University
of Dundee

University of Dundee

Quantitative trait loci mapping of polyphenol metabolites from a 'Latham' x 'Glen Moy' red raspberry (Rubus idaeus L) cross

McDougall, G. J.; Allwood, J. W.; Dobson, G.; Austin, C.; Verrall, S.; Alexander, C. J.

[Link to publication in Discovery Research Portal](#)

ORIGINAL ARTICLE

Quantitative trait loci mapping of polyphenol metabolites from a 'Latham' x 'Glen Moy' red raspberry (*Rubus idaeus* L) cross

G. J. McDougall[1] · J. W. Allwood[1] · G. Dobson[1] · C. Austin[1] · S. Verrall[2] · C. J. Alexander[3] ·
R. D. Hancock[4] · J. Graham[4] · C. A. Hackett[3]

Abstract

Objective The objective of this study was to investigate the genetic control of polyphenol accumulation in red raspberry (*Rubus idaeus* L).

Methods The levels of total anthocyanins and 37 individual polyphenol metabolites were measured over three years in a raspberry biparental mapping population. Quantitative trait loci (QTLs) for these traits were mapped onto a high-density SNP linkage map.

Results At least one QTL was detected for each trait, with good consistency among the years. On four linkage groups (LG), there were major QTLs affecting several metabolites. On LG1, a QTL had large effects on anthocyanins and flavonols containing a rutinoside or rhamnose group. On LG4, a QTL had large effects on several flavonols and on LG5 and LG6 QTLs had large effects on ellagic acid derivatives. Smaller QTLs were found on LG2 and LG3.

Conclusion The identification of robust QTLs for key polyphenols in raspberry provides great potential for marker-assisted breeding for improved levels of potentially health beneficial components.

Keywords Raspberry · Polyphenols · Anthocyanins · Ellagitannins · Genetic control · Heritability · QTLs

1 Introduction

World raspberry production has risen steadily from 520 000 tonnes in 2010 to 896 000 tonnes in 2020 [FAO figures] with 75% of production in Europe. The value of raspberry production quadrupled in the UK from 1995 to 2015 supported by a drive for increased class I cultivation for the fresh market [DEFRA figures] although class II fruits also supply the increasing popularity of raspberry juice products (Rajauria & Tiwari, 2018) as well as traditional uses in purees, jams, and preserves.

Raspberries are valued for their distinctive flavour and colour and raspberry fruit breeding has widened its focus from traits associated with agronomic performance and disease resistance towards those associated with fruit sensory quality (Jennings, 2018; Jennings et al., 2016) and potential health benefits (e.g., Mazzoni et al., 2016). The potential health benefits of raspberries have been associated with their polyphenol content and composition (e.g., Hancock et al., 2018). The distinctive coloration of red raspberries is due to their accumulation of anthocyanins during ripening, but these are accompanied by the presence of ellagitannins, ellagic acid derivatives, and flavonols (e.g., Carvalho et al., 2013; McDougall et al., 2005, 2014; Mullen et al., 2003). The anthocyanins and ellagitannins are the major polyphenol components and have received the most attention for potential health benefits (e.g. Rao & Snyder, 2010). The ellagitannins, in particular, contribute to astringency, which is an important factor in the characteristic sensory properties of this fruit (e.g. He et al., 2015).

✉ G. J. McDougall
gordon.mcdougall@hutton.ac.uk

1 Plant Biochemistry and Food Quality Group, Environmental and Biochemical Sciences Department, The James Hutton Institute, Dundee DD2 5DA, Scotland, UK

2 Ecological Sciences Department, The James Hutton Institute, Dundee DD2 5DA, Scotland, UK

3 Biomathematics and Statistics Scotland (BioSS), The James Hutton Institute, Dundee DD2 5DA, Scotland, UK

4 Cell and Molecular Sciences Department, The James Hutton Institute, Dundee DD2 5DA, Scotland, UK

The levels of anthocyanins, ellagitannins and other phenolics vary between different Rubus species and indeed different raspberry varieties (e.g. Bradish et al., 2012; Deighton et al., 2000). Although phenolic content and composition can vary with agronomic practice, location of growth, abiotic and biotic stresses, there is a strong genetic element in the control of accumulation (Anttonen & Karjalainen, 2005; Connor et al. 2005; Remberg et al., 2010; Dobson et al., 2012; Bradish et al., 2012).

Previous studies have used the genetic linkage map developed from a cross between red raspberry cultivars, Glen Moy and Latham (Graham et al., 2004), followed by Quantitative Trait Loci (QTL) mapping and have identified multiple markers for a broad range of important raspberry agronomical characteristics, which can lead to improved cultivars [e.g. McCallum et al., 2018; Scolari et al., 2021]. Also using this approach, studies at the James Hutton Institute established QTLs for total anthocyanin content and for the levels of selected major individual raspberry anthocyanins (Kassim et al., 2009; McCallum et al., 2010; Paterson et al., 2013). In other studies, consistent QTLs were also established for total phenol content and total anthocyanin content in year-on-year studies (Dobson et al, 2012). As total anthocyanin content is a sub-set of total phenol content, QTLs for total phenol which did not align with those for total anthocyanin content could be related to variation in the other polyphenol components.

The genetic linkage map used in these studies based on the Glen Moy-Latham cross has subsequently been enhanced using GbS (Hackett et al., 2018), providing a higher-density GbS map that allows QTLs to be identified more precisely. The establishment of a draft genome sequence for Glen Moy aided the development of the GbS map and also allows direct assignment of SNP marker information to the Glen Moy genome scaffolds. These resources provide the ability to perform more precise QTL analysis and genes associated with QTLs can also be examined against gene expression data across fruit development to correlate expression against the timing of polyphenol accumulation. This study examines the variation in individual polyphenol components in the progeny set of this cross identified by liquid chromatography mass spectrometry (LC–MS^n) across three years of cultivation. We combine this data with marker data to identify robust and consistent QTLs associated with specific polyphenol classes. The data presented offers opportunities for the development of markers to accelerate the development of raspberry varieties with specific increases in key polyphenolic antioxidants.

2 Materials and methods

2.1 Mapping population and map

The mapping population was the previously reported cross between the North American raspberry Latham and the Scottish raspberry Glen Moy, with 188 offspring (Graham et al. 2004). This cross has been studied in detail over many years at the James Hutton Institute, for traits such as fruit ripening and fruit characteristics including anthocyanin content (e.g. Graham et al., 2009; Kassim et al., 2009; McCallum et al., 2010). These traits have been mapped on a map of medium density, with up to 439 markers for the study of crumbly fruit by Graham et al. (2015). More recently the linkage map was extended by 2348 SNPs from genotyping-by-sequencing to give a high-density linkage map and the fruit ripening data was reanalysed using this map (Hackett et al., 2018). The high-density map has been used for the QTL interval mapping analysis below.

2.2 Metabolite analysis using liquid chromatography mass spectrometry (LC-MS^n)

Fruit from the parents and offspring of the mapping population were analysed for metabolite profiles over 3 years, 2010, 2011 and 2015. In 2010, two field replicates of the population were analysed and in 2011 and 2015, three field replicates were used for each year. The analysis was carried out using LC–MS^n, with one batch for each field replicate in 2010 and 2011. In 2015, extra quality control samples were included, and the samples were analysed in five batches. The weather conditions in the analysis years, 2010, 2011 and 2015 were substantially different in maximum temperature, distribution of rainfall and sun hours across the growing season (results not shown).

As before (Dobson et al., 2012), fruit was picked when ripe, placed in labelled bags stored on ice, transferred to the laboratory within one hour and frozen at -20 °C until extraction. Whilst still frozen, a representative subsample of fruit from each progeny (5–8 berries equivalent to 7–10 g) was weighed. The selected berries were cut in half, then extracted using a glass tissue homogenizer with a PTFE pestle with an equal volume to weight of ice-cold 100% acetonitrile (ACN) containing 0.1% formic acid (FA). This method does not crush the seeds, so it is possibly not representative of some components (e.g., ellagitannins) that are known to be enriched in the seeds (e.g., Arnold et al., 2022). The homogenate was transferred in a centrifuge tube and centrifuged at 2 750 g for 10 min at 5 °C (Eppendorf 5810 R) and the supernatant aliquoted into 1 mL amounts. The extracts were analysed for total anthocyanin content as described by McDougall et al. (2005) using a standard curve of cyanidin-3-*O*-glucose (Extrasynthese, Genay, France). The aliquots were dried in a Speed-Vac then stored at -20 °C.

The dried raspberry extracts were resuspended in 475 µL of 10% ACN containing 0.1% FA with vortex mixing then 25 µL of internal standard solution (0.5 mg/mL morin in methanol; Sigma Chem Co. Ltd) was added. After centrifugation at $10\,000 \times g$ for 10 min at 5 °C, the supernatants were removed and placed in 0.45 µM PTFE filter vials (Thomson Instrument Company, Bioprocess Engineering Services Ltd, Kent, UK) prior to analysis. Samples were analysed using a LC system consisting of an Accela 600 quaternary pump and Acela PDA detector coupled to an LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific Ltd.). Samples (10 µL) were injected onto a 2×150 mm (4 µm) Synergy Hydro-RP 80 fitted with a C18 4×2 mm Security Guard cartridge (Phenomenex Ltd, Macclesfield, UK.). Auto-sampler and column temperatures were maintained at 6 and 30 °C, respectively. The samples were analysed at a flow rate of 200 µL/min using a binary mobile phase of (A) 0.1% aqueous formic acid and (B) 0.1% formic acid in 50% acetonitrile/water with the following gradient: 0−5 min, 5% B; 5−22 min, 5−50% B; 22−32 min, 50−100% and 32–34 min 100% B. Mass detection was carried out using an LTQ Orbitrap mass spectrometer in positive ESI mode. Two scan events were employed; full-scan analysis was followed by data-dependent MS/MS of the three most intense ions using collision energies of 45 eV source voltage (set at 3.4 kV) in wide-band activation mode. The instrument was optimized by tuning against morin at a resolution of 100 000 in a range of 80−2000 mass units. For optimal electrospray ionization, the source conditions were set at a source temperature of 280 °C, sheath gas at 60 arbitrary units, and an auxiliary gas at 5 arbitrary units. Prior to analysis, the mass accuracy of the instrument was assured by calibration following the manufacturer's protocols. All predicted formula data presented were accurate at < 2 ppm.

The samples were analysed in a randomised order and the quality of the MS response was checked by monitoring blanks containing the internal standard and quality control samples of samples from the parental lines interspersed through the sequence of samples. After peak checking, raw peak areas for the major components were obtained using the resident Xcalibur software. After export of MS peak data to Microsoft Excel, the peak areas were ratioed against the internal standard (morin).

Metabolites were identified by their MS properties, the formulae derived from the exact mass data and fragmentation MS[2] data and compared against literature on raspberry components as discussed. Only cyanidin-3-*O*-glucoside, pelargonidin-3-*O*-glucoside, cyanidin-3-*O*-sophoroside, quercetin-3-*O*-glucoside and ellagic acid were confirmed against standard compounds (Extrasynthese Ltd).

2.3 Statistical analysis

2.3.1 Exploratory analysis

Each metabolite was analysed using a linear mixed model, fitted using REML (GenStat 20, VSN International, 2019). The metabolites were transformed using a $\log_{10}$ transformation before analysis to stabilise the variance. To estimate genotype means for QTL mapping, data from each year was first analysed separately and genotype was treated as a fixed effect. For 2015, a random effect of batch in addition to that of field replicate was tested but was not significant. A mean over the years was also estimated from a mixed model with genotype, year and their interaction as fixed effects and field replicate within year as a random effect. The generalised heritabilities of the genotypes and the genotype×year interactions were estimated using the GenStat VHERITABILITY procedure. For this analysis, genotype, year, and field replicate within year were treated as random effects. Pearson correlations between the metabolites were calculated for each year separately, based on the log-transformed genotype means.

2.3.2 QTL analysis

For each year separately, and for the mean across years, QTL interval mapping was carried out for each metabolite using the high-density SNP linkage map and the mapping approach from Hackett et al. (2018), based on a hidden Markov model (HMM) to estimate genotype probabilities. This method was shown in that paper to give better peak resolution than software such as MapQTL and GenStat for this population, where there is much more information about markers from the Latham parent than from Glen Moy. The analysis combines information across genetic markers along the chromosome to estimate the probabilities of each possible QTL genotype for each offspring at each position. In a cross such as this with outbreeding parents, the parental genotypes at a QTL are usually represented as ab × cd, with offspring genotypes ac, ad, bc, and bd. Genetic predictors for the Latham additive effect (P_1), the Glen Moy additive effect (P_2) and the dominance effect (D) can then be derived from the genotype probabilities *pr(ac)* etc. as:

$$P_1 = pr(bc) + pr(bd) - pr(ac) - pr(ad)$$

$$P_2 = pr(bd) + pr(ad) - pr(bc) - pr(ac)$$

$$D = pr(bd) - pr(bc) - pr(ad) + pr(ac)$$

A permutation study was carried out on this population for highly multivariate imaging data (Williams et al., 2021) and estimated the genome-wide 95% LOD threshold as 3.86; the same threshold has been used for all traits here. A

thinned SNP map was used to carry out a QTL × year analysis in GenStat, with a stepwise selection of multiple QTLs, as the high-density map was too computationally intensive for this analysis. This used the significance threshold of Li and Ji (2005), with a genome-wide significance of 0.05.

3 Results and discussion

3.1 LC–MS^n analysis of polyphenol metabolites

Thirty-seven metabolites were detected in all 3 years (Table 1). These comprised 12 anthocyanins, ten of which could be putatively identified from previous work on red raspberries (Bradish et al., 2012; McDougall et al., 2005; Mullen et al., 2003) i.e. cyanidin sophoroside, cyanidin glucoside, cyanidin sophoroside rhamnose /cyanidin glucosyl rutinoside, cyanidin sambubioside, cyanidin rutinoside/pelargonidin sophoroside, cyanidin sambubioside rhamnoside/cyanidin xylosyl rutinoside, pelargonidin glucoside, pelargonidin glucosyl rutinoside, pelargonidin sambubioside and pelargonidin rutinoside). Three of these anthocyanin components could not be distinguished by their molecular formulae or their MS^2 properties so are labelled with both possible structures (e.g. CyRut/PelSoph). Two other anthocyanin-like components ($CyC_{28}H_{30}O_{17}$ and $CyC_{34}H_{40}O_{21}$) were identified as possible cyanidin derivatives due to their fragmentation to *m/z* 287 [which is consistent with cyanidin (Cy)] and were named with their putative molecular formula. Four major ellagitannins were noted (Lambertianin C, Sanguiin H2, Sanguiin H6 and Sanguiin H10) and five ellagic acid derivatives (ellagic acid, ellagic acid acetyl pentose, ellagic acid pentose 1, ellagic acid pentose 2, and methyl ellagic acid pentose) (Gasperotti et al., 2010; McDougall et al., 2014). As the homogenisation method did not crush the seeds, the metabolic profile probably more reflects the flesh of the raspberries and may not capture the entirety of ellagitannin constituents, which are known to be enriched in the seeds (Arnold et al., 2022). However, the extractable components should be consistent across the progeny. Sixteen flavonol derivatives were noted (i.e. kaempferol glucuronide, galactoside and glucoside but also quercetin dihexosides, glucuronide, galactoside, glucoside, hexosyl rhamnosides, glucosyl rutinosides, pentose and malonyl glucoside). The putative identification of these components was also aided by SOPs delivered through the EUBERRY project (Dobson et al., 2010).

Photodiode array (PDA) profiles (280 nm; Fig. S1A) showed the major components in the parents. However, profiles at 520 nm (Fig. S1B) highlighted the anthocyanins with Latham generally showing higher levels of some of the later eluting anthocyanins (A5–A12).

3.2 Exploratory analysis of polyphenol metabolite data

Summary statistics for the metabolite data and for the total anthocyanins are presented in Table 2. No samples from the Glen Moy parent were analysed in 2011. The mixed model analysis showed few statistically significant (p < 0.05) differences between the levels in the parents. For example, CySam and QPent were significantly higher in Glen Moy, whilst CyRut/PelSoph, PelGlu, PelRut, PelGluRut, KGlcU, QGlcU, QGluRut1, QGluRut2 and QDihex1 had significantly higher levels in Latham.

