Article

Molecular Characterization of a Supergene Conditioning Super-High Vitamin C in Kiwifruit Hybrids

John McCallum[1,2], William Laing[3], Sean Bulley[3], Susan Thomson[1], Andrew Catanach[1], Martin Shaw[1], Mareike Knaebel[3], Jibran Tahir[3], Simon Deroles[3], Gail Timmerman-Vaughan [1] , Ross Crowhurst [4], Elena Hilario[4], Matthew Chisnall[2], Robyn Lee[2], Richard Macknight[2], Alan Seal[5]

[1] The New Zealand Institute for Plant & Food Research Limited, Private Bag 4704, Christchurch, NEW ZEALAND.

[2] Biochemistry Dept. University of Otago, Dunedin, NEW ZEALAND.

[3] The New Zealand Institute for Plant & Food Research Limited, Palmerston North, NEW ZEALAND.

[4] The New Zealand Institute for Plant & Food Research Limited, Auckland, NEW ZEALAND.

[5] The New Zealand Institute for Plant & Food Research Limited, Te Puke, NEW ZEALAND.

* Correspondence: John.mccallum@plantandfood.co.nz

Abstract: During analysis of kiwifruit derived from hybrids between the high AsA species *Actinidia eriantha* and *A. chinensis* var *chinensis*, we observed bimodal segregation of fruit AsA concentration suggesting major gene segregation. To test this hypothesis we performed whole-genome sequencing on pools of high and low AsA fruit from tetraploid *A. chinensis var. deliciosa x A. eriantha* backcross families. Pool-GWAS revealed a single QTL spanning more than 5 Mbp on chromosome 26, which we denote as qAsA26.1. A co-dominant PCR marker was used to validate this association in four diploid (*A. chinensis x A. eriantha*) x *A. chinensis* backcross families, showing that the *eriantha* allele at this locus increases fruit AsA levels by 250 mg/100 g fresh weight. Inspection of genome composition and recombination in other *A. chinensis* genetic maps confirmed that the qAsA26.1 region bears hallmarks of suppressed recombination. The molecular fingerprint of this locus was examined in leaves of backcross validation families by RNASEQ. This confirmed strong allelic expression bias across this region as well as differential expression of transcripts on other chromosomes. This evidence suggests that the region harboring qAsA26.1 constitutes a supergene, which may condition multiple pleiotropic effects on metabolism.

Keywords: kiwifruit, genomics, polyploidy, breeding, ascorbic acid, vitamin C

1. Introduction

Commercial kiwifruit varieties of species *Actinidia chinensis* var. *chinensis* are known as a rich source of dietary Vitamin C (AsA). However the related species *Actinidia eriantha* has AsA concentrations in its fruit of up to 800 mg/100 g fresh weight, but has small fruit with a bland flavor [1]. Recently a large-fruited high AsA *A. eriantha* cultivar ('White') has been described [2]. If this high concentration could be transferred by crossing to more palatable kiwifruit species, an ultra high health fruit could be developed. The availability of high-quality genome sequences for *A. eriantha* [3] as well as *A. chinensis var. chinensis* [4,5] provide the basis for functional and genetic approaches to aid such introgression.

The dominant pathway of AsA biosynthesis in *Actinidia* species including *A. eriantha* appears to be the L-galactose pathway [1], with AsA biosynthesis occurring early in fruit development, and then declining. The control of this pathway lies in an early committed step of biosynthesis in the

enzymes GDP-galactose phosphorylase (GGP) and GDP-mannose epimerase (GME), with some input from GDP mannose pyrophosphorylase (GMP) [6,7]. Transformation of plants to over-express GGP results in a several fold increase in fruit or tuber ascorbate [8] and over-expression of GME, which by itself has little effect, synergistically increases ascorbate yet further [9]. Oxidized ascorbate is also reduced by several enzymes which have also been implicated in controlling ascorbate concentrations, as have a range of transcription factors and other regulators [7]. In addition, the upstream open reading frame of the GGP gene has a role in controlling translation of the GGP gene and thus ascorbate concentration, forming a feed-back control loop in response to elevated ascorbate [9]. Thus a complex of enzymes and regulators controls ascorbate concentration in plants any of which may explain why *A. eriantha* has such very high ascorbate concentration.

Both in apples [10] and tomatoes [11] QTL mapping has successfully identified candidate genes for regulation of ascorbate content. In this paper we analyse the genetic basis for why *A. eriantha* has such high ascorbate by studying crosses between *A. eriantha* and other *Actinidia* species, and locate the chromosomal region conditioning super-high ascorbate levels in *A. eriantha*.

2. Results

2.1 Pooled Whole-Genome Sequencing and GWAS

Quantitative HPLC analysis of AsA levels in fruit harvested from tetraploid hybrid *Actinidia* backcross populations revealed evidence for bimodal segregation in all families as well as differences in family medians (Figure 1).

