlapping DNA sequence for the different genotypes varied in length indicated in base pair (bp): **MOY_34151** (235 bp), **MOY_36258** (186 bp), **MOY_35728** (285 bp), **s182_p91185_R6** (487 bp) and **s353_p21288_R19** (355 bp). The number of SNPs per sequence identified varied per locus. The contig for **MOY_34151** had four different SNP positions 139 bp, 146 bp, 158 bp and 183 bp. The contig for **MOY_36258** had five different SNP positions but none segregating in the selected population. The contig for **MOY_35728** carried three different SNP positions (i.e. 75 bp, 174 bp and 242 bp). The contig for the locus **s182_p91185_R6** had eleven different SNP positions (i.e. 110 bp, 129 bp, 139 bp, 147 bp, 148 bp 170 bp, 206 bp, 236 bp, 376 bp, 441 bp and 444 bp). The contig of the last marker, **s353_p21288_R19**, carried only three different SNP positions (i.e. 180 bp, 196 bp and 303 bp).

Pedigree				
♀ **mother**	**x**	♂ **father**	**genotype**	**phenotype**
complex hybrid	**x**	complex hybrid	Autumn Bliss	'crumbly'
Joan Squire	**x**	complex	Brice	'crumbly'
[1]Autumn bliss			Erika	'crumbly'

SCRI 6531/84	**x**	SCRI 6549/1	Glen Prosen	‘crumbly’
			Kweli	‘crumbly’
Nootka	**x**	Glen Prosen	Tulameen	‘crumbly’
			Obbard	‘crumbly’
Willamette	**x**	Cuthbert	Meeker	‘crumbly’
97134B1	**x**	8510A57	0867E-4	‘crumbly’
7326E1	**x**	7412H16	Glen Rosa	‘crumbly’
			Imara	‘crumbly’
Glen Rosa	**x**	SCRI 8605C-2	Glen Doll	no ‘crumbly’
Preussen	**x**	Lloyd George	Malling Minerva	no ‘crumbly’
			Malling Leo	no ‘crumbly’

[1]open pollinated

Table 3. Selected genotypes for the GWAS that did not give any amplification fragments for all the five SNPs markers, polymorphic between Glen Moy and Latham.

List of fourteen genotypes, out of the 63 selected for the GWAS and not containing the five markers strongly associated with the three ‘crumbly’ QTLs and then chosen for this study. For each genotype, when available, was reported the pedigree too.

Chi-square (χ^2) tests of independence were used to look for associations between the two ‘crumbly’ categories (i.e. ‘crumbly’ and non ‘crumbly’) and markers. Each marker SNP position was tested to see whether any of the various polymorphic forms detected in the GWAS population was significantly associated with either of the two traits (i.e. ‘crumbly’ and non ‘crumbly’). Ideally a $p = 0.05$ confidence level would be considered significant but due to the low numbers in the GWAS population higher levels would be considered significant (eg. $p=0.01$). If this was found, then that SNP position could have been considered linked with the genotypes showing ‘crumbly’ phenotype and that marker could be then analysed by a generalised mixed mode (Yu et al. 2006) specifically designed to address quantitative traits, such as ‘crumbly’ fruit and complex levels of relatedness within the analysed population.

None of the segregating SNP position, for the five markers, was highly significatively associated with the ‘crumbly’ phenotype in the studied population (see Tables 4 and 5 for further details). The use only of the forward primers to sequence the amplified fragment containing the selected markers limited the size of the region in which screen for segregating alleles. The use of both forward and reverse primers and the chromosome walking, along the region were the markers are located, could increase the size of the fragment to be sequenced and thus the number of SNP positions to enhance the chances of identification of segregating SNPs, potentially associated with the ‘crumbly’ phenotype, in the selected population.

The presence of expected values lower than 5 in any cell of the contingency table would make the probability test calculation unreliable (Steve, 2011). To bypass this issue, a permutation

test to calculate the significance probability for a chi-square test of the independence of rows and columns was performed. This procedure converts the usual chi-square test in a nonparametric alternative that, in situations where some values in the contingency table are lower than five, allows to strengthen the results of the analysis (GenStat manual, VSN International, UK). The permutation test simulates the random distribution of table values that may occur in tables that have the same overall distribution of numbers over the columns, and over the rows, as in the original table. Even by means of a more accurate and precise statistical analysis, none of the markers tested proved to be significantly associated with the 'crumbly' phenotype.