Histograms of the distributions of the log-transformed metabolites showed several metabolites had bimodal distributions, as recorded in Table 2. This was usually consistent across all three years, although for some there was no clear bimodality in 2015. As discussed below, these distributions are associated with major QTLs detected on linkage groups (LG) 1, 4 and 6. The generalised heritability for the genotypes varied from 0.31 to 0.89 (Table 2). The highest heritabilities were found mainly for the metabolites with bimodal distributions: the anthocyanins PelRut, PelGluRut, CySophRham/CyGluRut, CyXylRut/CySamRham, $CyC_{34}H_{40}O_{21}$, the flavonols KGal, QGal, QPent, QHexPent, QDihex1, QDihex2, QGluRut1, QGluRut2, QHexRham1, QHexRham2 and the ellagic acid derivative EllagicPent2. The lowest generalised heritabilities were found for KGlu, KGlcU, QGlcU and Ellagic, with values of 0.31–0.45. The generalised heritability of the genotype x year effect was much lower, varying from 0.00 to 0.28 apart from for KGlu, estimated as 0.47.

Because of the bimodal distributions for several metabolites, a simple correlation coefficient does not represent the relationship among the metabolites well in some cases. Supplementary Figures S2A, S2B and S2C show scatter plot matrices for the anthocyanins, flavonols and ellagitannins/ellagic acid derivatives respectively, based on the log-transformed genotype means for 2010 and Figure S2D shows relationships between the flavonols and anthocyanins. The figures for the other years (not shown) are similar. As an example, consider the correlation between CySoph and CyXylRut/CySamRham in Fig. S2A. The Pearson correlation coefficient was − 0.50 but the scatter plot showed that some genotypes had very low levels of CyXylRut/CySamRham, and the other genotypes formed two clusters, within each of which the correlation with CySoph was positive. Another interesting pair were CySoph and CySam. The individual distributions for these metabolites were unimodal but the scatter plot showed two clear clusters. Likewise, MeEllPent and EllAcetylPent in Fig. S2C showed clear clustering and a positive correlation within the clusters, but unimodal

Table 1 Components identified in raspberry progeny

Peak	RT	PDA max	m/z [M+H]$^+$	MS^2	Predicted formula [M+H]$^+$	Putative identity	Short name
A1	4.23	515	611.1603	287	$C_{27}H_{31}O_{16}$	Cyanidin3-O-sophoroside[a]	CySoph
A2	4.52	515	449.1075	287	$C_{21}H_{21}O_{11}$	Cyanidin-3-O-glucoside[a]	CyGlu
A3	4.62	515	757.2184	287, **611**	$C_{33}H_{41}O_{20}$	Cyanidin sophoroside rhamnoside/ Cyanidin glucosyl rutinoside	CySophRham/ CyGluRut
A4	4.69	515	581.1503	287	$C_{26}H_{29}O_{15}$	Cyanidin sambubioside	CySam
A5	4.94	505	639.1566	287	$C_{28}H_{31}O_{17}$	Cyanidin derivative	$CyC_{28}H_{30}O_{17}$
A6	4.97	510	595.1660	271, **287**, 449	$C_{27}H_{31}O_{15}$	Cyanidin rutinoside/ Pelargonidin sophoroside	CyRut/PelSoph
A7	5.05	510	727.2079	**287**, 433, 581	$C_{32}H_{39}O_{19}$	Cyanidin sambubioside rhamnoside/ Cyanidin xyloside rutinoside	CyXylRut/ CySamRham
A8	5.22	510	433.1127	271	$C_{21}H_{21}O_{10}$	Pelargonidin-3-O-glucoside[a]	PelGlu
A9	5.26	510	785.2134	**287**, 639	$C_{34}H_{41}O_{21}$	Cyanidin derivative	$CyC_{34}H_{40}O_{21}$
A10	5.44	510	741.2240	**271**, 417, 595	$C_{33}H_{41}O_{19}$	Pelargonidin sophoroside rhamnoside/ Pelargonidin glucosyl rutinoside	PelGluRut/ PelSophRham
A11	5.76	500	565.1540	ND	$C_{26}H_{29}O_{14}$	Pelargonidin sambubioside	PelSam
A12	5.73	510	579.1711	**271**, 433	$C_{27}H_{31}O_{14}$	Pelargonidin rutinoside	PelRut

Peak	RT	PDA max	m/z [M−H]$^-$	MS^2	Predicted formula [M−H]$^-$	Putative identity	Short name
13	5.98	270	783.0574*	**1265,** 935, 933, 897, 633, **301**	$C_{68}H_{47}O_{44}$	Sanguiin H10	Sanguiin H10
14	6.12	355	771.1839	**301**, 591	$C_{33}H_{39}O_{21}$	Quercetin glucosyl rutinoside	QGluRut1
15	6.34	350	771.1887	301, **591**	$C_{33}H_{39}O_{21}$	Quercetin glucosyl rutinoside	QGluRut2
16	6.44	350	625.1330	**301**, 445, 505	$C_{27}H_{29}O_{17}$	Quercetin dihexoside	QDihex1
17	6.60	350	625.1331	301	$C_{27}H_{29}O_{17}$	Quercetin dihexoside	QDihex2
18	7.14	355	433.0353	301	$C_{19}H_{13}O_{12}$	Ellagic acid pentoside	EllagicPent1
19	7.43	355	433.0347	301	$C_{19}H_{13}O_{12}$	Ellagic acid pentoside	EllagicPent2
20	7.22	260	1401.5901*	**633**, 933,1099, 1250, 1567, 1867	$C_{123}H_{79}O_{78}$	Lambertianin C	Lambertianin
21	7.31	365	300.9947	185, 229, **257**, 284	$C_{14}H_5O_8$	Ellagic acid[a]	Ellagic
22	7.47	260	551.0307*	529	$C_{48}H_{31}O_{31}$	Sanguiin H2	Sanguiin H2
23	7.55	260	1869.3168 (934.059*)	301, 633, 1235, **1567**	$C_{82}H_{53}O_{52}$	Sanguiin H6	Sanguiin H6
24	7.77	350	595.1219	**301**, 505	$C_{26}H_{27}O_{16}$	Quercetin hexosylpentoside	QHexPent
25	7.81	355	609.1374	**301**, 463	$C_{27}H_{29}O_{16}$	Quercetin hexosylrhamnoside	QHexRham1
26	7.98	355	463.0823	301	$C_{21}H_{19}O_{12}$	Quercetin galactoside	QGal
27	8.15	355	477.0608	301	$C_{21}H_{17}O_{13}$	Quercetin glucuronide	QGlcU
28	8.17	350	609.1381	**301**, 463	$C_{27}H_{29}O_{16}$	Quercetin hexosyl rhamnoside	QHexRham2
29	8.28	355	463.0824	301	$C_{21}H_{19}O_{12}$	Quercetin-3-O-glucoside[a]	QGlu

Table 1 (continued)

Peak	RT	PDA max	m/z [M − H]⁻	MS²	Predicted formula [M − H]⁻	Putative identity	Short name
30	8.81	355	433.0713	301	$C_{20}H_{17}O_{11}$	Quercetin pentoside	QPent
31	8.84	355	447.0866	**285**	$C_{21}H_{19}O_{11}$	Kaempferol galactoside	KGal
32	8.94	350	549.0812	**505**, 301	$C_{24}H_{21}O_{15}$	Quercetin malonyl glucoside	QMalGlu
33	8.96	350	447.0511	301, **315**	$C_{20}H_{16}O_{12}$	Methylellagic acid pentoside	MeEllPent
34	9.09	365	475.0455	301	$C_{21}H_{15}O_{13}$	Ellagic acid acetyl pentoside	EllAcetylPent
35	9.17	345	461.0656	285	$C_{21}H_{17}O_{12}$	Kaempferol glucuronide	KGlcU
36	9.18	345	447.0862	285	$C_{21}H_{19}O_{11}$	Kaempferol glucoside	KGlu
37	9.41	360	463.0812	301	$C_{21}H_{19}O_{12}$	Quercetin hexoside	QHex

*Doubly charged ion, m/z [M-H]²⁻

[a]component confirmed against standard

MS² fragments in bold are the dominant products

individual distributions and a negative correlation overall. These plots will be discussed in more detail in the QTL results below.

3.3 QTL mapping

The QTL analysis generally showed a good consistency over years, but there were also some significant QTL × year interactions. These were however generally associated with different sizes of QTL effects, rather than a different direction of the effect. QTLs were detected in each year on raspberry LG1-LG6, but not on LG7 (Table 3). This is in accordance with previous studies on this population (e.g. Hackett et al., 2018), which have seldom found QTLs on LG7. Figure 1 shows linkage maps with one- and two-LOD support intervals for the QTLs. The term 'major' QTL is used to refer to a QTL with LOD greater than 10, 'strong' QTL for a QTL with LOD between 5 and 10 and 'minor' QTL for a QTL with LOD above the permutation threshold of 3.8 but below 5. The results presented below are from the over-year averages unless otherwise stated. Linkage groups LG1, LG4 and LG6, with QTLs large enough to give bimodal distributions for some metabolites, are discussed first, followed by the groups with smaller QTLs.

3.3.1 LG1

There was a very significant region at the bottom of LG1 for all three years, with many metabolites mapping to positions from 106 to 113 cM, with LODs up to 55.1 and the % variance explained (R^2) by this QTL up to 67%. This is close to the RibHLH marker at 113 cM, which was identified by McCallum et al. (2010) as close to a QTL for a smaller set of anthocyanins than measured in this study. One group of metabolites had major QTLs (explaining between 22 and 67% of the trait variance) and showed a pattern where the levels were close to zero for one genotype (bd), higher for bc and highest for ac and ad (Fig. 2A). This pattern was shown by the anthocyanins, $CyC_{34}H_{40}O_{21}$, CySophRham/ CyGluRut, CyXylRut/CySamRham, PelGluRut, PelRut and CyRut/PelSoph and the flavonols, QHexRham2, QGluRut2, QGluRut1 and QHexRham1. All these flavonoid components contain attached rhamnose groups and even the unidentified cyanidin derivative $CyC_{34}H_{40}O_{21}$ gave a neutral loss of 146 under MS fragmentation which is consistent with the presence of a rhamnose group (Table 1). The presence of the major QTLs is responsible for the bimodal distribution shown by all these metabolites apart from CyRut/ PelSoph. The genotype bd, with the lowest values, represents a clear separate cluster in the scatter plot matrices (see supplementary Figs S2A, S2B, & S2D). The metabolites are all positively correlated, both for the overall correlation and for the correlation within the genotypes ac, ad and bc, which have higher values of each metabolite than bd (Fig. 2A).

A further set of metabolites mapped to the same location (109–114 cM) on LG1, but with smaller QTLs, and gave a different pattern of segregation and more variation over the years (Fig. 2B). All these components are flavonoid

Table 2 Summary statistics for metabolite levels and total anthocyanin data in the parents and offspring across the 3 years of harvest

Short name	Year	Latham	Moy	Mean	Min	Max	SD	H_G	H_GY	BM
TA	10	10.481	6.922	8.765	1.924	25.761	3.820	0.593	0.096	N
	11	16.084	NA	9.703	0.000	45.447	5.284			N
	15	8.812	10.093	8.320	1.584	26.799	3.777			N
CySoph	10	7.631	7.195	8.366	1.826	26.167	4.424	0.715	0.181	N
	11	7.301	NA	6.047	0.815	20.078	2.989			N
	15	5.257	6.135	6.151	1.076	15.954	2.508			N
CyGlu	10	2.132	1.950	2.139	0.503	9.710	1.190	0.771	0.089	N
	11	1.421	NA	1.406	0.258	5.012	0.826			N
	15	1.393	1.894	1.643	0.257	4.603	0.856			N
CySophRham/ CyGluRut	10	6.590	2.520	3.364	0.000	16.180	2.999	0.887	0.186	Y
	11	5.014	NA	1.925	0.000	11.318	1.870			Y
	15	2.637	2.579	1.526	0.000	7.150	1.626			Y
CySam	10	0.044	**0.252**	0.160	0.022	0.759	0.116	0.799	0.119	N
	11	0.036	NA	0.108	0.000	0.561	0.090			N
	15	0.025	**0.159**	0.109	0.007	0.378	0.079			N
CyC$_{28}$H$_{30}$O$_{17}$	10	0.146	0.197	0.322	0.000	1.465	0.239	0.634	0.066	N
	11	0.086	NA	0.130	0.007	0.574	0.097			N
	15	0.190	0.287	0.320	0.047	1.486	0.196			N
CyRut/ PelSoph	10	**3.075**	1.117	1.868	0.114	6.489	1.174	0.830	0.053	N
	11	2.796	NA	1.165	0.056	4.504	0.806			N
	15	**1.840**	1.305	1.025	0.069	5.221	0.748			N
CyXylRut/ CySamRham	10	0.063	0.112	0.095	0.000	0.576	0.101	0.879	0.155	Y
	11	0.042	NA	0.047	0.000	0.417	0.060			Y
	15	0.023	0.078	0.042	0.000	0.301	0.054			Y
PelGlu	10	**0.308**	0.063	0.189	0.014	1.406	0.195	0.795	0.205	N
	11	0.264	NA	0.123	0.000	1.512	0.179			N
	15	**0.224**	0.148	0.144	0.010	0.975	0.130			N
CyC$_{34}$H$_{40}$O$_{21}$	10	0.129	0.065	0.131	0.000	0.646	0.136	0.885	0.109	Y
	11	0.051	NA	0.038	0.000	0.343	0.049			Y
	15	0.099	0.107	0.081	0.000	0.978	0.106			Y
PelGluRut	10	**1.807**	0.149	0.350	0.000	3.078	0.397	0.891	0.135	Y
	11	1.965	NA	0.239	0.000	1.979	0.310			Y
	15	**1.066**	0.218	0.166	0.000	1.582	0.224			Y
PelSam	10	0.004	0.002	0.010	0.000	0.139	0.016	0.718	0.227	Y
	11	0.000	NA	0.004	0.000	0.113	0.012			Y
	15	0.005	0.005	0.007	0.000	0.053	0.008			N
PelRut	10	**0.221**	0.030	0.078	0.000	0.514	0.091	0.880	0.099	Y
	11	0.155	NA	0.035	0.000	0.587	0.060			Y
	15	**0.123**	0.082	0.038	0.000	0.554	0.063			Y
SanguiinH10	10	0.294	0.722	0.397	0.154	1.000	0.114	0.572	0.046	N
	11	0.104	NA	0.131	0.037	0.352	0.055			N
	15	0.225	0.309	0.301	0.073	0.854	0.112			N
QGluRut1	10	**0.074**	0.023	0.016	0.000	0.149	0.021	0.847	0.134	Y
	11	0.065	NA	0.012	0.000	0.073	0.016			Y
	15	**0.081**	0.045	0.023	0.000	0.171	0.031			Y
QGluRut2	10	**0.158**	0.070	0.064	0.000	0.491	0.070	0.894	0.164	Y
	11	0.136	NA	0.033	0.000	0.201	0.036			Y
	15	**0.119**	0.073	0.048	0.000	0.301	0.057			Y
QDihex1	10	**0.097**	0.015	0.022	0.000	0.156	0.026	0.861	0.139	Y
	11	0.069	NA	0.019	0.000	0.211	0.023			Y
	15	**0.140**	0.035	0.043	0.000	0.194	0.042			Y

Table 2 (continued)