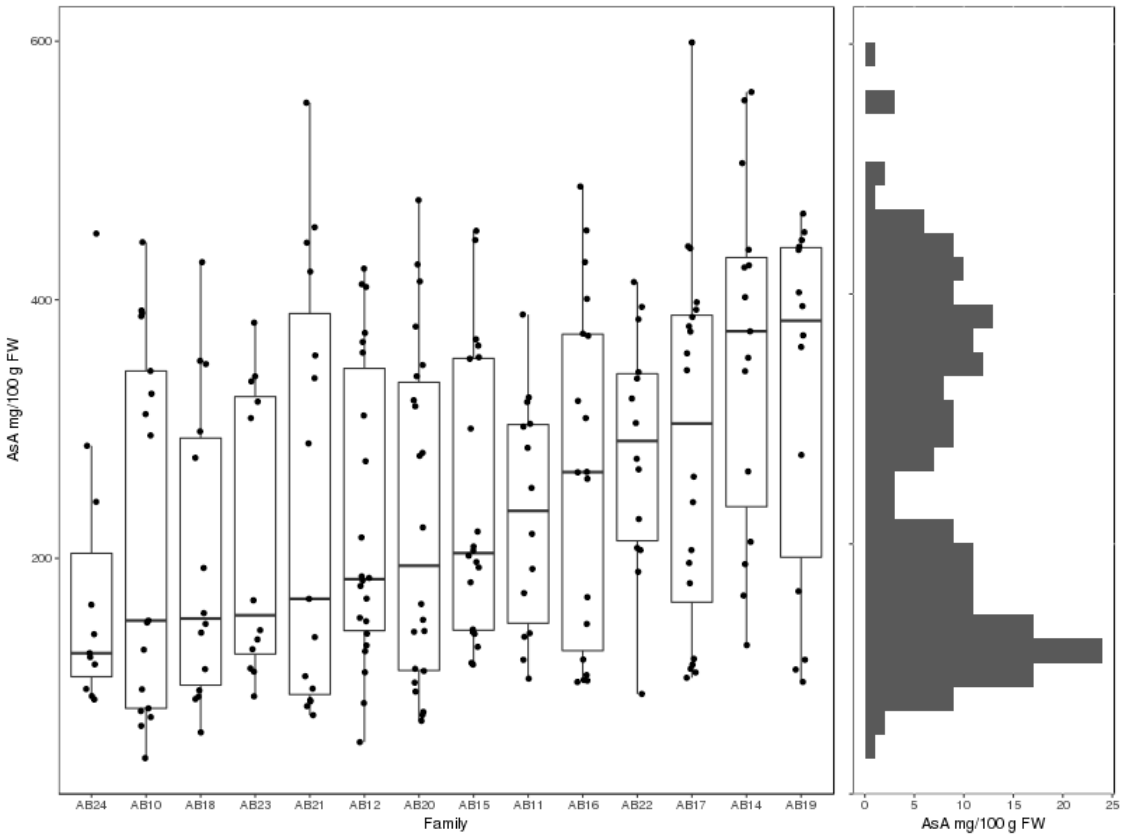

Figure 1 . Distributions of vine mean ascorbic acid concentration in tetraploid hybrid *Actinidia* families

The parents of these populations were selected from crosses between hexaploid *A. chinensis* var. *deliciosa* and diploid *A. eriantha* (Figure 2). Because of the complex polyploidy genome composition of these populations and the observation of bimodal segregation suggesting a major gene, we conducted

genetic analysis by pooled whole-genome sequencing, exploiting the availability of draft genome assemblies of *A. chinensis* var *chinensis* as reference [4,12]. Since there was wide variation for fruit weight and this was not correlated with AsA levels, we constructed pools by both traits to enable orthogonal, replicated tests of allele frequencies for each trait (Figure 3; Table 3).

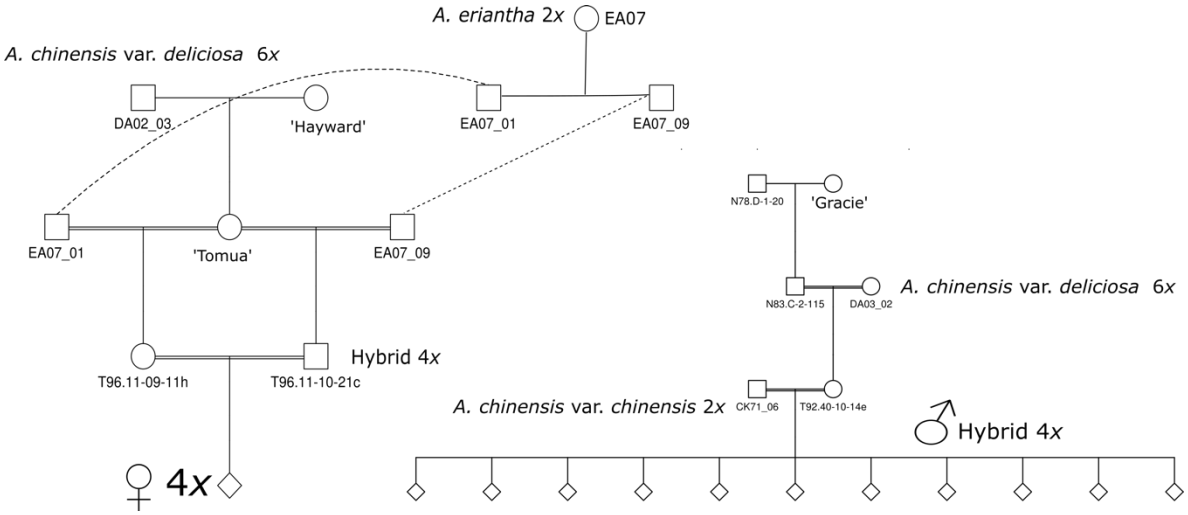

Figure 2 Pedigrees and ploidies of maternal parent (L) and of *A. chinensis* var. *deliciosa* x *A. chinensis* var. *chinensis* males (R) used to generate families used for pooled sequencing.