For the SSRs markers, while for **Rub2a1** and **ERubLR_SQ05.3_D11AOC** none of the segregating alleles showed significant association with the 'crumbly' phenotype with, respectively, p = 0.28 and p = 0.24 for the permutated chi-square test. The SSR marker Rub256e however showed significant association p = 0.02 for the permeated chi-square test.

marker	allele	*χ^2	†d.f.	p-value
MOY_34151	base_146	1.99	3	0.574
	base_139	2.62	4	0.662
	base_183	0.88	2	0.643
	base_158	2.56	2	0.217
[a]MOY_35	base_75	2.91	4	0.573
	base_174	2.52	2	0.284
	base_242	0.68	1	0.41
s182_p9185_R6	base_110	1.96	3	0.58
	base_129	2.93	2	0.231
	base_139	6.39	4	0.172
	base_147	1.52	2	0.467
	base_148	2.18	3	0.536
	base_170	3.19	2	0.203
	base_206	4.11	3	0.250
	base_236	2.01	2	0.367
	base_376	1.14	2	0.565
	base_441	2.18	3	0.536
	base_444	2.71	2	0.257
[b]s353_p21	base_196	3.69	3	0.297
	base_180	0.94	2	0.625

	base_303	2.9	2	0.235

*Pearson chi-square value - [t]degree of freedom
[a]**MOY_35728 - [b]s353_p21288_R16**

Table 4. Significance probability for a chi-square test.

For each marker was reported the number of alleles, the chi-square test statistic, the probability test (p-value) and the number of degrees of freedom for a chi-square test with two criteria (i.e. 'crumbly' and 'non crumbly') was calculated according the formula (m-1) x (n-1) where m is the number or rows and n the number of column of the contingency table.

marker	allele	$^{*}\chi^2$	[t]d.f.	p-value
MOY_34151	base_146	1.99	3	0.668
	base_139	2.62	4	0.736
	base_183	0.88	2	0.723
	base_158	2.56	2	0.343
[a]MOY_35	base_75	2.91	4	0.684
	base_174	2.52	2	0.379
	base_242	0.68	1	0.479
s182_p9185_R6	base_110	1.96	3	0.727
	base_129	2.93	2	0.256
	base_139	6.39	4	0.167
	base_147	1.52	2	0.612
	base_148	2.18	3	0.632
	base_170	3.19	2	0.240
	base_206	4.11	3	0.234
	base_236	2.01	2	0.367
	base_376	1.14	2	0.569
	base_441	2.18	3	0.609
	base_444	2.71	2	0.298
[b]s353_p21	base_196	3.69	3	0.346
	base_180	0.94	2	1
	base_303	2.9	2	0.242

*Pearson chi-square value - [t]degree of freedom
[a]**MOY_35728 - [b]s353_p21288_R16**

Table 5. Significance probability for a for a permutated chi-square test.

For each marker was reported the number of alleles, the chi-square test statistic, the probability test (p-value) and the number of degrees of freedom for a chi-square test with two criteria (i.e. ‘crumbly’ and ‘non crumbly’) was calculated according the formula (m-1) x (n-1) where m is the number or rows and n the number of column of the contingency table.

Discussion

The first ‘crumbly’ markers tested for association with specific genotypes from a population of 63 individuals did not show highly significant associations with the ‘crumbly’ fruit phenotype though some of these are promising and should be examined in future on a larger population and also considering other SNPs close by as due to time constraints and Covid-19 the first steps towards marker identification/validation were focused on the sequencing of amplified regions produced with only the forward primers. This produced a limited number of SNPs for analysis and in future the use of both primers (i.e. forward and reverse) and extending sequencing across a wider region might help with the identification of further alleles that might be significantly associated with ‘crumbly’ fruit.

The selection of markers associated to complex quantitative traits such as ‘crumbly’ fruit is not an easy task and the size and structure of the population taken under exam play an important role in the whole process.

Three ‘crumbly’ SSR markers were also selected and their primers were designed. In this case **Rub256e** that showed significant association with the ‘crumbly phenotype.

References

Collard B.C.Y., Jahufer M.Z.Z., Brouwer, J.B. & Pang, E.C.K. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica,* 142**,** 169-196.

Graham, J., Smith, K., McCallum, S., Hedley, P. E., Cullen, D. W., Dolan, A., Milne, L., McNicol, J. W. and Hackett, C. A. (2015). Towards an understanding of the control of 'crumbly' fruit in red raspberry. Springerplus, 4:223.

Hackett, C. A., Milne, L., Smith, K., Hedley, P., Morris, J., Simpson, C. G., Preedy, K. and Graham, J. (2018). Enhancement of Glen Moy x Latham raspberry linkage map using GbS to further understand control of developmental processes leading to fruit ripening. Bmc Genetics, 19:59.

Kebriyaee, D., Kordrostami, M., Rezadoost, M. H. and Lahiji, H. S. (2012). QTL Analysis of Agronomic Traits in Rice Using SSR and AFLP Markers. Not Sci Biol, 4:116-123. https://doi.org/10.15835/nsb427501

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, S. Y., Comertpay, G. et al. 2018. DNA molecular markers in plant breeding: current status and recent advancements in genomiccselection and genome editing. Biotechnology & Biotechnological Equipment, 32:261-285, DOI:10.1080/13102818.2017.1400401

Scolari, L. M. (2020). Understanding the genetic and physiology controls of ‘crumbly’ fruit in red raspberry (*Rubus idaeus*). AHDB PhD project final report. https://ahdb.org.uk/horticulture

Steve, L. 2011. How to use statistics, Gosport.