Short name	Year	Latham	Moy	Mean	Min	Max	SD	H_G	H_GY	BM
QDihex2	10	0.046	0.024	0.050	0.001	0.417	0.068	0.837	0.171	N
	11	0.032	NA	0.030	0.000	0.377	0.046			N
	15	0.047	0.024	0.055	0.003	0.380	0.062			N
EllagicPent1	10	0.162	0.215	0.171	0.047	0.360	0.058	0.798	0.016	N
	11	0.063	NA	0.069	0.018	0.157	0.026			N
	15	0.194	0.242	0.243	0.075	0.568	0.095			N
EllagicPent2	10	0.073	0.031	0.048	0.000	0.160	0.040	0.843	0.107	Y
	11	0.026	NA	0.022	0.002	0.075	0.016			Y
	15	0.120	0.044	0.095	0.009	0.413	0.079			Y
Lambertianin	10	2.699	3.578	2.457	0.847	4.286	0.629	0.648	0.071	N
	11	0.578	NA	0.579	0.151	1.315	0.222			N
	15	0.917	1.223	1.086	0.419	2.111	0.321			N
Ellagic	10	0.057	0.040	0.044	0.009	0.155	0.021	0.446	0.000	N
	11	0.013	NA	0.011	0.001	0.101	0.008			N
	15	0.391	0.284	0.308	0.039	1.505	0.219			N
SanguiinH2	10	0.036	0.034	0.029	0.012	0.058	0.008	0.544	0.130	N
	11	0.017	NA	0.015	0.003	0.033	0.006			N
	15	0.018	0.014	0.015	0.007	0.048	0.005			N
SanguiinH6	10	5.496	5.965	5.020	2.147	7.419	0.914	0.630	0.085	N
	11	1.297	NA	1.323	0.604	2.889	0.391			N
	15	2.918	2.992	2.977	1.695	4.682	0.598			N
QHexPent	10	0.008	0.075	0.037	0.000	0.193	0.041	0.864	0.250	Y
	11	0.005	NA	0.028	0.000	0.134	0.031			Y
	15	0.014	0.078	0.064	0.001	0.245	0.060			N
QHexRham1	10	0.074	0.035	0.021	0.000	0.147	0.031	0.802	0.278	Y
	11	0.027	NA	0.012	0.000	0.077	0.017			Y
	15	0.041	0.044	0.023	0.000	0.156	0.032			N
QGal	10	0.232	0.086	0.072	0.001	0.376	0.069	0.838	0.182	Y
	11	0.106	NA	0.053	0.000	0.263	0.052			Y
	15	0.142	0.109	0.104	0.000	0.450	0.096			N
QGlcU	10	**1.356**	0.431	0.457	0.038	1.186	0.244	0.417	0.128	N
	11	0.506	NA	0.280	0.015	1.042	0.151			N
	15	**0.633**	0.345	0.516	0.096	1.356	0.204			N
QHexRham2	10	0.067	0.036	0.034	0.000	0.145	0.031	0.894	0.148	Y
	11	0.031	NA	0.015	0.000	0.105	0.016			Y
	15	0.038	0.039	0.023	0.000	0.178	0.027			Y
QGlu	10	0.124	0.097	0.062	0.011	0.255	0.035	0.590	0.101	N
	11	0.058	NA	0.045	0.002	0.164	0.025			N
	15	0.057	0.060	0.077	0.015	0.221	0.035			N
QPent	10	0.009	**0.134**	0.064	0.000	0.315	0.076	0.857	0.219	Y
	11	0.002	NA	0.046	0.000	0.261	0.051			Y
	15	0.013	**0.095**	0.107	0.001	0.431	0.101			Y
KGal	10	0.029	0.007	0.006	0.000	0.050	0.008	0.846	0.185	Y
	11	0.010	NA	0.005	0.000	0.050	0.006			Y
	15	0.014	0.010	0.009	0.000	0.062	0.010			Y
QMalGlu	10	0.038	0.033	0.018	0.001	0.087	0.011	0.616	0.051	N
	11	0.018	NA	0.014	0.000	0.046	0.009			N
	15	0.013	0.018	0.018	0.001	0.061	0.012			N
MeEllPent	10	0.055	0.074	0.062	0.015	0.169	0.029	0.817	0.000	N
	11	0.018	NA	0.022	0.004	0.069	0.011			N
	15	0.079	0.082	0.086	0.018	0.251	0.041			N

Table 2 (continued)

Short name	Year	Latham	Moy	Mean	Min	Max	SD	H_G	H_GY	BM
EllAcetylPent	10	0.087	0.123	0.072	0.003	0.235	0.048	0.815	0.214	N
	11	0.027	NA	0.021	0.000	0.092	0.017			N
	15	0.080	0.116	0.078	0.002	0.284	0.051			N
KGlcU	10	**0.210**	0.038	0.044	0.001	0.204	0.034	0.367	0.077	N
	11	0.068	NA	0.028	0.000	0.169	0.021			N
	15	**0.073**	0.032	0.047	0.003	0.177	0.030			N
KGlu	10	0.043	0.027	0.008	0.000	0.080	0.013	0.311	0.472	Y
	11	0.014	NA	0.010	0.000	0.045	0.008			N
	15	0.016	0.016	0.020	0.001	0.071	0.012			N
QHex	10	0.025	0.010	0.012	0.001	0.038	0.008	0.763	0.092	N
	11	0.010	NA	0.005	0.000	0.022	0.004			N
	15	0.013	0.007	0.011	0.002	0.045	0.007			N

10, 11 and 15 represent samples from years 2010, 2011 and 2015

TA = total anthocyanin content by the colorimetric method, values in mg cyanidin glucoside equivalents/100 g fresh weight. Metabolites are denoted by their abbreviated names, see Table 1. NA = value not available due to no Moy samples that year. Mean, Min, Max, SD = mean, minimum, maximum and standard deviation of the offspring

H_G = generalised heritability of the genotype ($\log_{10}$ scale); H_GY = generalised heritability of the genotype × year effect ($\log_{10}$ scale). BM = Y if the metabolite distribution was clearly bimodal, N otherwise. Metabolites in bold were significantly higher in that parent (p < 0.05) in that year

glucosides with glucose groups directly attached to the flavonoid aglycone (e.g. CyGlu) or with further glycosyl groups attached to the flavonoid glucosyl core (e.g. cyanidin 3-sambubioside; CySam or cyanidin sophoroside; CySoph). A major QTL for the flavonol QDihex2 was also detected in this region for all three years. Major QTLs were detected in 2010, with smaller QTLs that are consistent in position and sign in 2011 and 2015, for the anthocyanins CySoph, $CyC_{28}H_{30}O_{17}$, and PelSam. A major QTL for CySam and strong QTLs for CyGlu and PelGlu were also detected in this region in 2010 only, together with a minor QTL nearby, at 96 cM, for QGlu in 2010 only. For all these metabolites, the difference in the genotype means is due to the allele received from Latham, with lower values for the 'a' allele (in contrast to the flavonoids shown in Fig. 2A), where the 'a' allele had high levels.

All the anthocyanins that map here showed positive correlations (Fig S2A), particularly PelGlu and PelSam, with a correlation of 0.90 in 2010, and CyGlu and CySam, with a correlation of 0.88. The correlations were also high in 2011 and 2015. These correlations may be expected as the glucosides could be metabolic precursors of the sambubioside derivatives, which have xylose groups added at position 2 of the glucose moiety which is attached to the anthocyanidin (Bradish et al., 2012). In addition, there was a strong correlation between PelGluRut and PelRut (0.95) which also could reflect a -precursor relationship (Fig. S2A). The strong correlation (0.93) between the potential cyanidin derivative $CYC_{34}H_{44}O_{21}$ and CyXylRut could also suggest

a precursor-intermediate relationship. However, the strong correlation (0.95) between PelRut and CyXylRut is not so readily explained.

For some pairs of metabolites, the overall correlations do not give a satisfactory representation of the true relationships between these anthocyanins and those with the alternative pattern above. For example, the overall correlation between CyXylRut/CySamRham and CySoph was − 0.50 in 2010 (Fig. S2A). However, apart from the cluster of values from genotype bd, which has very low levels of CyXylRut/CySamRham (see red arrow in Fig. S2A), the relationship within the other three clusters is positive. Indeed, CySam and CyXylRut/CySamRham (overall r = − 0.18 in 2010) and CyGlu and CyXylRut/CySamRham (overall r = − 0.10 in 2010) showed the same pattern with low levels of CyXylRut/CySamRham due to the genotype bd.

Supplementary Figure S2D shows the relationships between anthocyanins and flavonols. There were many near-zero correlations, but some components showed high positive correlations. For example, CySophRham/CyGluRut had strong correlations with QGluRut2 (r = 0.97), QHexRham2 (0.97) and QGluRut1 (0.72). A similar pattern of correlation was seen between PelGluRut with QGluRut2 (0.90), QHexRham (0.89) and QGluRut1 (0.61) and between PelRut with QGluRut2 (0.91), QHexRham (0.93) and QGluRut1 (0.71). In all these cases, the lower correlation with QGluRut1 was influenced by low levels associated with one genotype. The biochemical feature that links these components is

Table 3 QTLs detected for each metabolite. This is based on the average values for each genotype over the three years

LG	Trait	Pos (cM)	LOD	%var	M_ac	M_ad	M_bc	M_bd	se_ac	se_ad	se_bc	se_bd	t_P1	t_P2	t_dom
LG1	$CyC_{34}H_{40}O_{21}$	113	55.1	66.5	− 2.87	− 3.11	− 4.73	− 9.63	0.339	0.262	0.337	0.278	14.2	8.7	7.9
	CyXylRut/CySamRham	113	52.3	64.9	− 3.12	− 3.38	− 4.89	− 9.34	0.322	0.248	0.320	0.264	13.8	8.4	7.5
	CySophRham/CyGluRut	113	53.4	64.8	0.66	0.34	− 1.09	− 6.57	0.376	0.290	0.374	0.309	13.3	8.9	7.9
	QHexRham2	113	50.1	63.2	− 3.76	− 4.04	− 4.87	− 8.13	0.232	0.179	0.231	0.191	12.9	8.7	7.4
	PelRut	109	42.4	61.0	− 3.91	− 3.85	− 5.25	− 10.10	0.380	0.282	0.360	0.292	12.0	7.6	7.7
	QGluRut2	109	50.2	65.6	− 3.41	− 3.49	− 4.17	− 9.61	0.347	0.257	0.329	0.267	11.9	9.5	9.3
	PelGluRut	113	47.2	60.4	− 2.06	− 2.15	− 3.61	− 9.47	0.433	0.334	0.430	0.355	11.8	7.9	7.7
	QGluRut1	113	23.2	39.0	− 5.76	− 5.67	− 6.30	− 10.30	0.419	0.323	0.416	0.344	7.1	5.4	5.6
	CyRut/PelSoph	106	13.6	27.3	0.40	0.25	0.00	− 0.51	0.099	0.075	0.095	0.077	6.9	4.0	2.2
	QHexRham1	113	10.9	21.6	− 4.84	− 4.93	− 5.58	− 6.91	0.269	0.208	0.268	0.221	5.8	3.0	2.6
	Lambertianin	113	3.9	7.5	0.14	0.11	0.04	− 0.03	0.038	0.029	0.037	0.031	3.8	1.7	0.8
	CyGlu	114	4.0	7.7	0.28	0.25	0.48	0.52	0.067	0.052	0.067	0.055	− 4.2	− 0.1	− 0.6
	CySam	114	7.1	14.3	− 2.56	− 2.62	− 2.15	− 2.09	0.096	0.075	0.095	0.079	− 5.7	0.0	− 0.7
	PelSam	114	9.6	19.6	− 7.01	− 6.66	− 5.65	− 5.69	0.189	0.147	0.187	0.155	− 7.1	− 0.9	1.2
	$CyC_{28}H_{30}O_{17}$	112	10.2	20.8	− 1.85	− 1.86	− 1.41	− 1.41	0.071	0.056	0.072	0.058	− 7.2	0.1	0.0
	CySoph	112	11.6	23.0	1.63	1.62	1.94	2.01	0.053	0.042	0.054	0.043	− 7.4	− 0.6	− 1.0
	QDihex2	113	29.1	48.7	− 4.49	− 4.33	− 3.49	− 2.86	0.115	0.089	0.115	0.095	− 12.3	− 3.9	− 2.3
	QGlu	96	4.8	9.7	− 3.10	− 3.05	− 2.91	− 2.81	0.057	0.043	0.059	0.046	− 4.5	− 1.5	− 0.5
	QHex	36	5.0	9.2	− 5.06	− 5.26	− 4.89	− 4.78	0.082	0.082	0.082	0.074	− 4.1	0.6	− 2.0
LG2	QDihex2	5	3.9	7.3	− 3.50	− 3.63	− 3.73	− 4.25	0.132	0.132	0.144	0.138	3.1	2.4	1.4
	CySoph	12	7.3	14.7	1.91	1.93	1.77	1.57	0.049	0.050	0.051	0.050	5.1	1.9	2.3
	TA	16	6.5	13.5	2.32	2.35	2.19	2.09	0.038	0.038	0.039	0.038	5.3	1.0	1.7
	$CyC_{28}H_{30}O_{17}$	19	5.7	11.9	− 1.51	− 1.47	− 1.68	− 1.91	0.066	0.065	0.069	0.066	4.6	1.5	2.0
	QGlcU	105	4.7	8.8	− 1.08	− 1.19	− 0.88	− 1.06	0.050	0.050	0.048	0.047	− 3.4	3.0	0.7
	KGlcU	105	4.2	7.8	− 3.52	− 3.75	− 3.35	− 3.51	0.067	0.067	0.065	0.063	− 3.2	3.0	− 0.5
LG3	MeEllPent	61	4.0	7.6	− 3.08	− 3.39	− 3.22	− 3.06	0.050	0.069	0.054	0.065	− 1.7	1.3	− 4.2
	SanguiinH10	66	5.1	8.89	− 0.632	− 0.713	− 0.629	− 0.607	0.013	0.018	0.014	0.017	− 3.8	2.0	− 3.5
	KGlu	95	6.4	13.2	− 5.79	− 5.64	− 4.98	− 4.89	0.128	0.148	0.137	0.167	− 5.5	− 0.9	0.2
	QGlu	99	6.2	12.5	− 3.11	− 3.04	− 2.85	− 2.77	0.043	0.049	0.047	0.058	− 5.4	− 1.5	− 0.1
	Lambertianin	124	4.8	9.6	0.15	0.10	0.00	− 0.04	0.029	0.033	0.033	0.038	4.7	1.4	− 0.1
	SanguiinH6	124	5.6	11.4	0.99	0.97	0.89	0.83	0.022	0.025	0.025	0.029	5.0	1.8	0.9
LG4	PelSam	89	7.9	15.4	− 5.59	− 6.42	− 5.93	− 6.97	0.175	0.156	0.185	0.173	2.7	5.5	0.6
	CySoph	88	4.0	7.5	1.95	1.82	1.79	1.64	0.054	0.048	0.055	0.052	3.3	2.8	0.2
	TA	58	4.9	9.8	2.36	2.22	2.25	2.09	0.038	0.036	0.040	0.043	3.1	3.9	0.1
	PelGlu	58	9.0	18.3	− 1.93	− 2.22	− 2.48	− 2.96	0.107	0.102	0.111	0.120	5.9	3.5	0.9
	CyGlu	47	16.2	31.2	0.66	0.52	0.20	0.05	0.051	0.047	0.053	0.055	9.1	3.0	0.1
	CyRut/PelSoph	34	4.9	9.6	0.28	− 0.05	0.07	− 0.34	0.090	0.088	0.102	0.098	2.6	4.0	0.4
	CySam	30	18.3	35.7	− 2.07	− 1.99	− 2.71	− 2.83	0.072	0.068	0.079	0.075	10.2	0.4	1.4
	QDihex2	78	6.9	13.7	− 3.29	− 3.89	− 3.66	− 4.34	0.129	0.126	0.132	0.141	3.2	5.0	0.3
	QGlu	59	8.6	17.5	− 2.79	− 2.93	− 2.96	− 3.23	0.046	0.045	0.049	0.051	4.9	4.4	1.4
	KGal	28	45.7	65.5	− 5.23	− 5.01	− 8.48	− 7.03	0.141	0.137	0.163	0.151	18.2	− 5.8	− 4.2
	QGal	29	45.4	65.0	− 2.58	− 2.49	− 5.54	− 3.90	0.124	0.120	0.143	0.133	17.2	− 6.8	− 6.1
	QDihex1	29	40.5	60.0	− 3.92	− 3.76	− 7.41	− 4.64	0.160	0.155	0.185	0.172	13.3	− 8.9	− 7.9
	QMalGlu	28	16.1	31.5	− 4.01	− 4.04	− 4.41	− 4.56	0.048	0.047	0.056	0.052	9.2	1.8	1.1
	QHexRham1	26	12.5	25.4	− 4.52	− 5.11	− 6.97	− 6.29	0.225	0.214	0.259	0.237	7.9	− 0.2	− 2.8
	QGluRut1	25	10.4	21.5	− 5.52	− 6.31	− 9.72	− 7.90	0.407	0.382	0.479	0.425	7.0	− 1.2	− 3.2
	QHex	29	19.1	36.0	− 5.03	− 5.35	− 4.36	− 5.03	0.064	0.062	0.074	0.069	− 7.4	7.5	2.6
	QGlcU	29	5.5	11.0	− 1.10	− 1.13	− 0.81	− 1.11	0.046	0.045	0.053	0.050	− 3.2	3.4	2.9
	QHexPent	29	67.7	80.2	− 7.53	− 3.18	− 4.87	− 3.04	0.132	0.128	0.152	0.142	− 10.3	− 22.8	9.3
	QPent	33	67.1	78.6	− 6.69	− 2.54	− 5.46	− 2.70	0.136	0.130	0.151	0.147	− 3.9	− 25.0	5.0
LG5	MeEllPent	8	10.2	21.0	− 3.37	− 3.22	− 3.25	− 2.88	0.052	0.056	0.055	0.051	− 4.5	− 5.0	− 2.1
	EllAcetylPent	16	27.2	47.4	− 2.87	− 3.10	− 3.21	− 4.34	0.087	0.095	0.098	0.087	8.8	7.5	5.0
	KGlu	19	5.2	10.6	− 5.01	− 5.27	− 5.21	− 5.93	0.142	0.153	0.149	0.138	3.0	3.4	1.6
	KGlcU	48	4.1	8.0	− 3.49	− 3.69	− 3.31	− 3.59	0.068	0.064	0.067	0.064	− 2.1	3.8	0.6
	PelSam	75	6.7	13.8	− 6.71	− 6.74	− 5.78	− 5.72	0.186	0.158	0.170	0.180	− 5.7	− 0.1	− 0.3
	PelGlu	76	6.3	12.7	− 2.76	− 2.55	− 2.06	− 2.02	0.121	0.102	0.113	0.123	− 5.5	− 1.1	0.8
	SanguiinH2	78	6.0	12.0	− 4.00	− 3.97	− 4.06	− 4.18	0.032	0.027	0.031	0.032	4.4	1.6	2.6
LG6	MeEllPent	65	6.0	12.3	− 3.04	− 3.03	− 3.34	− 3.32	0.055	0.054	0.057	0.061	5.4	− 0.4	− 0.1
	EllagicPent1	80	13.4	26.8	− 1.90	− 1.88	− 2.27	− 2.25	0.043	0.042	0.046	0.048	8.5	− 0.4	0.0
	EllAcetylPent	80	7.5	15.5	− 3.09	− 3.07	− 3.75	− 3.82	0.111	0.110	0.119	0.125	6.1	0.2	0.4
	EllagicPent2	80	55.2	73.9	− 3.93	− 3.96	− 2.76	− 2.71	0.051	0.050	0.054	0.057	− 23.2	− 0.2	− 0.8
	KGlu	81	11.9	23.7	− 5.01	− 4.81	− 5.88	− 5.99	0.129	0.127	0.137	0.143	7.8	− 0.3	1.2

Pos is the position of the maximum LOD, %var is the %variance explained by that QTL. M_ac, M_ad, M_bc and M_bd are the estimated means for offspring genotypes ac, ad, bc and bd respectively and se_ac etc. are the corresponding standard errors. t_P1 is a t-test for the significance of the Latham additive effect, t_P2 is a t-test for the Glen Moy additive effect and t_dom for the dominance effect. The significant effects are shaded

Table 3 (continued)

red if they are positive or green if they are negative. Anthocyanin components are shaded in red, flavonols in blue, with ellagitannins and ellagic acid derivatives in grey

the addition of similar glycosyl groups to the flavonol or anthocyanidin core, so it is intriguing to suggest that these steps are under common genetic control. In addition, QDihex2 also had a correlation of 0.82 with CySoph, again suggesting a commonality of added glycosyl groups.