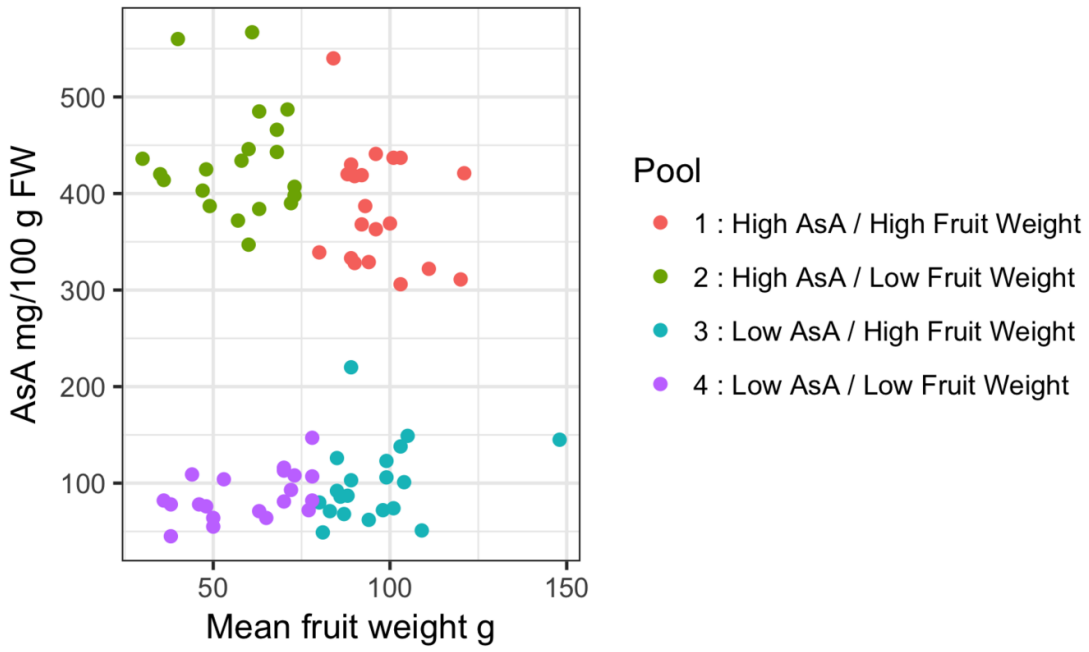

Figure 3 Allocation of samples to sequencing pools

Small insert paired end Illumina sequencing over two lanes yielded 965,452,550 reads with 92.8% Q30, and 86% of reads mapped to the Red5 PS1.1.68.5 pseudomolecules. Pool-GWAS (genome-wide association study) scans performed on both normalised and non-normalised read count data using Popoolation2 [13] revealed a single major QTL for AsA content on Chromosome 26 (Figure 4), but no significant associations with fruit weight (data not shown). Closer inspection of the Chromosome 26 region and windowed analysis using QTLseqR [14] revealed a broad distribution of significant scores for AsA on chromosome 26 (Figure 5; Table A1). SNPs showing association with pool AsA were observed over an interval of 7 Mbp . We denote this major QTL as qAsA26.1.

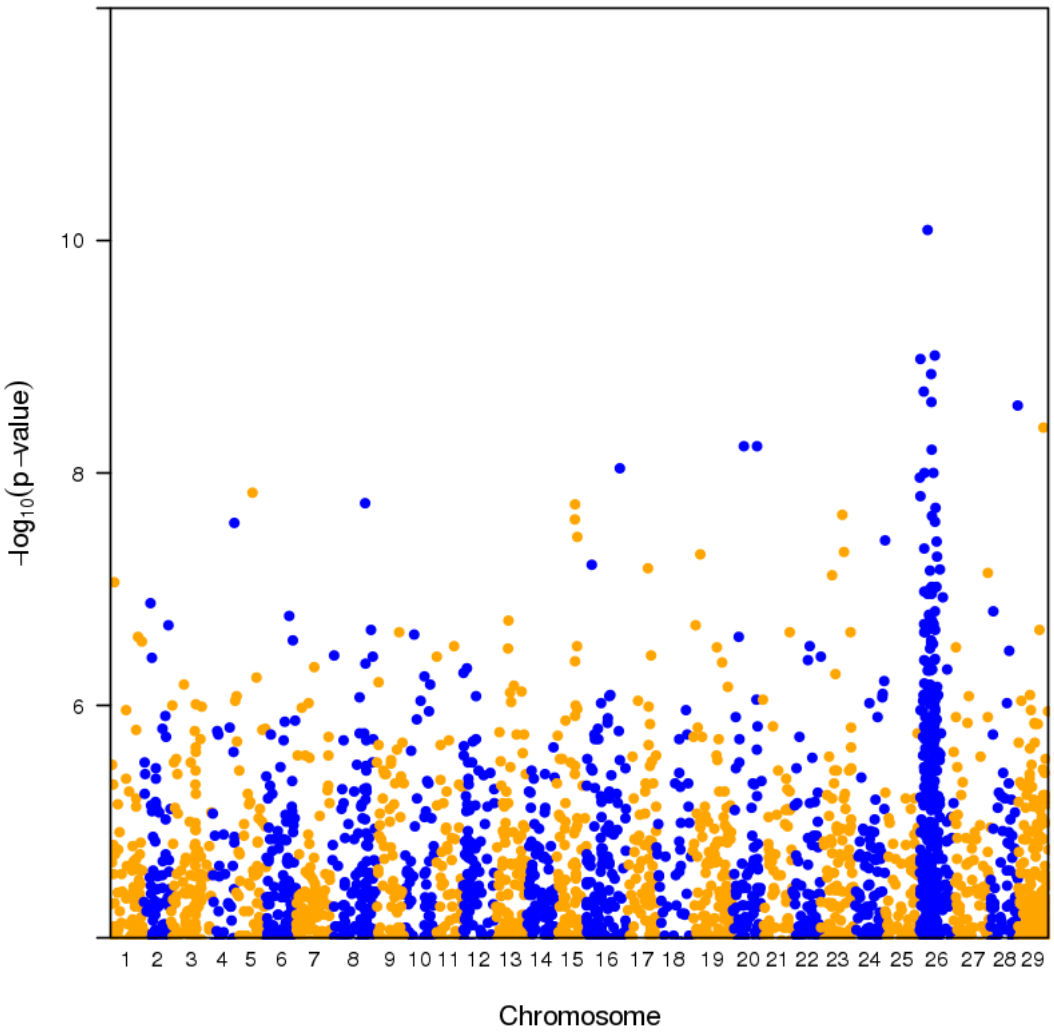

Figure 4 Pool-GWAS scan for fruit AsA concentration level using Popoolation2 . Symbols denote significance tests for association of individual SNPs with pool AsA level by Cochran-Mantel-Haenszel (CMH) Chi-Squared Test with normalised allele counts.