Yu, J., Pressoir, G., Briggs, W. H., Vroh BI, I., Yamasaki, M. et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet, 38, 203-8.

Appendices

Graham J, Hackett CA, Smith K, Woodhead M, Hein I, McCallum S (2009) Mapping QTLs for developmental traits in raspberry from bud break to ripe fruit. Theor Appl Genet 118 (6):1143-1155

Table A1. List of 63 different genotypes of which 45 specifically selected for their tendency to show 'crumbly' symptoms and 18 on the contrary for having never displayed the symptoms.

Pedigree					
♀ mother	x	♂ father	genotype		phenotype
DJ1185	x	8510A21	0560E11	FC	crumbly
Tulameen	x	8510A28	0565F3	FC	crumbly
Malling Minerva	x	8510A41	0663RE3	FC	crumbly
R4A1	x	Glen Fyne	0946/4	FC	crumbly
7826C1	x	8627RE7	9059D-2	FC	crumbly
9421A4	x	9434B-1	99105RC-2	FC	crumbly
9426C-5	x	9429E-2	99116E-4	FC	crumbly
SCRI 8631D-1	x	SCRI 8605C-2	Glen Fyne	FC	crumbly
9059D-2	x	8510A73	0019E2	FC	crumbly
9455F-2	x	8510A5	0304F6	FC	crumbly
0096RF-4	x	8510A6	04101A5	FC	crumbly
0003RB1	x	8510A9	0433F2	FC	crumbly
9046RA2	x	8510A22	0511F1	FC	crumbly
Glen Fyne	x	8510A29	0573B5	FC	crumbly
0312E3	x	8510A35	0658C5	FC	crumbly
0312E3	x	8510A32	0658E-1	FC	crumbly
0312E3	x	8510A37	0658F-7	FC	crumbly
Glen Ample	x	8510A43	0671D-4	FC	crumbly
97134B1	x	8510A57	0867E-4	FC	crumbly
R4A1	x	Glen Fyne	0946/19	FC	crumbly
8735J-7	x	8626RJ-2	9050RD3	FC	crumbly
8820E3	x	88K-7	9238D5	FC	crumbly
EM5961/1	x	7826C1	9350F3	FC	crumbly
WSU1068	x	ORUS 2078	97134B1	FC	crumbly
9349F5	x	9349A4	9764F-3	FC	crumbly
9351D-3	x	9350E1	9769RD1	FC	crumbly
	x		Autumn Bliss	PF	crumbly
Autumn Bliss[1]	x		Erika	PF	crumbly
7326E1	x	7412H6	Glen Ample	FC	crumbly
00123A5	x	0019B11	Glen Dee	FC	crumbly
SCRI 7331/1	x	SCRI 7256/1	Glen Lyon	FC	crumbly
SCRI 688/12	x	SCRI 6815/113	Glen Moy	FC	crumbly

SCRI 6531/84	x	SCRI 6549/1	Glen Prosen	FC	crumbly
7326E1	x	7412H16	Glen Rosa	FC	crumbly
7741D4	x	7919B11	Glen Shee	FC	crumbly
	x		Imara	PF	crumbly
Autumn Bliss	x		Joan Squire	PF	crumbly
	x		Kweli	PF	crumbly
King	x	Louden	Latham	FC	crumbly
Willamette	x	Cuthbert	Meeker	FC	crumbly
(unknown)	x		Obbard	PF	crumbly
Nootka	x	Glen Prosen	Tulameen	FC	crumbly
9431G-8	x	8510A71	Sanibelle	FC	no crumbly
0015D3	x	8510A14	0453C4	FC	no crumbly
	x		00123A7	FC	no crumbly
Glen Rosa	x	8510A20	0550E4	FC	no crumbly
0304F6	x	Autumn Treasure	0925B4	PF	no crumbly
0304F6	x	Autumn Treasure	0925B8	PF	no crumbly
0304F6	x	Autumn Treasure	0925D-15	PF	no crumbly
8003G10	x	8003C1	8605C-2	FC	no crumbly
8020E8	x	8631D-1	9025A1	FC	no crumbly
complex	x		Autumn Britten	PF	no crumbly
EM6304/36	x	EM6330/96	Autumn Treasure	PF	no crumbly
SCRI 9422C-4	x	SCRI 9434B-1	Glen Cally	FC	no crumbly
0030E-12	x	8510A16	Glen Carron	FC	no crumbly
Glen Rosa	x	SCRI 8605C-2	Glen Doll	FC	no crumbly
SCRI 9422C-4	x	SCRI 9434B-1	Glen Ericht	FC	no crumbly
	x		Malling Leo	FC	no crumbly
Joan Squire	x	complex	Brice	PF	crumbly
	x		Chief	FC	no crumbly
Joan J	x	complex	Joan Irene	PF	crumbly
EM3689	x	Gaia	Malling Hiesta	FC	crumbly
EM selection	x	SCRI selection	Malling Minerva	FC	no crumbly
Lloyd George	x	Preussen	Schoenemann	FC	crumbly

[1]open pollinated - FC (floricane) – PF (primocaine)