There were some candidate genes underlying this QTL region (LG1, 106–114 cM) that could be involved in the control of anthocyanin and flavonol biosynthesis. Firstly, there was a homologue (>80% homology with genes from *Rosa chinensis*) of a leucine-rich repeat protein kinase that has been associated with the accumulation of polyphenols in Tannat grapes (Da Silva et al., 2013). There was a homologue of a MYB39 transcription factor (>80% homology to genes from *R. chinensis* and *Fragaria vesca*) that has been associated with microRNA control of phenylpropanoid biosynthesis in *Arabidopsis* (AT3G61250; Sharma et al., 2016) and another MYB3-like transcription factor (homology to AT1G22640.1; ≥85% homology with *R. chinensis* and *F. vesca* genes) that was upregulated in polyphenol-rich blueberry skins vs flesh (Plunkett et al., 2018). There was a homologue of a C2H2-type zinc finger protein (>75% homology to *R. chinensis* and *F. vesca* genes), which was found to be upregulated in ripe versus unripe grapes (Ahn et al., 2019). Also, there was a homologue to a structural biosynthetic gene that could be directly involved in the formation of the anthocyanidins, i.e. an anthocyanidin reductase (AT1G61720.1; ANR; >80% homology with Prunus genes), which was upregulated in the polyphenol-rich skin of blueberries versus blueberry flesh (Plunkett et al., 2018) and implicated in polyphenol accumulation in grapes (Bogs et al., 2005).

Also, there was a homologue to an Arabidopsis gene (~80% homology to *R. chinensis and F. vesca* genes) that encoded a UDP-glycosyl transferase (UDPGT) that has been implicated in polyphenol biosynthesis in ripe grape berries (Ahn et al., 2019). This could be responsible for the attachment of glycosyl groups directly to the flavonoid aglycone (e.g. to the cyanidin group in CyGlu) and/or in adding extra glycosyl groups to the flavonoid glucoside (as discussed for components in Fig. 2B).

Despite the QTLs for individual anthocyanins, no QTL was detected for Total Anthocyanin content (TA) on LG1. This appears to be because the increase in the levels of some of the individual anthocyanins is balanced by the decrease in the levels of others. This was also the case in a previous study (McCallum et al., 2010) which examined

levels of eight major raspberry anthocyanins. A strong QTL for QHex was also detected at 36 cM on LG1. This was an effect mainly of the Latham parent.

3.3.2 LG4

There was a very significant area on LG4 from 25 to 33 cM for all three years, especially for the flavonol metabolites, with QTLs with LODS up to 67.7 and R^2 up to 80%. This is close to the region detected by Kassim et al. (2009) and McCallum et al. (2010). Markers ERubLR_SQ8.1_H09HMGR, JHIRi49055_GIG and ERubLR_11_2C02_ADH are in this region, also marker RiCAD is nearby. The flavonol metabolites mapping to this region showed three distinct genotype patterns. The most significant metabolites were QHexPent and QPent, which had major QTLs in this region in all three years (with LODs of 67.1 to 67.7 and R^2 from 78.6 to 80.2%) with associated bimodal distributions. For these metabolites genotype ac had the lowest level, then bc, with ad and bd highest (Fig. 3A). They are also very highly correlated (r = 0.90 in 2010; Fig. S2B) which could indicate a precursor- product relationship. QHex also had a major QTL here but the genotype pattern was different, with ad having the lowest level, ac and bd having intermediate levels and bc having the highest level. QGlcU also had a strong QTL in this region with the same genotype pattern as QHex.

The relationship between the pentose containing flavonols and the genotypes could be explained if this QTL overlaid genes involved in sugar addition to the flavonols. However, the metabolic relationship between QHex and QGlcU is less obvious.

Other flavonols KGal, QGal and QDihex1 also had major QTLs in this region, with LODs of 40.5 to 45.7 and R^2 from 60.0 to 65.5%. For these, the lowest levels were for genotype bc, higher for bd and the highest for ac and ad (Fig. 3B). This region was associated with the observed bimodal distribution for these metabolites. KGal, QGal and QDihex1 were positively correlated, with correlations of at least 0.88 in 2010 (Fig S2B), and similarly high for the other years. QHexRham1 and QGluRut1 also had major QTLs here (LODs of 10.4–12.5) with a similar genotype pattern. QMalGlu also mapped with a major QTL here, with a similar pattern but less distinction between bc and bd. It is intriguing that these flavonol components share attached hexosyl groups (i.e. either glucosyl

Fig. 1 Linkage maps showing QTL locations. The boxes show the one-LOD support interval, and the whiskers show the two-LOD interval. Major QTLs are shown as solid boxes, strong QTLs as diagonally filled boxes and minor QTLs as unfilled boxes. Red = anthocyanins, blue = flavonols, black = other traits. **A–F** show linkage groups (LG) 1–6

Fig. 1 (continued)

C
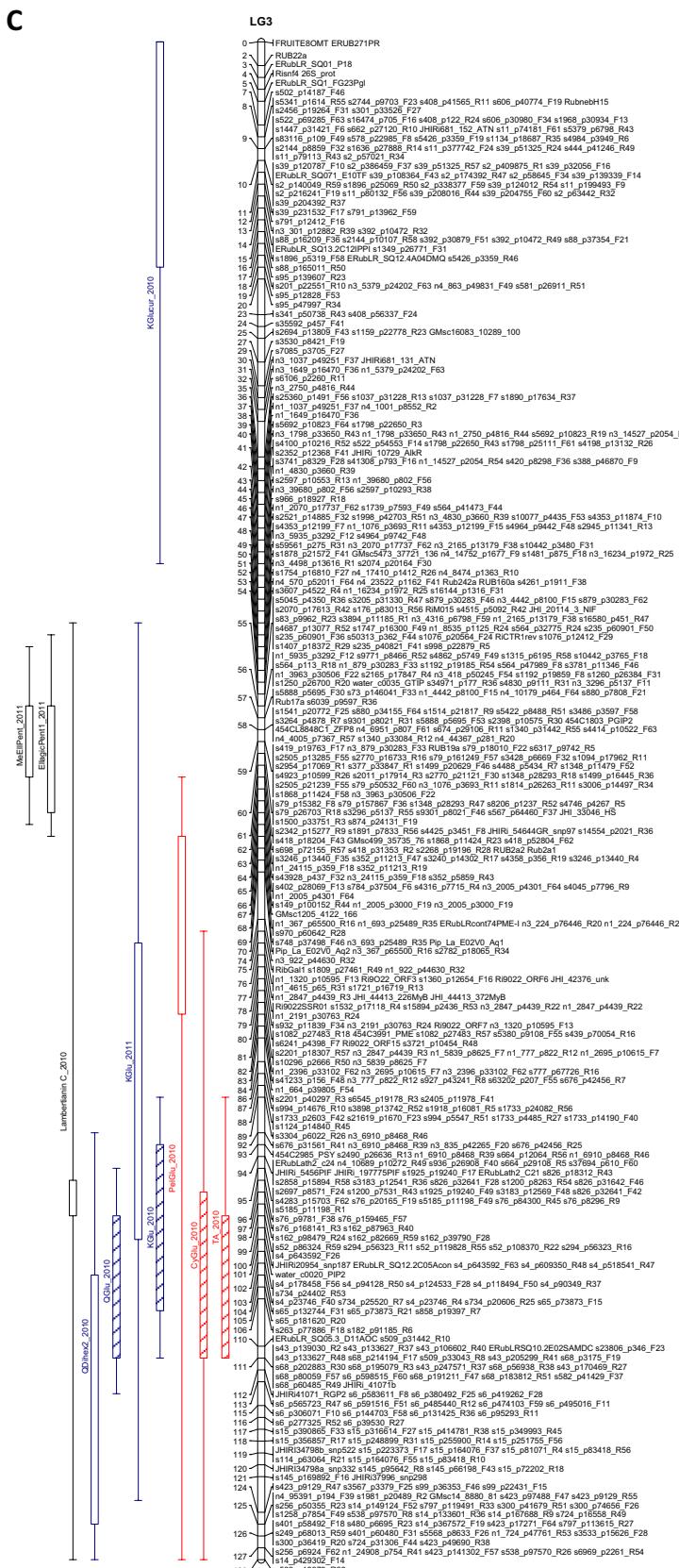
Fig. 1 (continued)
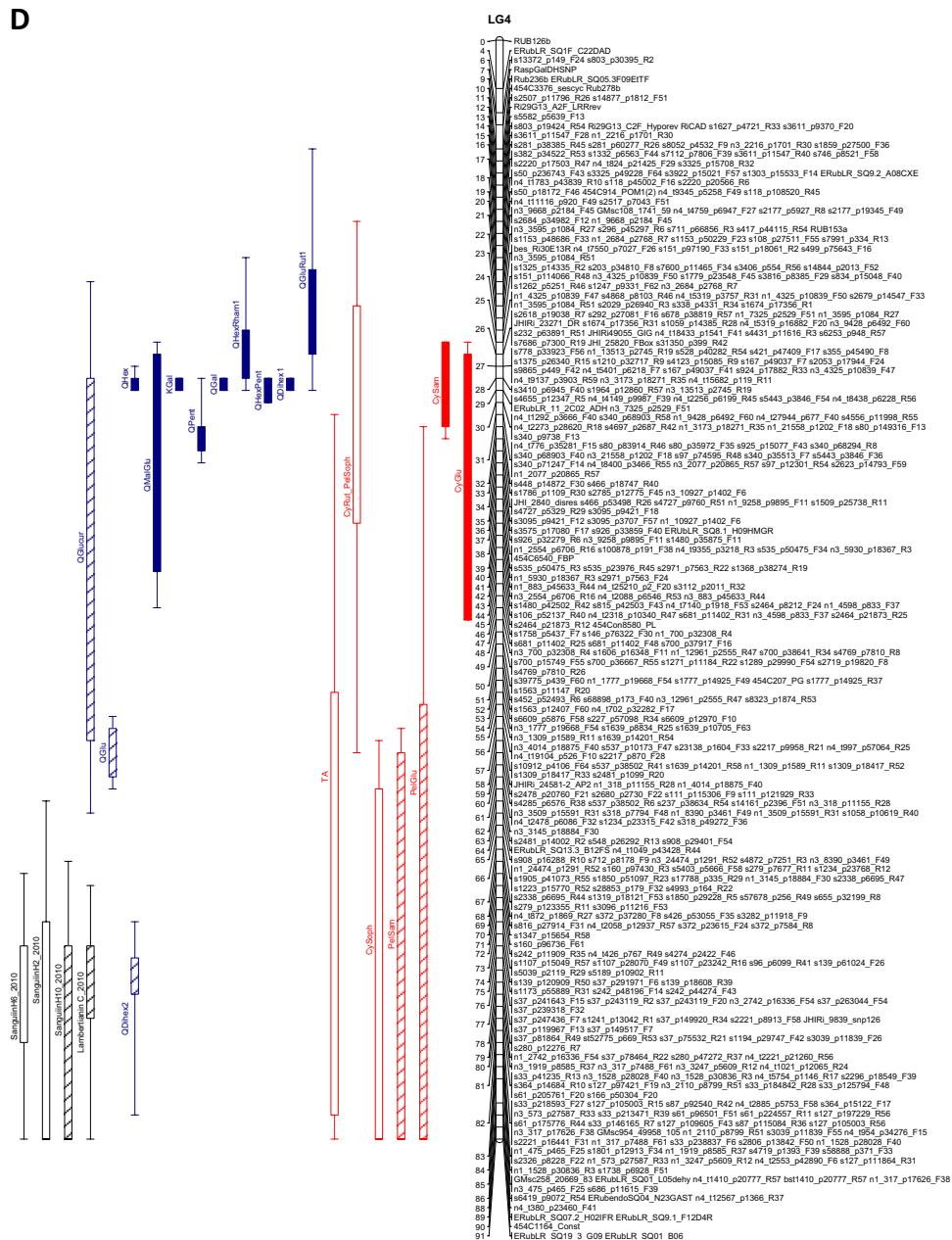
Fig. 1 (continued)

or galactosyl) and perhaps this region controls the addition of such sugar groups.