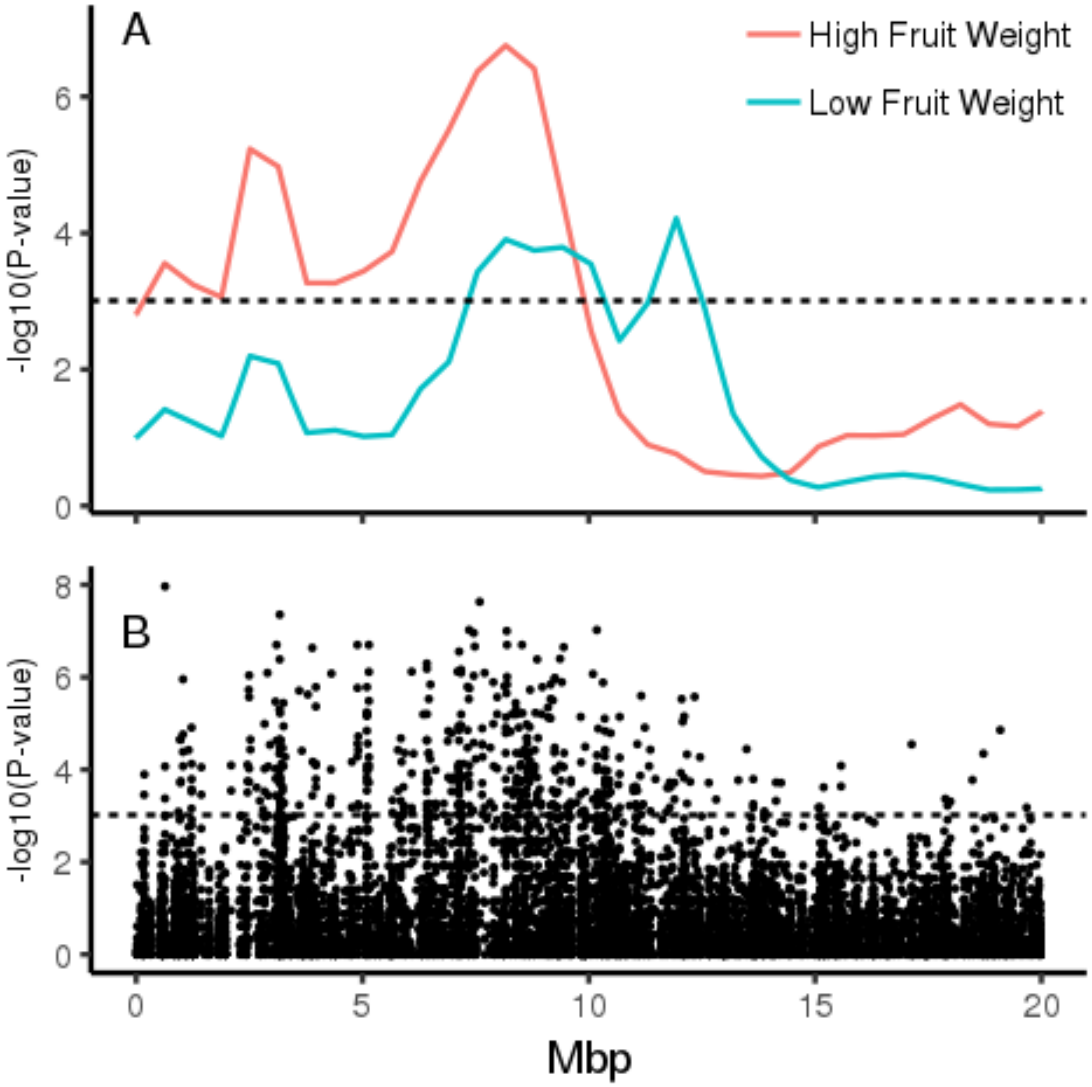

Figure 5 Pool GWAS analysis of fruit AsA levels on Chromosome 26 using A) 1 Mbp windowed analysis using QTLseqr performed separately in high and low fruit weight pools. B) CMH tests at individual SNPs using Popoolation2. Sites were restricted to fixed variants between *A. eriantha* and *A. chinensis*. Dashed lines denote false discovery rate (FDR) cutoff at $p < 0.001$.

2.2 Validation in Diploid Backcross Populations

To validate this association, we designed a set of high resolution melting (HRM) assays (Table 3). to the associated variants on chromosome 26 which were homozygous in the low AsA pools (Table A1). Marker KCH00062 targetting the polymorphisms at 7647158-7647167 bp, exhibited agreement with pool AsA levels in 78/80 samples used to construct sequencing pools. A two-way ANOVA model of fruit AsA concentration showed that marker dosage and paternal family explained 79% ($P < 2e\text{-}16$) and 10%(1.21e-07) respectively of total variance.

This marker was evaluated in a further six diploid backcross families: three (*A. eriantha x A. chinensis*) x *A. chinensis* and three (*A. chinensis x A. eriantha*) x *A. chinensis* (Figure 6). The maternal parent 11-06-16e of the EACK2 family used by Fraser *et al.* [15] was homozygous for the *A. chinensis* var. *chinensis* allele and the family did not have any high AsA (> 400 mg/100g FW) fruit. In the AI247 and AJ247 families (totaling N=196 progeny), ANOVA analysis indicated that the marker explained 78% of the phenotypic variance and residual analysis revealed 3/196 (1.5%) recombinants. The presence of the *eriantha* allele is associated with an increase in AsA content of approximately 250 mg/

100 g FW.