Scatterplots of these flavonols show a strong structure with up to four groups corresponding closely to the genotypes at about 30 cM on LG4 (Fig. 4). This is generally backed up by the scatter plots for flavonols shown in Fig. S2B. The ac genotype (black) and ad genotype (red), which inherit the "a" allele from Latham, both had high levels of KGal, QGal, QDihex1, QMalGlu and QHex but the ac genotype had much higher levels than the ad genotype for QPent

and QHexPent. The ac genotype and bc genotype (green) inheriting allele c from Glen Moy both had high QPent and QHexPent but the bc genotype (green) had lower levels of KGal, QGal, QDihex1 and QMalGlu. There were strong correlations (~ 0.9) between the levels of QGal and QDihex1 across the progeny in 2010, which could reflect precursor relationships (Fig. 4 & Fig. S2). This fits with the addition of a second hexose group to QGal. A similar precursor-product relationship may exist between the kaempferol and quercetin hexosides 1 (probably Gal), which differ by one hydroxyl

E

LG5

Side bracket QTL labels (left to right): MeElPent · ElAcetyPent · KGlu · KGlucur · PelSam · PelGlu · SanguinH2

| Pos	Markers
0	RUB289a s1741_p18900_F43 s6131_p42_F24
1	s6131_p1400_F39 s11293_p1568_R19 s376_p49208_R19 s376_p49208_R48 n1_2571_p24320_R14
2	ERubLR_SQ06_2_E01pol n3_1110_p29233_R30 s3614_p10204_F30 n4_3375_p10641_F61
3	n1_7338_p4933_F38 n1_1110_p29233_R30 ERubEndoSQ07_P15XL1 RhM021
4	n3_7338_p4933_F38 s30663_p719_R20 s5480_p11037_F39 n3_1734_p792_F64
5	s2260_p10042_F58
8	MADS_01
9	n3_491_p42383_F14 s3793_p12381_F13
10	s3081_p12455_R20 s2792_p8523_R7 s7098_p6968_R18 n1_1734_p792_F64 n4_10859_p3937_F41
	ERubLR_SQ4_4_C101433 n1_29468_p889_R37 n3_29468_p889_R37 s47_p193053_F33 s437_p57372_F27
11	s271_p37634_F57 s4865_p11663_R26 s437_p54141_F49 s447_p100951_F15
12	n4_5580_p1571_R45 n3_2571_p24320_R14 n1_4004_p3252_R41
13	n1_7765_p1212_F7 s8484_p5110_F29 n1_4131_p7778_F48 n1_491_p42383_F14 n1_3793_p7375_R14
	n1_607_p20626_F13 s3081_p12455_R38 s1145_p24711_F35 s1646_p53095_F54 n4_4837_p23593_F47
14	s3192_p11663_F52 s2631_p10855_R51
15	n4_630_p51257_R14 n3_607_p20626_F13 n3_3762_p12611_F53 n3_4131_p7778_F48 n1_3436_p82_F61
	n1_3038_p15419_F20 n3_3436_p82_F61
16	n3_7765_p1212_F7 n3_3793_p7375_R14 n4_5347_p17987_F51 n3_884_p10907_F45 n4_2927_p24993_R44
	n3_26338_p1258_R20
17	s715_p101809_R15 s11536_p3456_R13 n3_3038_p15419_F20 s1757_p7917_R50 s3192_p11663_F62
	s54868_p407_F20 n4_7396_p1236_R19 s2350_p39473_F9 RUB268b
18	Ri25D10_262 n3_1399_p31063_R26 s2012_p43649_R52 s6670_p7252_F32 RUB35a
	n3_4004_p3252_R41 s7069_p3849_F42 n1_26338_p1258_R20 s5910_p1806_F57 n1_1399_p31063_R26
	n4_8718_p871_F45 n4_398_p59545_R39 s398_p46263_F44 n4_709_p33338_R33 s2350_p39473_F10
19	s3038_p15419_R11 s2903_p24220_R3 s11536_p3456_R37 n1_884_p10907_F45 n4_25239_p205_R15
	s398_p45247_F52
20	n4_10085_p3404_F13 s398_p25565_R10 s5901_p9439_F15 n4_4554_p2209_F62 s4355_p8090_R38
	n3_5425_p2724_F49 n1_5425_p2724_F49
21	n1_4620_p11928_R3 s42755_p112_F61 s4551_p7037_R55 s5523_p8458_R40 n1_1476_p9064_R3
	n1_16629_p1874_R40 s11290_p1872_F8
	s677_p20116_F7 s860_p18881_F31 s6007_p1751_F17 n1_5561_p1843_R35 n1_16629_p1874_R10
	n1_8764_p4764_R21 s4921_p5038_F38 s2709_p8927_F19 s860_p18881_F21 s522_p69285_F34
	s7000_p6933_R17 n1_1846_p25821_F38 n3_1846_p25821_F38 n3_16629_p1874_R10 s144860_p123_F21
	s1851_p8098_F21 n1_16629_p1874_R40 s18859_p1386_F13 n3_3734_p7563_R4 n3_8764_p4764_R21
	s4491_p11879_F25 s4551_p7037_R54
22	s2465_p15602_R33 n1_30659_p1188_R10 n4_9292_p13612_F31 n1_2755_p28892_R5 n3_2194_p18919_F15
	n1_2523_p15848_R30
23	s2831_p13560_R9 n3_6210_p692_F41 s4491_p11879_F35 n4_14229_p993_F29 n3_1984_p25602_F51
	n4_9878_p3974_R55 n3_4620_p11928_R3 n4_2827_p9540_F32 GMsc1824_128621_300
24	n4_25776_p490_R57 s1578_p18055_F24 n4_1774_p22559_F47 s679_p48524_R35 s1578_p18055_R11
	n1_8917_p2354_F25 n3_30659_p1188_R10 s1024_p11673_R28 n4_5805_p12685_R3 n4_8549_p4959_F63
	s1578_p14063_R23 n3_2523_p15848_R30 s1100_p21956_R32 s1547_p19112_R3 n4_5280_p10375_F45
	33L10SPBL3K_snp200r
25	n1_3734_p7563_R4 n1_6210_p692_F41 n3_3344_p4599_F22 s840_p31143_R3 n3_8917_p2354_F25
	JHI_14173DR n1_1984_p25602_F51 s1791_p21904_R44 s840_p39230_R19 ERubLR_SQ01_M20r
	s840_p31143_F45 n3_3878_p14285_R1 s1815_p9435_R59 n3_1476_p9064_R3 s1100_p8063_R27
	s878_p12079_R49 n1_2194_p18919_F15
26	s109335_p98_F30 n4_87004_p147_F55 n4_3846_p6183_F24 s4313_p9743_R46 n3_2755_p28892_R5
	RUB275a s3204_p16964_F10 n3_5561_p1843_R35 n3_3344_p10126_R47 s878_p12079_F18
	GMsc1824_128621_108 s451_p25846_R3 n1_3344_p4599_F22 s14395_p2931_R20 454C4050_HMG
	s1863_p13643_R58 JHIRi12783b_HMGR1 n1_3878_p14285_R1 33L10SPBL3K_snp182r n3_840_p34928_R33
	s9569_p1993_F6 n1_840_p34928_R33 s2820_p14616_F43
27	s2820_p14169_R16 s3880_p15515_F31 n1_3344_p10126_R47 s3382_p22659_R27 s1218_p33490_R23
	s78166_p151_R40 s5724_p9047_R37
28	n1_7302_p3280_R14 ERubLR_SQ01_M20 n3_3566_p4625_R56 n4_106802_p126_F14
	s566_p9755_F36 n3_7302_p3280_R14 n3_4423_p14377_F60 s843_p15065_F62 JHIRi_30994a_s186r
	n1_8965_p3814_R4
29	n1_3303_p11709_R46 n1_7909_p6824_F32 n4_7244_p7821_F24 n4_5388_p6351_R46 JHIRi_36325a_snp87
30	s272_p61276_F61
32	n3_7909_p6824_F32 n3_8177_p3260_R39 n3_3303_p11709_R46 n3_1011_p29982_R46 n3_8965_p3814_R4
33	s1761_p15087_F38 s758_p25728_R40 s8908_p5516_R46 s1761_p15087_F16 n1_51621_p338_F38
34	s832_p40126_R57 s247_p33804_R54 n4_2603_p9683_F22
35	s3258_p8588_R24 n4_6118_p2359_F32 s2603_p2249_R55 n1_8177_p3260_R39 n4_247_p10344_F23
36	n4_1324_p17078_F54 n4_11360_p6026_R58 s919_p8362_F20
37	s1240_p16963_F22 n3_51621_p338_F38
38	s1985_p16078_R33 n1_4423_p14377_F60
39	s2495_p241_F42
40	s4729_p8464_R30 s4729_p11705_F63
41	s4729_p8464_R39 s1302_p27855_R49
42	s39735_p266_R15 n4_13217_p2471_R59 454Con8186_PMEi n1_1011_p29982_R46 s8057_p2065_F36
	s5282_p25595_F41 s8057_p2065_R55
43	s608_p18443_F46 s3358_p5309_F56 s571_p48111_F27 s1011_p28214_F41 n4_7015_p7482_R41
	n4_1483_p29134_R14 s524_p29708_F37
44	n11_1657_p47676_F34 s5041_p11413_R19 s4468_p6531_R1
45	s10873_p2641_F10 s4745_p469_F26 n4_3475_p11300_F41 n3_1657_p47676_F34 RUB117b
	s335_p46718_F19
46	s9217_p10257_R8 s7292_p13146_R28 s7292_p11826_F39
47	s1011_p31013_F46 s8448_p6355_R30
48	s336_p9161_R58
49	Rub222e s608_p42360_F17 s5786_p4510_F61 RUB45c s608_p46787_R59
50	n3_6869_p7817_R14
51	n3_3209_p2389_F36 s524_p22539_R30 s117_p24297_R38 s3982_p308_F18 s667_p27736_F18
52	s1280_p29430_F27 s2051_p13770_R9
53	n1_3209_p2389_F36 s2247_p12625_R3 s2993_p21170_R2
54	s77_p61737_R40 s1092_p9869_R7 s2051_p13770_R19 s58841_p523_F16 s661_p25318_R37
	n1_859_p17446_F19 s661_p26374_F63 n3_497_p42363_R5
55	s8151_p8502_R17 s859_p25988_F6 ERubLR_SQ05.3_E02
56	s1365_p5989_R4 s735_p29161_R27 s735_p43384_R36
57	s696_p43451_R13 ERubLR_SQ13.1F09HMGS ERubLR_SQ13.4_H10PG n4_9823_p404_F6 s6886_p8846_R7
	s667_p21657_F64 s667_p21866_F56 s667_p21454_R51 s696_p40565_R31 s3418_p9472_F7
58	n1_10011_p6921_R22 s6886_p12719_F7 n1_497_p42363_R5 s1091_p17893_R53 s7872_p8875_R55
	s4296_p11476_R18 n1_4871_p14878_R35
59	n3_10011_p6921_R22 ERubLR_SQ19.1_A05 454CL4590c1_DXR n3_7757_p6948_F24 n3_859_p17446_F19
	RubnebAo9b s1586_p45665_F10 n1_7757_p6948_F24 n3_4871_p14878_R35 s661_p24838_R40
60	n1_6869_p7817_R14 s859_p26224_R25 s1405_p18960_F42 n1_32571_p1066_F22 s1405_p23098_R57
61	RiCellulase n3_32571_p1066_F22
62	n4_10176_p2911_F15 s2650_p21103_F49 n3_31624_p604_R9 s391_p52019_R52 n1_474_p78356_R56
63	s22606_p1645_R52
64	n4_745_p1747_F61
65	s743_p30341_R57 GMsc16725_11091_80 s10039_p165_R36
66	s745_p9481_R20 s745_p19242_F36 s99_p88968_F32 s745_p34108_F30 n3_31624_p604_R42
	s2889_p24647_R28 s5853_p10773_F26 s99_p135758_F20
67	n4_99_p64282_R33 s745_p35569_R56 s67_p83806_R19 s67_p63108_R55 n1_506_p53181_R4
	s67_p83806_F49
68	s67_p61764_R43 n1_381_p3284_R43 n4_755_p41754_R19
69	n3_381_p3284_R43 s3373_p19318_R2 s589_p30821_R28 n1_31624_p604_R42 n1_31624_p604_R9
	s1382_p24098_R51 s381_p35491_R20 n3_506_p53181_R4 s3373_p19318_R23 s381_p19407_R9
70	n3_474_p78356_R56 s1225_p15580_F10
71	s236_p83065_F43 s10_p124354_R50 RGsc2GRF_snp151 454c5803bHOT1 RGsc2SexLet_snp243
72	s374_p27676_F61 s1838_p21919_F22 s10_p389528_R50 s10_p299033_F46
	s12_p405880_R16 s10_p334829_F50 s10_p251729_F35 s10_p21337_F62 s12_p229804_R31
	s12_p300127_R29 s506_p28651_F13 s10_p275885_R5 s638_p35885_R30 s14405_p1382_R58
	s842_p17746_R27 s10_p359692_F12 s647_p46430_F31 s12_p272490_F34 s2888_p11983_R29
	s12_p326525_F28 s12_p454873_R30 s647_p7985_R48 s12_p488204_R27 s12_p47643_R28
	s838_p19224_F61 s10_p167917_R45 s12_p299465_R29 s374_p13454_R27 s10_p149674_F14
	s10_p165243_F46 s508_p16311_R51 s3494_p8069_R53 s12_p315238_F47 s12_p462752_F40
	s374_p38224_R53 s647_p45958_F31 s647_p45958_R34 s10_p339677_R28 s2063_p1325_R34
	s10_p53221_R48 s12_p68458_R37 s2888_p6457_R42 s236_p1626_R32 s236_p1626_F20
73	RGsc2BHLH_snp245 s261_p65753_F44 454C7840_CDF1ZnT RGsc2Cell5_snp82 s12_p397968_F15
74	s10_p56290_R53 s816_p24156_R19 s1932_p556_F62 n4_1767_p23944_F57 s1767_p8958_F18
75	s633_p30062_F20
76	n3_36018_p425_R28 s1235_p11903_F26 n4_13129_p138_F35 n1_2044_p13445_R12
77	s156_p34556_R26 ERubLR_SQ13.1_F06F3H s20209_p651_R34 n3_2044_p13445_R12 s156_p73813_F52
	s50059_p600_R54 s3390_p12252_R4 s2044_p13445_F43 s156_p39659_F57 s107_p112293_R11
	s1561_p16339_R12 ERubParaSQ005_K23 454c5803HOT1 s1986_p25996_F16 s562_p25127_F63
	s562_p44190_R47 n3_1593_p15423_R29 s688_p24569_F48 s1362_p12733_F44 s4751_p18220_F57
78	n1_36018_p425_R28 s1593_p15423_F56 s3390_p13510_R23 s156_p12214_R58 s156_p98835_F31
79	n3_1986_p26488_R22 s1637_p16517_R11 s7346_p4627_R36
80	s1986_p25996_F53
81	RiM019
83	n1_1986_p26488_R22
84	s562_p44863_R36 n1_1593_p15423_R29
85	s4751_p5333_R7

Fig. 1 (continued)

F

LG6

Position	Marker
0	RUB43a
2	454C6475_Arabino
3	s5412_p6684_R47 s26492_p501_F32 s4321_p5696_R54
4	s6149_p1292_F37 s781_p15166_F32 s1043_p25396_F59 s862_p13000_F8
	RUB1b s1931_p23773_F31 s5202_p10092_F12 s1992_p21230_R31 454CL3132C1_ISPE
5	n1_3752_p10538_R43 s1043_p25396_R31
	s3837_p10402_F10 s1043_p25396_F24 s1931_p23773_F51 s1931_p8787_F18 s1391_p12747_F13
6	s4040_p11597_R43 s6600_p4041_R15 s3241_p12352_F13 s3056_p10335_R10 s4321_p7236_R3
	s1224_p8816_R56 s3127_p17325_R54 s3127_p19578_F42 s1224_p8816_R58 s210_p22268_R8
	s534_p4071_R3
7	n3_3752_p10538_R43 s2529_p36484_R2 n1_34582_p737_F41 n4_8265_p3181_F32 n3_34582_p737_F41
8	s36_p119077_F42 s297_p69726_R10 s907_p6241_R28
9	n4_7706_p1824_F32
11	s666_p20259_R9 s4665_p10912_F43 s10127_p3936_R52 s122_p93074_F57
12	s406_p58134_R52 s406_p55971_F62 s4064_p9160_R30
13	s6158_p15611_R4 s545_p18618_F34 s3056_p10335_R45
14	s846_p37946_R19 s3127_p19578_F15 s21460_p612_F8
15	s628_p34344_R26
16	s8635_p2535_R20
18	s590_p17226_R12 s1877_p26854_R9
20	s194_p34353_F17 s187_p53138_R11 s754_p10366_F28 s194_p44868_R19
21	s187_p35159_F48
23	n3_5153_p5835_F10 s327_p23718_R16
24	s291_p4625_R43 s327_p53810_F15 s1119_p12847_F7 s1877_p26854_R29 s1119_p30702_F48
25	s192_p60780_F35 n4_4926_p4165_F53
26	n1_5153_p5835_F10 s180_p82607_F53 s1417_p17440_R17
27	s1055_p6647_F52
28	n4_6995_p6383_F10 n1_19938_p309_F61
30	n1_975_p16625_F13 s1486_p28139_R48 n3_975_p16625_F13
31	n4_326_p27133_F13 s326_p33608_R7
32	s326_p33608_F60 n3_19938_p309_F61
33	s326_p63968_R7
37	s1211_p34231_R40 n1_15984_p1339_R8
40	n3_5959_p5471_F10
41	s347_p27545_R9
42	s34540_p1067_F11
43	s12910_p1952_F59 n3_15984_p1339_R8
44	n1_5959_p5471_F10
46	s75795_p432_F40
47	454C2458ShikMee32
49	s75795_p432_R52
50	n3_4722_p10559_F14
51	n1_1396_p20001_F54 s434_p49279_F62
52	n1_4722_p10559_F14 s2375_p17738_R44
54	s1579_p14718_F19
55	s90_p32000_F19
56	n3_1396_p20001_F54 s90_p126391_R45
57	s90_p92747_F40
58	s141_p46366_F40
60	s8500_p3629_F36 s228_p27949_R7 s228_p47674_R53 s228_p27949_R31
61	s78_p109830_R12 s93_p86749_F8 s93_p38473_R36 s78_p94878_R33 s78_p170628_F47
63	s137_p23139_R59
64	s137_p4224_F9 s26_p268043_F56 s40_p308382_R8 s26_p323445_R14 s643_p40024_R3
	GMsc26_21683_snp259 s26_p390056_R43
65	s26_p226546_R47 s26_p245996_R32 s75_p30405_R48 s26_p80331_F37 s26_p54780_F43
	s40_p47457_R18 s40_p188760_F9 s40_p273960_R37 s26_p99338_F43
66	s40_p14815_F8 s411_p24277_F39
67	s732_p489_R16 s411_p56014_F42
68	s11_p461365_F17
69	s9_p172840_F19 s154_p57940_F23 s9_p328755_F58
70	s9_p220159_R4
	s9_p460010_F60 s27_p272599_R16 s27_p510423_R2 s9_p354834_F49 s9_p379558_R1
71	s41_p281245_R32
72	s252_p82330_R16 s27_p272599_R30 s49_p39208_R40 s41_p7066_F19
	s49_p34284_F34 s27_p43490_F37 s49_p22618_F62 s49_p132605_R33 s27_p59558_F16
73	s27_p227542_F60 JHIRi_38712a_snp167
	s49_p193292_R30 s49_p240489_R29 s23_p57404_F29 s23_p75516_R57 s23_p291249_F12
74	s49_p78590_F19 s822_p15045_R1 s23_p239682_R2 s109_p148468_R54 s49_p166051_R59
	s46_p23716_R22 s46_p89188_R36 s23_p346475_R20 s23_p51794_R15 s109_p1338_F32
	s18_p288705_R58 s7_p437825_R18 s7_p523119_R41
	s18_p300088_F50 s63_p78381_R36 s23_p353018_F64 s46_p43314_F56 s23_p179062_F25
	s23_p105714_R25 s18_p297797_R58 s231_p66829_R37 s23_p57404_F19 s121_p3038_F63
	s23_p38117_F34 s70_p146018_R24 s7_p369710_R1 s55_p24285_F35 s822_p14608_R25
	s23_p238221_R57 s7_p413316_R48 s62_p198202_F38 s63_p89937_F44 s46_p272389_R10
	s7_p66018_F27 s46_p142420_R20 s23_p247265_F57 s23_p76646_R34 s23_p247265_F54
75	s63_p190775_F13 s28_p197539_F55 ERubLR_SQ8.4_E06 s109_p56643_R45 s70_p13599_F49
	s55_p42325_R19 s23_p163644_F47 s23_p105714_F13 s28_p58847_F20 s18_p315510_R6
	s7_p145405_R30 s23_p265616_F21 s55_p140190_R40 s62_p147494_R51
	s18_p48651_R24 s34_p277903_F50 s28_p205442_R32 s18_p291231_F42 s28_p171922_R5
	s34_p23458_R21 s18_p320722_R59 s55_p225204_F45 s345_p78531_R47 s18_p102599_F55
76	s28_p205442_R51 s18_p250908_F42 s70_p146018_F64 GMsc18_13670_snp153 s70_p196724_F39
	s345_p45903_R55 s28_p275314_R21 s28_p320704_R31 s18_p315582_F43 s18_p71755_F19
	s70_p4027_F6 s48_p206695_F28 s120_p124813_R16 GMsc18_13610_snp101
	s48_p270775_R4 s34_p58728_F35 s188_p28001_F64 s120_p125593_F49 s32_p12523_F45
	s188_p50820_R12 s75_p296048_F31 s48_p57357_R37 s17_p169772_F41 102B35N18fB(2)
77	s47_p53388_F38 s75_p199538_F44 s284_p61400_R48 454SUSKC0003 s284_p60550_F61
	s75_p153071_F46 s17_p463956_R41 s32_p83680_F34 s47_p258368_R34 s48_p57805_R50
	s32_p184662_F48 s32_p300575_R22 s47_p200389_F59 s17_p138977_R25 s120_p32525_F15
	s17_p21368_F61 s32_p234670_F19 s7228_p6894_R58 s120_p81161_R7 s48_p60633_R55
78	s32_p150041_R48 s120_p133140_R33 s32_p57707_F22 s17_p368062_R1 s17_p314965_R25
	s48_p62495_F59 s32_p23624_F23 s47_p18631_F28 s75_p253801_R6 s17_p169772_R59
	s53_p220354_F37 s17_p248581_R53 s17_p457428_F47
79	GMsc17_12614_snp86 s53_p85073_R51 s246_p22918_R40 s17_p401530_F16 s53_p215100_R9
	s17_p255505_R25 s53_p125869_F21 102B35N18fB(1) Ri11G23_Glyhydro
80	Ri11G23_cont5_SSR ERubLR_SQ11_3_A11Aux
82	454C1027_lycCyc 454c2137bHS
83	38J3SSR2-1A12 454a2137HS 38J3SSR2-1G02 bes_Ri38J3R
	Ri38J34038clava bes_Ri38J3F RUB118b2 Ri38J34034-knopecu Ri38J35275stre
84	Ri38J35250zf
85	Ri38J34046_dicer Ri38J34070_sugart Leaf97b
86	Ri38J3_5262_NAP Ri38J35271MCM5 Ri38J35269sugar
87	Bac38J3-3-1H04r
89	Ri38J35356oxid

QTL bars (left): SanguiinH10_2010, EllAcetylPent, MeEllPent, EllagicPent2, EllagicPent1, KGlu

Fig. 1 (continued)

Fig. 2 **A** Genotype means from the over-years analysis (on $\log_{10}$ transformed data) for metabolites at the QTLs at the bottom of LG1. The genotype bd results in lower levels of these metabolites than the other genotypes. Each genotype mean is standardised by subtracting the overall mean and then dividing by $\sqrt{}$(QTL residual mean square) for comparability. **B** Genotype means from the over-years analysis

(on $\log_{10}$ transformed data) for metabolites at the QTL at the bottom of LG1. The ac and ad genotypes have lower levels of the components, an effect mainly due the Latham parent. Each genotype mean is standardised by subtracting the overall mean and then dividing by $\sqrt{}$(QTL residual mean square) for comparability
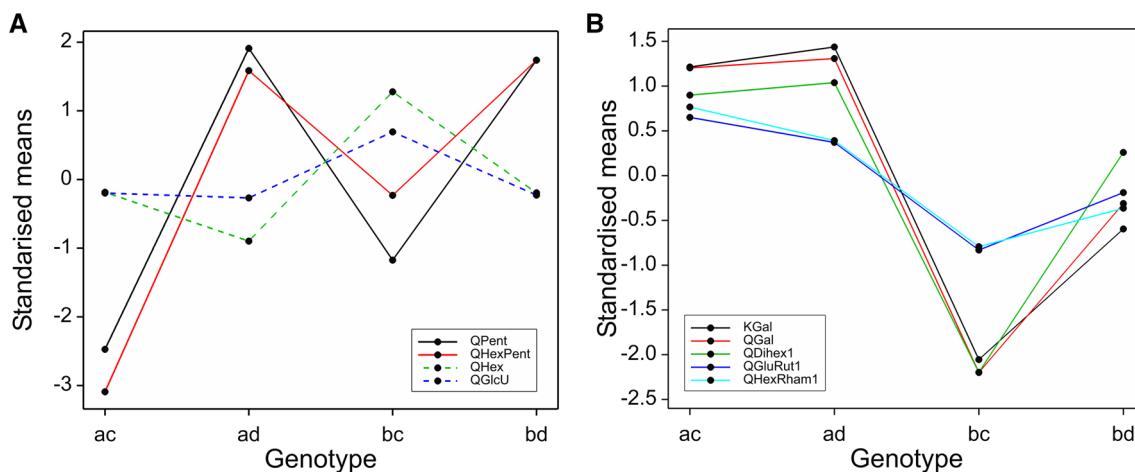
Fig. 3 **A** Genotype means from the over-years analysis (on $\log_{10}$ transformed data) for the QTL at 25–33 cM on LG4 (and the metabolites where the QTL effects are different from Fig. **B**). Each genotype mean is standardised by subtracting the overall mean and then dividing by $\sqrt{}$(QTL residual mean square) for comparability. **B** Genotype

means from the over-years analysis (on $\log_{10}$ transformed data) for the QTL at 25–33 cM on LG4. Each genotype mean is standardised by subtracting the overall mean and then dividing by $\sqrt{}$(QTL residual mean square) for comparability. The levels of the metabolites were lowest in genotype bc

group on the A-ring. Indeed, a similar strong correlation (r = 0.92) was noted between KGlcU and QGlcU (Fig. S2c). This could suggest metabolic interconversion of kaempferol to quercetin glycosides.

Strong QTLs were also detected for QGlu and QDihex, but these were well separated from the above, with peak LODs at 59 cM and 78 cM respectively. Both had highest levels for genotype ac, intermediate for ad and bc and lowest for bd.

Some QTLs for anthocyanins were also detected on LG4, with good consistency over years, but the QTL positions for each anthocyanin varied more. A major QTL for CySam mapped to 30 cM with LOD 18.3, and R^2 of 35.7%. This was mainly an effect of the Latham parent. Another major QTL for CyGlu was mapped to 47 cM, with LOD 16.2, and R^2 of 31.2%, and a strong QTL for PelGlu and a minor QTL for Total Anthocyanin mapped to 58 cM. All these showed a similar pattern of genotype means, with ac having the highest level, ad and bc having intermediate levels and bd having
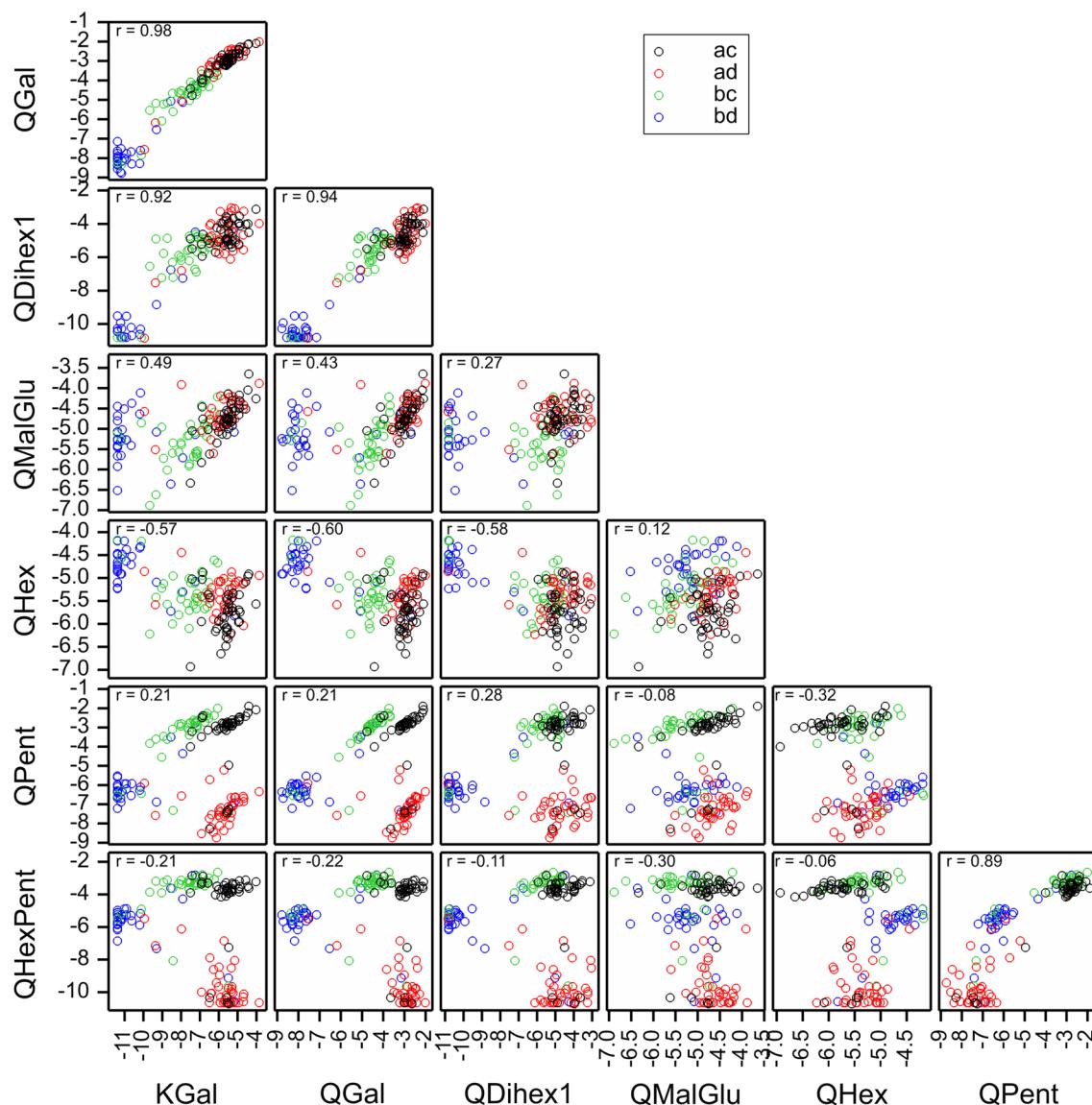
Fig. 4 Scatterplot matrix of flavonols with major QTLs on LG4, labelled by the genotypes at 30 cM (based on SNPs s340_p9738_F13 from Latham and n1_9428_p6492_F60 from Glen Moy). The colour coding reflects the alleles inherited from the parents, ac, ad, bc & bd

the lowest level. PelSam had a strong QTL at 89 cM and CySoph had a minor QTL at 88 cM, with similar patterns to the previous three anthocyanins.

An interesting pattern was noted in the scatterplots for the data in 2010 (Fig. S2a, top left box) for CySoph and CySam. CySam is a disaccharide consisting of xylose and glucose attached to cyanidin whereas CySoph is cyanidin with an attached disaccharide of glucose. These showed two distinct parallel clusters related to the genotypes. The marker that corresponds best to the separation is JHIRi49055_GIG at 26 cM on LG4, which segregates in Latham only: the clusters correspond closely to the offspring receiving the a and b alleles from Latham. Similar patterns were observed in the other years (not shown).

There were some genes underlying the QTL region at 25–33 cM on LG4 that have been implicated in the control of polyphenol biosynthesis. There was a homologue of a cytochrome P450 that has been implicated in anthocyanin accumulation in purple Brassicas (Zhu et al, 2021). Also, there was a gene with homology to a dihydroflavonol 4-reductase (DFR)-like protein in *Arabidopsis* (Yuan et al., 2007) (also 85% homology with the *R. chinensis* gene) which could be involved in the final steps of anthocyanidin biosynthesis (see scheme 1A).

However, most interestingly there were homologues of genes encoding uridine diphosphate glycosyltransferases (UDPGTs) which have been implicated in the glycosylation of flavonoids, and in particular anthocyanins. All the

Scheme 1 A Pathway for anthocyanin and flavonol biosynthesis. *Insert of possible biosynthetic pathway to ellagitannins and ellagic acid.* Enzymes are abbreviated as follows: ADH, arogenate dehydrogenase; ADT, arogenate dehydratase; ANS, anthocyanin synthase; CHI, chalcone isomerase; CHS, chalcone synthase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumarate ligase; CM, chorismate mutase; DFR, dihydroflavonol reductase; F3H, flavonone-3-hydroxylase; F5H, ferulate 5-hydroxylase; F3′H, flavonoid-3′-hydroxylase; F3′5′H, flavonoid-3′,5′-hydroxylase; FLS, flavonol synthase; PAL, phenylalanine ammonia lyase; PDH, prephenate dehydratase; SK = Shikimate kinase and CS = chorismate synthase. For clarity, only enzymes that produce the flavonoid aglycones are shown, not the glycosyltransferases and acyl transferases that decorate the aglycones. Methylated anthocyanidins (i.e. peonidin, petunidin and malvidin) and flavonols (isorhamnetin) are also not shown as they are not present in raspberry. **B** Structures of ellagitannins and ellagic acid found in raspberry