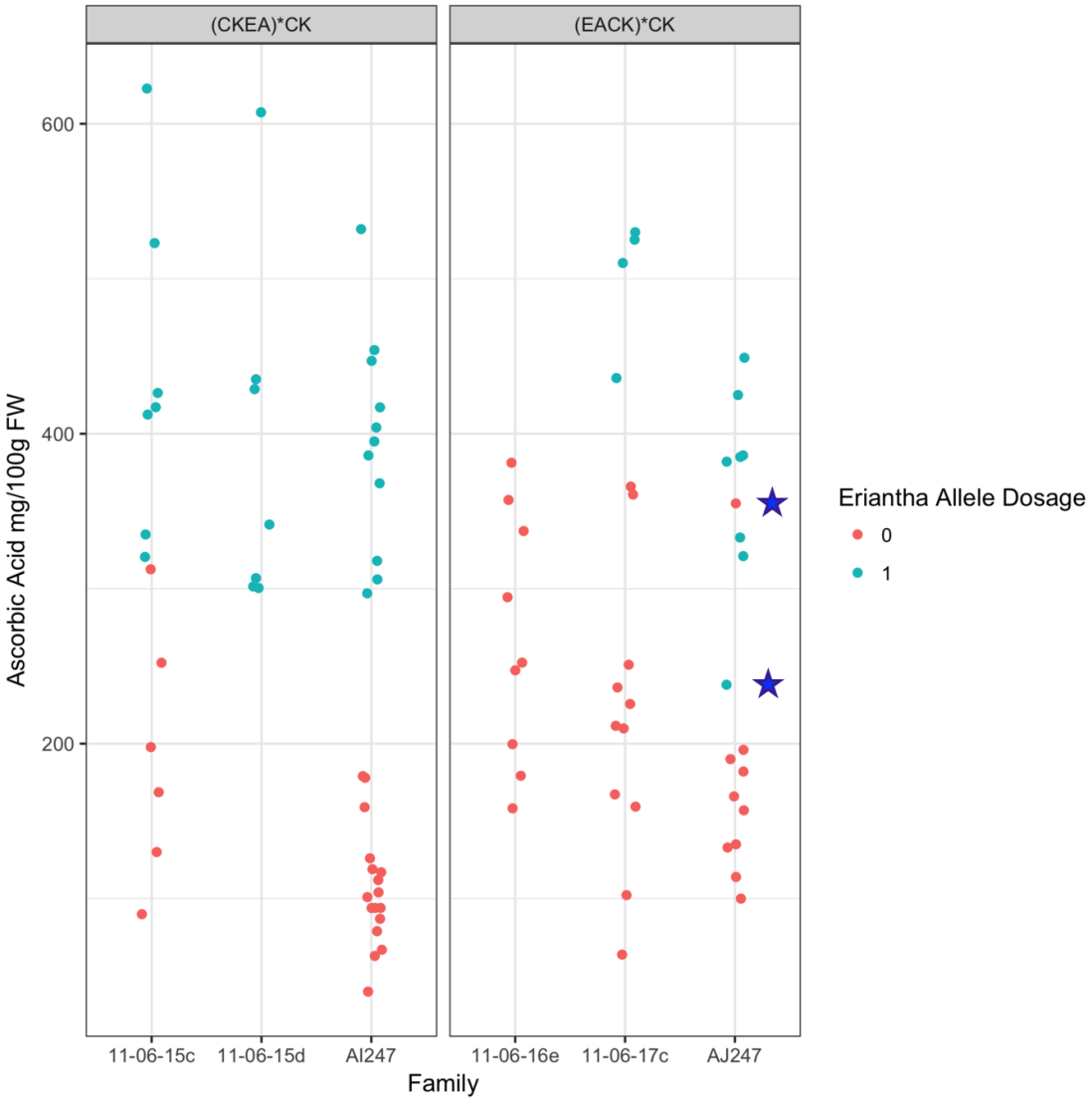

Figure 6 Segregation of the HRM marker KCH00062 in relation to fruit AsA content in six diploid backcross *Actinidia* families. Panels denote maternal cross type (*A. chinensis x A. eriantha* CKEA; *A. eriantha x A. chinensis* EACK). Two putative recombinants are denoted by stars.

Additional HRM markers were evaluated from targets in the 8.2-8.5 Mb interval and three informative markers were identified targetting SNPs at 8193148, 8453577 and 8874229 bp. The marker at 8453577 bp exhibited 10% recombination in the tetraploid families but the others exhibited complex segregation patterns and could not be scored. Efforts to design further co-dominant HRM markers in the 0-7 Mbp region were unsuccessful, suggesting that other marker types may be better suited to these highly heterozygous polyploid hybrids.

2.3 Genome Architecture of Actinidia Chromosome 26

Inspection of chromosome 26 repeat density and recombination estimates from genetic mapping [17] shows that the location of qAsA26.1 coincides with the boundary of a region with high repeat density and lower recombination (Figure 7). Alignment of the chromosome 26 pseudomolecules from the assemblies of *A. chinensis* 'Red5' [4] and the *A. eriantha* 'White' [3] indicate that these are highly collinear apart from differences in the terminal repeat-rich region (Figure A2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 June 2019 doi:10.20944/preprints201906.0293.v1

Peer-reviewed version available at Plants 2019, 8, 237; doi:10.3390/plants8070237