B

Lambertianin C

Sanguiin H6

Sanguiin H10

Ellagic acid

Sanguiin H2
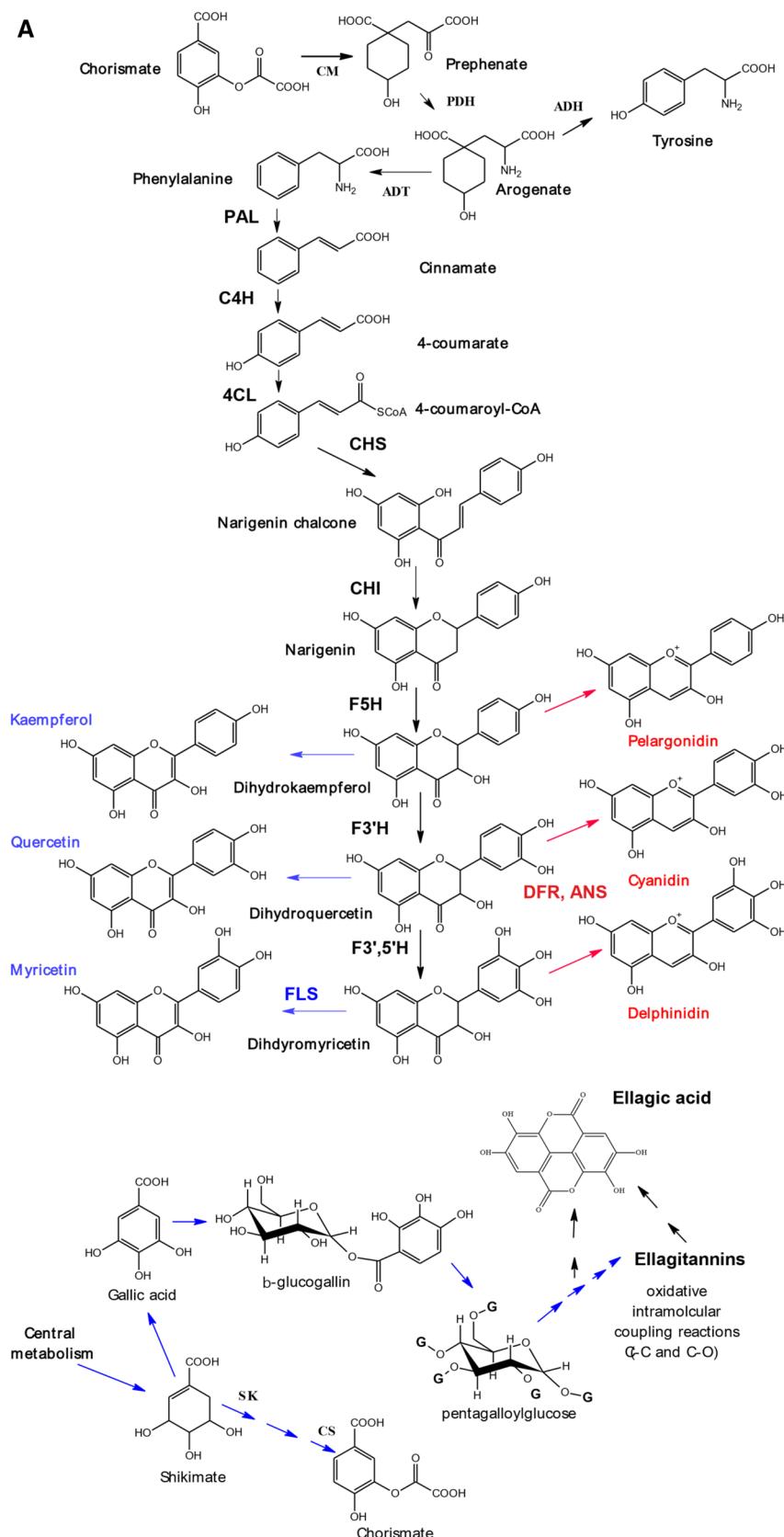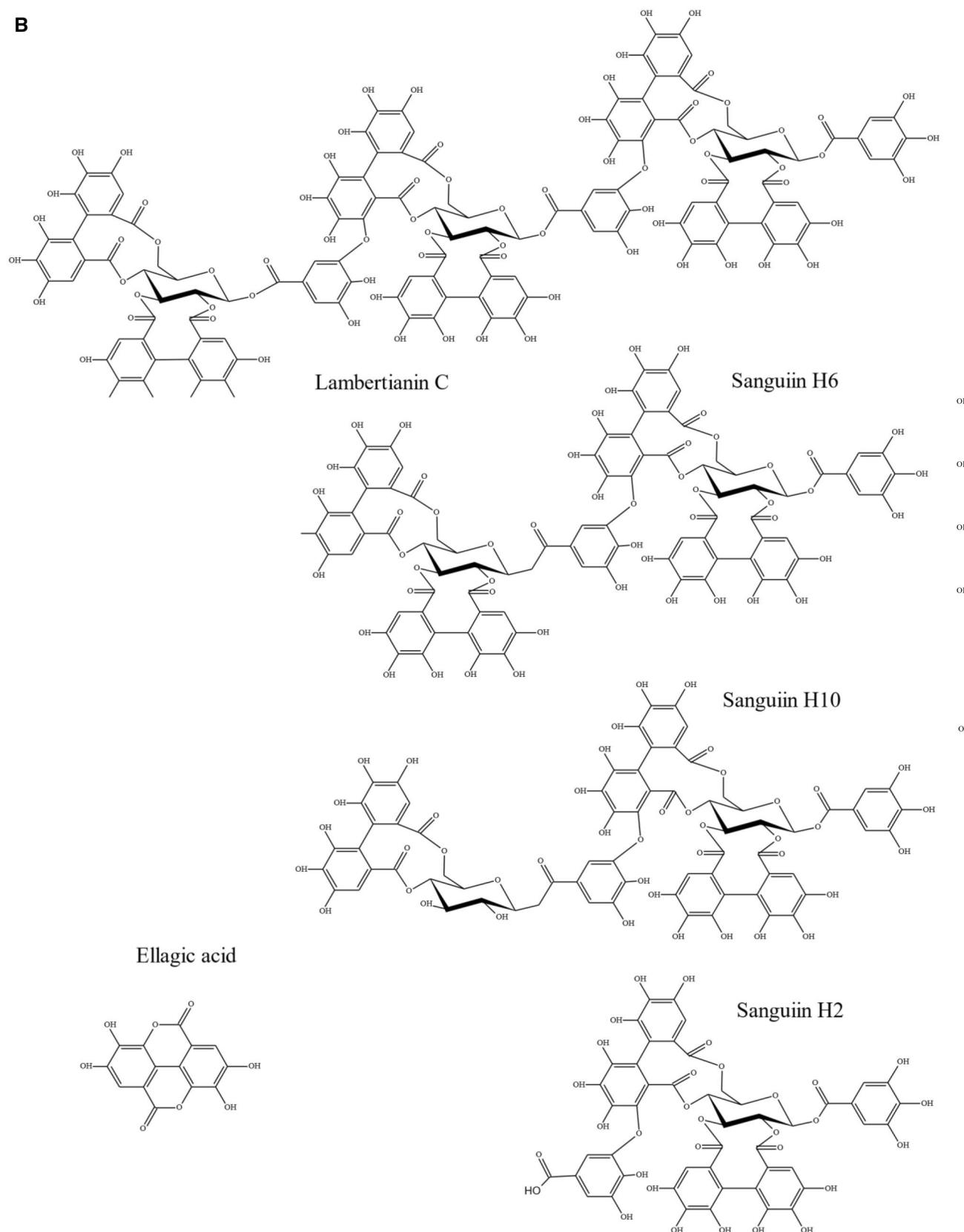
Scheme 1 (continued)

homologies were > 80% with the genes from *R. chinensis* or *F. vesca*. Indeed, one hit was noted with UDPGT 74E2 (AT1G05680.1) which has been described as an anthocyanin 3-*O*-glucoside: 2″-*O*-xylosyl-transferase that converts cyanidin 3-*O*-glucoside to cyanidin 3-*O*-xylosyl-glucoside in *Arabidopsis* [AT5G54060(UF3GT) (arabidopsis.org)]. This is interesting as CySam (or cyanidin 3-*O*-xylosyl-glucoside) has a major QTL in this region. Also, the transfer of xylose would also be relevant to the formation of the flavonol pentoses (QPent and QHexPent), which also have major QTLs here. It would be intriguing if this UDPGT homologue could also be involved with the transferal of hexosyl units to form the other anthocyanidin and flavonol hexose components associated with this QTL region.

In 2010 at 77 cM on LG4, there was a strong QTL for the ellagitannins Lambertianin C and Sanguiin H10 and minor QTLs for Sanguiin H6, and Sanguiin H2. These had minor QTLs in a similar region in 2011 (77–89 cM) and there was a minor QTL for Lambertianin C at 84 cM in 2015 only. However, no significant QTLs were detected here for the over-years means. These metabolites are all positively correlated from 0.52 up to a maximum of 0.86 for the correlation between Lambertianin C and Sanguiin H6 in 2010 (Fig. S2c). Correlations for other years were similar (results not shown). Ellagitannins (ETs) originate from gallic acid, itself formed from the central metabolite, shikimate (Scheme 1A). The production of galloyl glucose is the first committed step in the formation of Ets (e.g. Schulenburg et al., 2016) then further galloyl groups are added to form pentagalloyl glucose. Ets are then produced by defined intramolecular oxidative coupling reactions on pentagalloyl glucose with a C–C coupling of galloyl groups producing the hexahydroxydiphenic (HHDP) acid unit characteristic of Ets (Yamada et al. 2017). Indeed, pentagalloyl glucose with two HHDP groups is also called casuarictin. The high correlation between Sanguiin H6 and Lambertianin C (R = 0.86) may reflect a precursor relationship as Sanguiin H6 can be described as a dimer of casuarictin linked through the gallic acid residue and a HHDP acid unit whereas Lambertianin C has been described as a trimer of casuarictin (Scheme 1B). The slightly less strong correlation (R = 0.76) between Sanguiin H2 and Sanguiin H6 also suggests a precursor relationship and indeed Sanguiin H6 differs from H2 by the addition of an HHDP-glucose unit (Scheme 1B). The less strong correlation (R = 0.56) between Sanguiin H6 and Sanguiin H10 may also indicate a less immediate precursor relationship as they differ by a HHDP unit.

Ellagic acid arises when HHDP groups released from Ets undergo spontaneous lactonization (Schulenburg et al., 2016). The correlation between the ET Sanguiin H2 and ellagic acid was relatively high (R = 0.72), which may indicate some metabolic connection. The correlations were less strong for the larger Ets (Sanguiin H10 (R = 0.40), Sanguiin H6 (R = 0.40) and Lambertianin C (R = 0.21) and ellagic acid, which implies a greater metabolic distance between the larger Ets and ellagic acid.

The correlations between ellagic acid and some of the ellagic acid derivatives was also interesting. There was a strong correlation between EllagicPent1 and MeEll pent (R = 0.82), which suggests a precursor relationship between these compounds. However, the correlation between EllagicPent2 and MeEll pent was negative (R = -0.29). The consensus is that the first eluting ellagic pentose is likely to be a xyloside and the latter eluting one an arabinoside (Gasperotti et al., 2010; Mullen et al., 2003). Therefore, this correlation may reflect that the pentose attached to EllagicPent1 is most probably xylose and arabinose in EllagicPent2. Indeed, there was a negative correlation between Ellagic pent 1 and 2 (R = − 0.6) with a bimodal pattern (Fig. S2C).

3.3.3 LG6

LG6 has few markers from Glen Moy and so most QTLs are due to differences between the alleles inherited from Latham. LG6 shows a very significant area for ellagic acid derivatives close to 80 cM in each year. There was a major QTL for EllagicPent2 (arabinose) for each year (giving it a bimodal distribution), with LOD 55 for the over-years average and R^2 of 74%. This is in the same region as a QTL for total phenol content identified in previous work (Dobson et al., 2012). EllagicPent1 (xylose) also had a major QTL here, with LOD 13.4, and R^2 of 27%, but with the opposite effect to EllagicPent2. MeEllPent and EllAcetylPent also have strong QTLs with the same direction as EllagicPent1 (Xyl) in this region, and the flavonol Kglu has a major QTL (LOD 11.9, R^2 24%). Figure 5A shows a scatter plot of the over-year genotype means for the ellagic acid derivatives, labelled by marker Ri11G23_cont5_SSR at 80 cM, displaying the separation of the distribution of EllagicPent2 (Ara) between the two genotypes. The bimodal distribution for EllagicPent2 (Ara) complicates correlations; for example, the overall correlation between EllagicPent1 (Xyl) and EllagicPent2 (Ara) in 2010 was -0.60, but within each of the clusters corresponding to the genotypes of Ri11G23_cont5_SSR the correlation is positive. There was also a strong QTL for the ellagitannin Sanguiin H10 in 2010 only, with the peak LOD of 5.9 at 71 cM.

There are some genes underlying the QTL region (75–80 cM) on LG6 that could be implicated in the control of transfer of sugars to flavonoids. Once again there are homologues (> 80% homology) with *Arabidopsis* genes that encode uridine diphosphate glycosyltransferases (UDPGTs), one of which has been described as specific for the transfer of glucose to quercetin to form
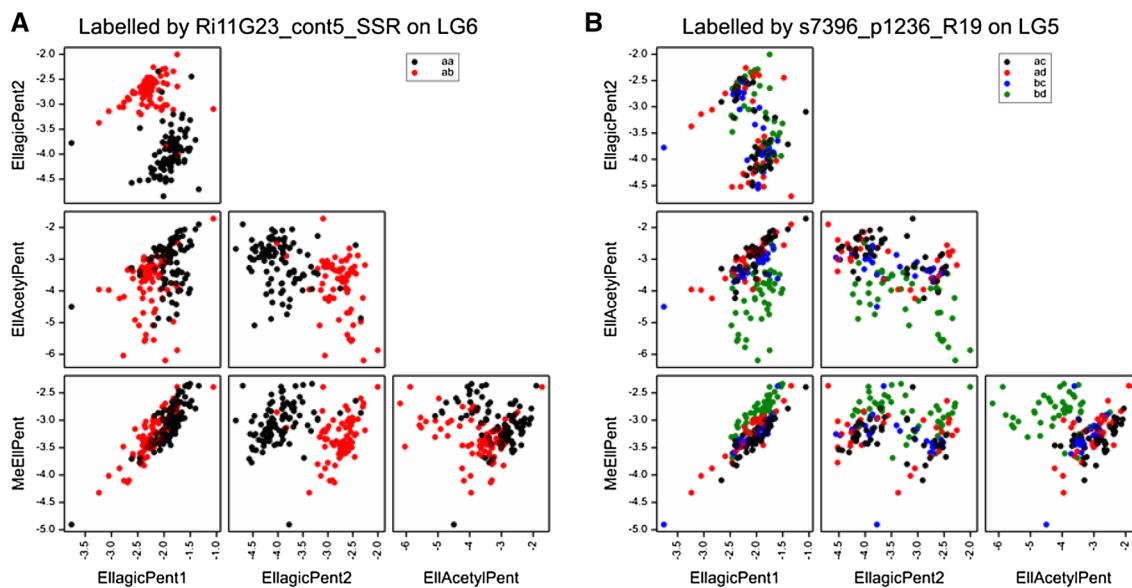
Fig. 5 Scatterplots of the ellagic acid derivatives, labelled by markers close to the major QTLs on **A** LG6 and **B** LG5. The colour coding reflects the alleles inherited from the parents, ac, ad, bc & bd

quercetin 7-O-glucoside (AT1G05675.1–rabidopsis.org) but another is described as specific for transfer to benzoic acids, including salicylates (AT2G43820.1). However, homologues in other species have not been described with these specificities so it is possible that the raspberry gene encodes a transferase that could be involved in the transfer of pentoses (arabinose or xylose) to ellagic acid.

3.3.4 LG5

LG5 also has major QTLs in the over-years analysis for ellagic acid derivatives EllAcetylPent (LOD 27.2, R^2 47%) and MeEllPent (LOD 10.2, R^2 21%) at 8–16 cM. For EllAcetylPent, offspring with genotype ac had the highest level, those with genotype ad and bc had intermediate levels and bd had the lowest level. The pattern was reversed for MeEllPent. Figure 5B shows a scatter plot of the ellagic acid derivatives, labelled by marker s7396_p1236_R19 at 17 cM on LG5. Neither EllAcetylPent nor MeEllPent have bimodal distributions, but their scatterplot shows a separation of the bd genotype (shown in green) from the other three genotypes. Once again, these metabolites have an overall negative correlation (Fig. S2C), but a positive correlation within the QTL genotype clusters. KGlu has a strong QTL nearby, at 19 cM, with a similar pattern to EllAcetylPent.

There were also strong QTLs in the over-years analysis for PelSam and PelGlu mapping to 75–76 cM on LG5, with similar effects due to the allele from Latham.

3.3.5 LG2

No major QTLs for metabolites were detected on LG2. Strong QTLs (LODs 5.7–7.3) were detected in the over-years analysis in the region 12–19 cM, for the anthocyanins CySoph, $CyC_{28}H_{30}O_{17}$ and total anthocyanins. All showed a similar pattern of effects, with ac and ad having the highest levels, bc intermediate and bd lowest. A minor QTL with a similar pattern was detected nearby, at 5 cM, for QDihex2.

A second region was found at the other end of LG2 for KGlcU and QGlcU, centred at 105 cM. This matches with the position of a QTL previously identified for total phenol content (Dobson et al., 2012). However, that study was based on fewer progeny and the significance is lower. For these flavonols, genotype bc had the highest level, genotypes ac and bd were intermediate and genotype ad was lowest.

3.3.6 LG3

Unlike the other linkage groups, the QTLs on LG3 were not consistent from year to year. No QTLs for were detected here in 2015, few in 2011 but more in 2010. Therefore, results for this linkage group are given from the analysis of data from each year separately in Fig. 1 and in the discussion below.