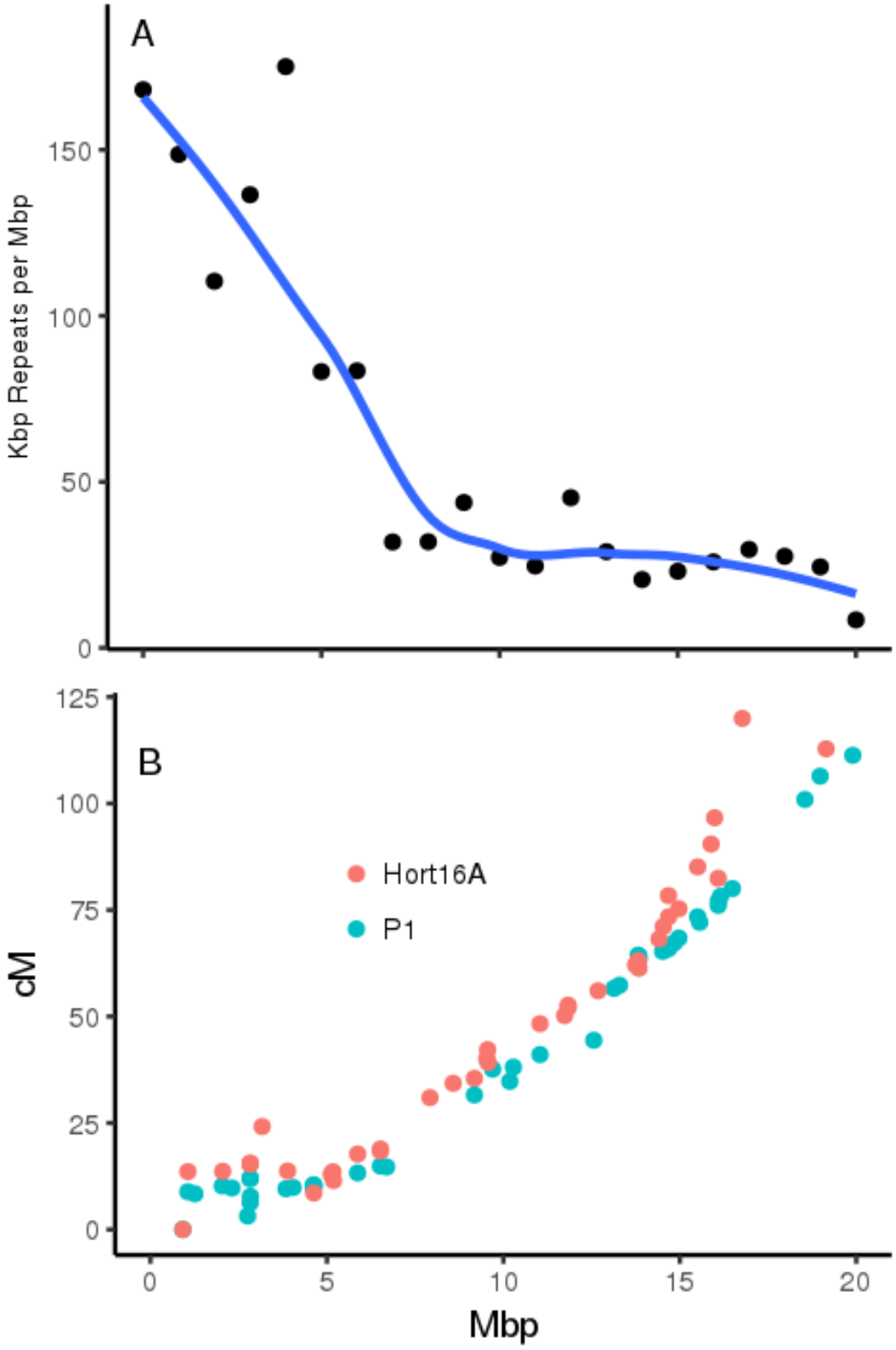

Figure 7 Composition and recombinational landscape of *A. chinensis* chromosome 26. **A**. Total repeat annotations based on the 'Red 5' assembly **B.** Physical versus recombination distance on *A. chinensis* 'Red5' Chromosome 26 estimated in male (green) and female (red) parental maps in the 'Hort 16A x P1' family .

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 June 2019 doi:10.20944/preprints201906.0293.v1

Peer-reviewed version available at Plants 2019, 8, 237; doi:10.3390/plants8070237

2.4 Characterising the qAsA26.1 Introgression in Leaf Tissues

To better characterise the qAsA26.1 introgression we compared the leaf transcriptome and metabolome of low and high AsA progeny in the diploid *A. chinensis* x *A. eriantha* backcross AI247 and AJ247 families used for marker validation. We chose analysis of leaf tissues for ease of reproducible sampling and because it has been shown that *A. eriantha* also exhibits very high leaf AsA levels [1]. Zhang *et al.* [17] have reported segregation for leaf AsA content in the cross between hexaploid *A. chinensis var. deliciosa* and a diploid *A. eriantha* × *A. chinensis* var. *chinensis.* We confirmed by HPLC that leaf AsA levels were higher in samples of immature leaves from backcross progeny carrying the introgression (ACH0007 homozygotes 10.4 mg/100g FW *versus* 25.3 mg/100 gFW in heterozygotes ; $p < 0.025$ by T-test). These analyses were performed on tissue samples collected in RNALater without the precautions necessary for good preservation of AsA, and are therefore lower than previously observed [1].

2.4.1 Pooled RNASEQ

RNASEQ was performed on three pools of backcross progeny with high fruit AsA which were heterozygous for the introgression and three pools of low AsA progeny lacking it, yielding 21.4-24.4 million reads per library. To determine the patterns of allelic expression on chromosome 26 we performed read assignment using PolyCat [18] based on a set of SNPs identified between *A. chinensis* and *A. eriantha.* This revealed that *A. eriantha* reads were essentially absent in low AsA pools across the first 10 Mbp of Chromosome 26 (Figure 9A), providing additional genetic evidence that recombination is strongly suppressed in this region.

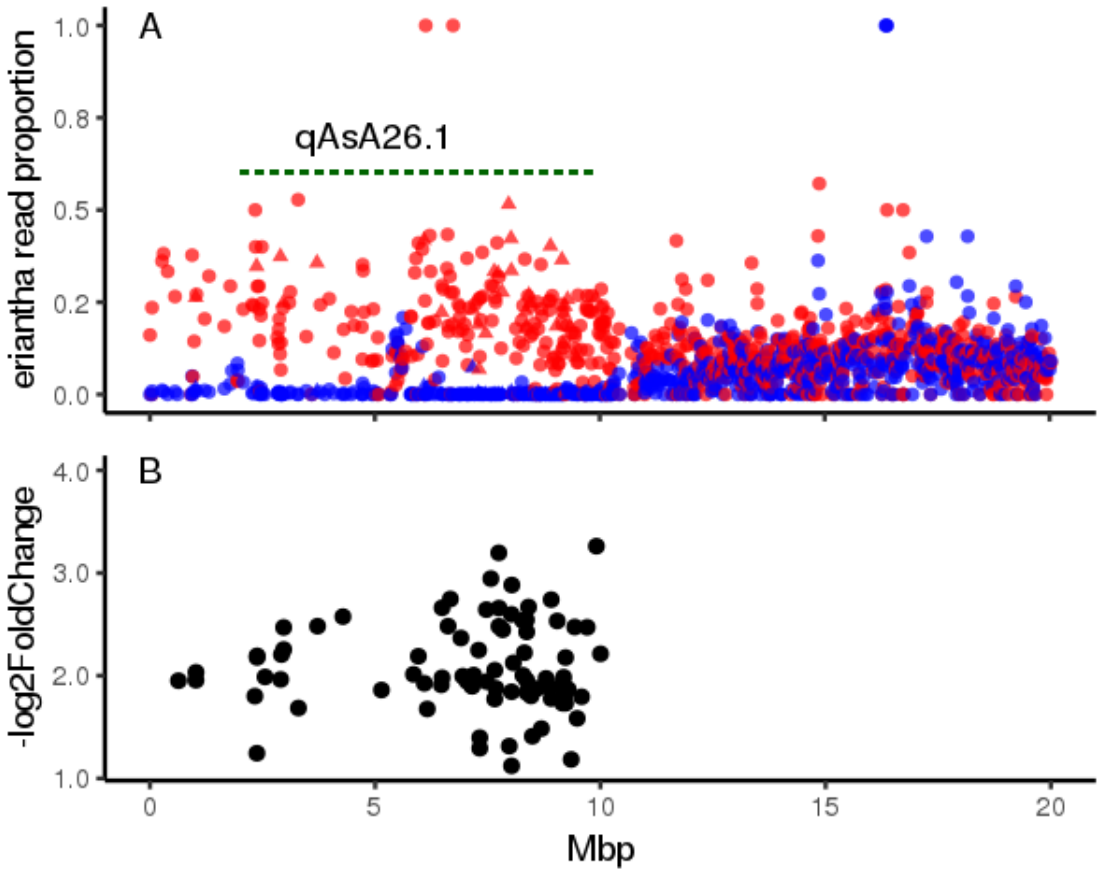

Figure 8 RNASEQ analysis of gene expression on *Actinidia* Chromosome 26 . Points denote gene models on the *A. chinensis* 'Red 5' assembly . A) Allelic expression proportion based on PolyCat read assignment . Red and blue symbols denote high and low AsA pools, respectively. Triangular points denote gene models with transcripts exhibiting differential expression. B) Genomic coordinates and Log 2-fold expression differences of genes showing significant differential expression.

Differential expression analysis revealed 113 differentially expressed transcripts (DETs) with between high and low AsA pools at an FDR < 0.05 (Supplementary Table 1). Of these, 82 mapped to the qAsA26.1 region of Chromosome 26 (Figure 8B) of of these, 61 mapped to annotated gene models. Because of the degree of allelic divergence between the two *Actinidia* species, transcript analysis based on the *de novo* assembly we used would be expected to frequently reveal novel alleles or splice variants absent in *A. chinensis.*

Prior to library construction, qPCR analysis of individual samples for GGP (GDP-L-galactose phosphorylase), GMD (GDP-D-mannose-4,6-dehydratase), DHAR (dehydro-AsA reductase) showed no evidence for differences between high and low AsA samples (P > 0.35 for all T-tests). This observation was confirmed in the RNASEQ data, which showed no evidence of differential expression in transcripts annotated as GGP (KEGG orthology number K14190) , GME (KO K10046), DHAR (KO K08232) , VTC4 (K10047) and GMD (K01711) (Supplementary Table 2).

Peer-reviewed version available at Plants 2019, 8, 237; doi:10.3390/plants8070237