In 2010, a strong QTL was detected for total anthocyanins at 101 cM, together with a strong QTL for CyGlu at 111 cM. This matches with a QTL for total anthocyanins noted in previous studies (Dobson et al., 2012; McCallum et al., 2010). A strong QTL for QGlu and minor QTLs for QDihex2, KGlu and Lambertianin C were also detected in the region 99–111 cM. All were predominantly Latham

effects, and all had effects in the same direction, apart from Lambertianin C, which showed the opposite effect. There was also a minor QTL detected for PelGlu at 76 cM, where the significant effect was from Glen Moy. Finally, there was a minor QTL for KGlcU at 8 cM (a significant effect of Latham). In 2011, a strong QTL was detected for Sanguiin H10 at 53 cM, and minor QTLs were detected for MeEllagicPent and EllagicPent1 (presumed xyloside) at 61 cM, all with effects from both parents. Interestingly, the QTL for Sanguiin H10 overlaps a QTL for total phenol content noted previously (Dobson et al., 2012), albeit from a smaller progeny set. There was also a minor QTL at KGlu at 98 cM, an effect of Latham.

4 Conclusions

Major and consistent QTLs were established for many of the polyphenol components of red raspberry, which have been previously implicated in health beneficial effects. This indicates that there is considerable genetic control of the levels of these components and provides confidence that markers could be deployed in breeding programmes to accelerate breeding of cultivars with defined polyphenol profiles. Many of these QTLs have underlying candidate genes that could influence accumulation of specific polyphenols. For example, the QTL region at 25–33 cM on LG4 had major QTLs for flavonoid components (QHexPent, QPent and CySam) which contained pentose sugars and was underlaid by a homologue of a UDP-glycosyltransferase implicated in the transfer of xylosyl groups to produce CySam from CyGlu in Arabidopsis [AT5G54060(UF3GT) (arabidopsis.org)]. This region could be examined for markers which could be used in accelerated breeding of new raspberry varieties with elevated levels of these specific anthocyanins and flavonols. Future work could examine the expression of some of the candidate genes indicated across the ripening stages of the parents (e.g., Scolari et al., 2021) to confirm if their expression matched the known accumulation of the specific polyphenols.

There was good correlation across years between various ellagic acid derivatives which suggested metabolic inter-connectivity. Similar inter-connectivity was noted for flavonol components, some following expected biosynthetic pathways but others less easily explainable. The correlations in levels of ellagitannins could reflect metabolic patterns and inter-relationships in the synthesis of these large polyphenol components. This is interesting because the biosynthetic pathways that lead to the formation of large ellagitannin structures are still not well defined despite pioneering work which suggested controlled oxidation of pentagalloyl glucose groups by specific laccases (see Niemetz & Gross, 2005). Closer examination of the genes that underly the

strong QTLs for the ellagitannins, Sanguiin H6 and Lambertianin C, on LG3 at 124 cM may reveal candidate genes involved in the formation of these abundant and biologically active components. Once again, backing up with candidate gene expression studies could be interesting as ellagitannins are known to be accumulated early in the fruit formation (Beekwilder et al., 2005).

Declarations

References

Ahn, S. Y., Kim, S. A., & Yun, H. K. (2019). Differentially expressed genes during berry ripening in de novo RNA assembly of *Vitis flexuosa* fruits. *Horticulture, Environment, and Biotechnology, 60*, 531–553. https://doi.org/10.1007/s13580-019-00148-2

Anttonen, M. J., & Karjalainen, R. O. (2005). Environmental and genetic variation of phenolic compounds in red raspberry. *Journal of Food Composition and Analysis, 18*, 759–769. https://doi.org/10.1016/j.jfca.2004.11.003

Arnold, H., Dobson, G., Foito, A., Austin, C., Sungurtas, J., Allwood, J. W., Stewart, D., & McDougall, G. J. (2022). Assessing available phytochemicals from commercial blackcurrant and raspberry

pomaces. *Journal of Berry Research, 12*, 415–431. https://doi.org/10.3233/JBR-220017

Beekwilder, J., Jonker, H., Meesters, P., Hall, R. D., van der Meer, I. M., & Ric de Vos, C. H. (2005). Antioxidants in raspberry: On-line analysis links antioxidant activity to a diversity of individual metabolites. *Journal of Agriculture and Food Chemistry, 53*, 3313–3320. https://doi.org/10.1021/jf047880b

Bogs, J., Downey, M. O., Harvey, J. S., Ashton, A. R., Tanner, G. J., & Robinson, S. P. (2005). Proanthocyanidin synthesis and expression of genes encoding Leucoanthocyanidin reductase and Anthocyanidin reductase in developing grape berries and grapevine leaves. *Plant Physiology, 139*, 652–663. https://doi.org/10.1104/pp.105.064238

Bradish, C. M., Perkins-Veazie, P., Fernandez, G. E., Xie, G., & Jia, W. (2012). Comparison of flavonoid composition of red raspberries (*Rubus idaeus* L.) grown in the Southern United States. *Journal of Agriculture and Food Chemistry, 60*, 5779–5786. https://doi.org/10.1021/jf203474e

Carvalho, E., Franceschi, P., Feller, A., Palmieri, L., Wehrens, R., & Martens, S. (2013). A targeted metabolomics approach to understand differences in flavonoid biosynthesis in red and yellow raspberries. *Plant Physiology and Biochemistry, 72*, 79–86. https://doi.org/10.1016/j.plaphy.2013.04.001

Connor, A. M., McGhie, T. K., Stephens, M. J., Hall, H. K., & Alspach, P. A. (2005). Variation and heritability estimates of anthocyanins and their relationship to antioxidant activity in a red raspberry factorial mating design. *Journal of the American Society for Horticultural Science, 130*, 534–542. https://doi.org/10.21273/JASHS.130.4.534

Da Silva, C., Zamperin, G., Ferrarini, A., Minio, A., Dal Molin, A., Venturini, L., Buson, G., Tononi, P., Avanzato, C., Zago, E., Boido, E., Dellacassa, E., Gaggero, C., Pezzotti, M., Carrau, F., & Delledonne, M. (2013). The high polyphenol content of grapevine cultivar Tannat berries is conferred primarily by genes that are not shared with the reference genome. *The Plant Cell, 25*(4777), 4788. https://doi.org/10.1105/tpc.113.118810

Deighton, N., Brennan, R., Finn, C., & Davies, H. V. (2000). Antioxidant properties of domesticated and wild *Rubus* species. *Journal of the Science of Food and Agriculture, 80*, 1307–1313. https://doi.org/10.1002/1097-0010(200007)80:9%3c1307

Department for Environment, Food & Rural Affairs (2022), Agriculture in the United Kingdom data sets https://www.gov.uk/government/statistical-data-sets/agriculture-in-the-united-kingdom

Dobson, G., McDougall, G. J. & Stewart, D. (2010) Analysis of polyphenolics in raspberry, blackberry, blueberry, black currant and strawberry by high pressure liquid chromatography-photodiode array-mass spectrometry (HPLC-PDA-MS/MS). https://www.euberry.univpm.it/sites/www.euberry.univpm.it/files/euberry/documenti/JHI%20SOP%20compilation%2020130412.pdf

Dobson, P., Graham, J., Stewart, D., Brennan, R., Hackett, C. A., & McDougall, G. J. (2012). Over-seasons analysis of quantitative trait loci affecting phenolic content and antioxidant capacity in raspberry. *Journal of Agriculture and Food Chemistry, 60*, 5360–5366. https://doi.org/10.1021/jf3005178

Gasperotti, M., Masuero, D., Vrhovsek, U., Guella, G., & Mattivi, F. (2010). Profiling and accurate quantification of *Rubus ellagitannins* and ellagic acid conjugates using direct UPLC-Q-TOF HDMS and HPLC-DAD analysis. *Journal of Agriculture and Food Chemistry, 58*, 4602–4616. https://doi.org/10.1021/jf904543w

Graham, J., Hackett, C. A., Smith, K., Woodhead, M., Hein, I., & McCallum, S. (2009). Mapping QTLs for developmental traits in raspberry from bud break to ripe fruit. *Theoretical and Applied Genetics, 118*, 1143–1155. https://doi.org/10.1007/s00122-009-0969-6

Graham, J., Smith, K., MacKenzie, K., Jorgenson, L., Hackett, C., & Powell, W. (2004). The construction of a genetic linkage map of red raspberry (*Rubus idaeus* subsp. idaeus) based on AFLPs, genomic-SSR and EST-SSR markers. *Theoretical and Applied Genetics, 109*, 740–749. https://doi.org/10.1007/s00122-004-1687-8

Graham, J., Smith, K., McCallum, S., Hedley, P. E., Cullen, D. W., Dolan, A., Milne, L., McNicol, J. W., & Hackett, C. A. (2015). Towards an understanding of the control of 'crumbly' fruit in red raspberry. *Springerplus, 4*, 223. https://doi.org/10.1186/s40064-015-1010-y

Hackett, C. A., Milne, L., Smith, K., Hedley, P., Morris, J., Simpson, C. G., Preedy, K., & 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. https://doi.org/10.1186/s12863-018-0666-z

Hancock, R. D., Petridis, A., & McDougall, G. J. (2018). Raspberry fruit chemistry in relation to fruit quality and human nutrition. In J. Graham & R. Brennan (Eds.), *Raspberry.* Cham: Springer. https://doi.org/10.1007/978-3-319-99031-6_7

He, M., Tian, H., Luo, X., Qi, X., & Chen, X. (2015). Molecular progress in research on fruit astringency. *Molecules, 20*, 1434–1451. https://doi.org/10.3390/molecules20011434

Jennings, S. N. (2018). Advances in Rubus breeding. In J. Graham & R. Brennan (Eds.), *Raspberry.* Cham: Springer. https://doi.org/10.1007/978-3-319-99031-6_2

Jennings, S. N., Graham, J., Ferguson, L., & Young, V. (2016). New developments in raspberry breeding in Scotland. *Acta Horticulture, 1133*, 23–28. https://doi.org/10.17660/ActaHortic.2016.1133.4

Kassim, A., Poette, J., Paterson, A., Zait, D., McCallum, S., Woodhead, M., et al. (2009). Environmental and seasonal influences on red raspberry anthocyanin antioxidant contents and identification of quantitative trait loci (QTL). *Molecular Nutrition & Food Research, 53*, 625–634. https://doi.org/10.1002/mnfr.200800174

Li, J., & Ji, L. (2005). Adjusting multiple testing in multi-locus analyses using the eigenvalues of a correlation matrix. *Heredity, 95*, 221–227. https://doi.org/10.1038/sj.hdy.6800717

Mazzoni, L., Perez-Lopez, P., Giampieri, F., Alvarez-Suarez, J. M., Gasparini, M., Forbes-Hernandez, T. Y., et al. (2016). The genetic aspects of berries, from field to health. *Journal of the Science of Food and Agriculture, 96*, 365–371. https://doi.org/10.1002/jsfa.7216

McCallum, S., Simpson, C., & Graham, J. (2018). QTL Mapping and marker assisted breeding in Rubus spp. In J. Graham & R. Brennan (Eds.), *Raspberry.* Cham: Springer. https://doi.org/10.1007/978-3-319-99031-6_8

McCallum, S., Woodhead, M., Hackett, C. A., Kassim, A., Paterson, A., & Graham, J. (2010). Genetic and environmental effects influencing fruit colour and QTL analysis in raspberry. *Theoretical and Applied Genetics, 121*, 611–627. https://doi.org/10.1007/s00122-010-1334-5

McDougall, G. J., Conner, S., Pereira-Caro, G., Gonzalez-Barrio, R., Brown, E. M., Verrall, S., et al. (2014). Tracking (poly)phenol components from raspberries in ileal fluid. *Journal of Agriculture and Food Chemistry, 62*, 7631–7641. https://doi.org/10.1021/jf502259j

McDougall, G. J., Dobson, P., Smith, P., Blake, A., & Stewart, D. (2005). Assessing potential bioavailability of raspberry anthocyanins using an in vitro digestion system. *Journal of Agriculture and Food Chemistry, 53*, 5896–5904. https://doi.org/10.1021/jf050131p

Mullen, W., Lean, M. E., & Crozier, A. (2002). Rapid characterization of anthocyanins in red raspberry fruit by high-performance liquid chromatography coupled to single quadrupole mass spectrometry.

Journal of Chromatography A, 966, 63–70. https://doi.org/10.1016/s0021-9673(02)00699-4

Mullen, W., Yokota, T., Lean, M. E., & Crozier, A. (2003). Analysis of ellagitannins and conjugates of ellagic acid and quercetin in raspberry fruits by LC-MS[n]. *Phytochemistry, 64*, 617–624. https://doi.org/10.1016/S0031-9422(03)00281-4

Niemetz, R., & Gross, G. G. (2005). Enzymology of gallotannin and ellagitannin biosynthesis. *Phytochemistry, 66*, 2001–2011. https://doi.org/10.1016/j.phytochem.2005.01.009

Paterson, A., Kassim, A., McCallum, S., Woodhead, M., Smith, K., Zait, D., & Graham, J. (2013). Environmental and seasonal influences on red raspberry flavour volatiles and identification of quantitative trait loci (QTL) and candidate genes. *Theoretical & Applied Genetics, 126*, 33–48. https://doi.org/10.1007/s00122-012-1957-9

Plunkett, B. J., Espley, R. V., Dare, A. P., Warren, B. A. W., Grierson, E. R. P., Cordiner, S., et al. (2018). MYBA from blueberry (Vaccinium section Cyanococcus) is a subgroup 6 type R2R3MYB transcription factor that activates anthocyanin production. *Frontiers in Plant Science, 11*, 1300. https://doi.org/10.3389/fpls.2018.01300

Rajauria, G., & Tiwari, B. K. (2018). *Fruit juices, extraction, composition, quality, and analysis*. London, UK: Academic Press.

Rao, A. V., & Snyder, D. M. (2010). Raspberries and human health: A review. *Journal of Agriculture and Food Chemistry, 58*, 3871–3883. https://doi.org/10.1021/jf903484g

Remberg, S. F., Sønsteby, A., Aaby, K., & Heide, O. M. (2010). Influence of post flowering temperature on fruit size and chemical composition of glen ample raspberry (*Rubus idaeus* L.). *Journal of Agriculture and Food Chemistry, 58*, 9120–9128. https://doi.org/10.1021/jf101736q

Schulenburg, K., Feller, A., Hoffmann, T., Schecker, J. H., Martens, S., & Schwab, W. (2016). Formation of β-glucogallin, the precursor of ellagic acid in strawberry and raspberry. *Journal of Experimental Botany, 67*, 2299–2308. https://doi.org/10.1093/jxb/erw036

Scolari, L. M., Hancock, R. D., Hedley, P. E., Morris, J., Smith, K., & Graham, J. (2021). Combining QTL mapping and gene expression analysis to elucidate the genetic control of 'crumbly' fruit in red raspberry (*Rubus idaeus* L.). *Agronomy, 11*, 794–799. https://doi.org/10.3390/agronomy11040794

Sharma, D., Tiwari, M., Pandey, A., Bhatia, C., Sharma, A., & Trivedi, P. K. (2016). MicroRNA858 is a potential regulator of phenylpropanoid pathway and plant development. *Plant Physiology, 171*, 944–959. https://doi.org/10.1104/pp.15.01831

The Food and Agriculture Organization (FAO) database (2022). https://www.fao.org/faostat/en/

VSN International. (2019). *Genstat for Windows* (20th ed.). Hemel Hempstead, UK: VSN International.

Williams, D., Hackett, C. A., Karley, A., McCallum, S., Smith, K., & Britten, A. (2021). Seeing the wood for the trees, hyperspectral imaging for high throughput QTL detection in raspberry, a perennial crop species. *Fruit Research*. https://doi.org/10.48130/FruRes-2021-0007

Yamada, H., Hirokane, T., Ikeuchi, K., & Wakamori, S. (2017). Fundamental methods in ellagitannin synthesis. *Natural Product Communications*. https://doi.org/10.1177/1934578X1701200846

Yuan, T., Fujioka, S., Takatsuto, S., Matsumoto, S., Gou, X., He, K., et al. (2007). BEN1, a gene encoding a dihydroflavonol 4-reductase (DFR)-like protein, regulates the levels of brassinosteroids in *Arabidopsis thaliana*. *The Plant Journal, 51*, 220–233. https://doi.org/10.1111/j.1365-313X.2007.03129.x

Zhu, M.-C., Hu, R., Zhao, H.-Y., Tang, Y.-S., Shi, X.-T., Jiang, H.-Y., et al. (2021). Identification of quantitative trait loci and candidate genes controlling seed pigments of rapeseed. *Journal of Integrative Agriculture, 20*, 2862–2879. https://doi.org/10.1016/S2095-3119(20)63377-9