Contig	log2FoldChange	adjusted P-value	Normalised read count High Pool Rep 1	Normalised read count High Pool Rep 2	Normalised read count High Pool Rep 3	Normalised read count Low Pool Rep 1	Normalised read count Low Pool Rep 2	Normalised read count Low Pool Rep 3	Chrom	Gene Model	Annotation
TRINITY_DN125630_c0_g2_i1	-1.6	0.0	197.71	344.66	225.3	70.25	67.17	76.34	CHR3	Acc2955.1	Laccase-7, Precursor (putative)
TRINITY_DN125630_c0_g5_i2	-1.5	0.0	260.83	333.18	217.43	74.26	103.72	69.07	CHR3	Acc2955.1	Laccase-7, Precursor (putative)
TRINITY_DN125630_c0_g5_i3	-1.33	0.01	162.44	199.49	172.18	49.17	76.06	59.08	CHR3	Acc2955.1	Laccase-7, Precursor (putative)
TRINITY_DN129828_c1_g5_i5	1.77	0.02	25.06	4.18	8.85	65.23	70.14	117.24	CHR3	Acc3372.1	T-complex protein 1 subunit epsilon (TCP-1-epsilon) (putative)
TRINITY_DN118492_c0_g1_i4	1.87	0.0	14.85	83.55	42.31	334.19	309.19	179.05	CHR3	Acc3845.1	Probable beta-glucosidase btgE, Precursor
TRINITY_DN118884_c0_g7_i1	1.87	0.01	1.86	12.53	10.82	90.32	50.38	55.44	CHR3	Acc3860.1	Cytochrome c1 2, heme protein, mitochondrial (Cytochrome c-1 2), Precursor (putativ
TRINITY_DN129466_c0_g6_i6	2.12	1.86e-05	45.48	250.66	124.95	1171.16	1034.26	664.38	CHR3	Acc3864.1	Magnesium-protoporphyrin IX monomethyl ester [oxidative] cyclase, chloroplastic (Mg
TRINITY_DN117715_c1_g3_i1	1.43	0.02	174.51	378.09	266.63	1407.0	832.74	500.78	CHR3	Acc3890.1	Tubulin beta-4 chain
TRINITY_DN129466_c0_g6_i3	1.9	0.0	65.9	237.09	94.45	1040.7	832.74	370.82	CHR5	Acc5209.1	Magnesium-protoporphyrin IX monomethyl ester [oxidative] cyclase, chloroplastic (Mg
TRINITY_DN102173_c0_g2_i1	1.79	0.01	36.2	22.98	25.58	350.24	108.66	65.44	CHR6	Acc12049.1	Polyadenylate-binding protein 8 (PABP-8) (putative)
TRINITY_DN129080_c0_g2_i1	-1.17	0.04	426.98	416.73	234.16	141.5	148.17	139.06	CHR8	Acc9639.1	Anthocyanidin reductase ((2S)-flavan-3-ol-forming) (VvANR) (putative)
TRINITY_DN121164_c0_g1_i4	-1.56	0.01	297.03	318.55	188.9	65.23	37.54	108.15	CHR9	Acc10699.1	Germin-like protein 5-1, Precursor (putative)
TRINITY_DN130940_c2_g1_i1	-2.87	9.73e-09	125.31	80.42	61.0	5.02	5.93	1.82	CHR12	Acc29528.1	Mitochondrial import inner membrane translocase subunit TIM17-2 (similar to)
TRINITY_DN126502_c0_g1_i1	2.33	1.86e-05	4.64	27.16	30.5	159.57	401.06	134.51	CHR13	Acc14777.1	Calcium uniporter protein 6, mitochondrial, Precursor (similar to)
TRINITY_DN123616_c1_g1_i1	-2.33	8.05e-06	73.33	69.98	72.81	3.01	9.88	8.18	CHR16	Acc18424.1	CUB and EGF-like domain-containing protein 1
TRINITY_DN126733_c0_g1_i7	-1.12	0.03	325.8	245.44	301.06	122.44	97.79	144.51	CHR18	Acc20147.1	Zinc finger protein CONSTANS-LIKE 5 (probable)
TRINITY_DN123292_c0_g2_i2	-1.65	0.0	270.11	115.93	124.95	37.13	37.54	47.26	CHR18	Acc20170.1	Putative GDP-L-fucose synthase 2 (AtGER2)
TRINITY_DN116626_c0_g2_i1	-1.78	0.04	44.55	27.16	7.87	0	0	0	CHR25	Acc28491.1	L10-interacting MYB domain-containing protein (probable)
TRINITY_DN122105_c0_g1_i1	-1.37	0.01	187.5	146.22	121.01	57.2	46.43	45.44	CHR25	Acc28707.1	Ubiquinol oxidase 1a, mitochondrial, Precursor (putative)
TRINITY_DN126144_c0_g3_i6	-1.83	0.02	34.34	17.76	15.74	0	0	0	CHR25	Acc29051.1	COBW domain-containing protein 1 (COBP) (probable)

TRINITY_DN117186_c0_g2_i2	-1.82	0.02	65.9	38.64	38.37	10.04	4.94	1.82	CHR25	Acc29489.1	UPF0162 protein PD_0709 (probable)
TRINITY_DN128356_c0_g1_i6	-1.47	0.02	107.67	206.8	122.98	53.19	39.51	30.9	CHR25	Acc29080.1	Pectin acetylesterase 8, Precursor (putative)
TRINITY_DN113632_c0_g1_i1	-1.44	0.03	451.11	1051.75	476.19	107.38	263.75	209.04	CHR25	Acc12497.1	BURP domain protein RD22, Precursor (similar to)
TRINITY_DN130185_c0_g1_i4	-2.3	2.26e-05	47.34	67.89	82.64	4.01	8.89	5.45	CHR29	Acc33009.1	CRM-domain containing factor CFM3, chloroplastic/mitochondrial (ZmCFM3), Precursor

169 **Table 1** Genome locations and associated annotations of *de novo* assembled transcripts not mapping to kiwifruit chromosome 26 which exhibited differential expression
170 between high and low AsA pools at FDR $p < 0.05$.

Significant expression differences were observed for 31 transcripts mapping to chromosomes other than chromosome 26, of which 26 mapped to gene models (Table 1). These include one DET of particular interest. The transcript TRINITY_DN123292_c0_g2_i2 maps to the gene model Acc20170.1 (GenBank: PSS04323.1) encoding a putative GDP-L-fucose synthase 2 homologous to *Arabidopsis* GER2 (KEGG K02377; [19]). We previously reported significantly higher expression of GER in *A. eriantha* compared to *A. chinensis* [1; Figure 4, Panel D] (Authors note: This panel is mis-labelled as GMD) . Because fucose synthesis draws upon the same substrate pool as ascorbate [6]; [7], this may have implications for regulation of mannose channeling to ascorbate. The observation of association between qAsA26.1 genotype and transcript expression at this locus suggests that qAsA26.1 contains transcriptional regulators of carbohydrate metabolism. DETs annotated as beta-glucosidase (Acc03845.1) and pectin acetylesterase (Acc29080.1) were also observed. A further DET of potential functional relevance to AsA metabolism is TRINITY_DN120596_c0_g1_i7 mapping to Acc29025.1 on chromosome 26. This gene is annotated as a component of the dolichol-phosphate mannose synthase complex which mediates mannosylation of glycans [20]. Similar associations between competing carbohydrate metabolic pathway expression and fruit AsA have been reported in studies of tomato interspecific introgressions [21] and ripening[22].

Differential expression was also observed of homologs of laccase (Acc02955.1) and anthocyanidin reductase (Acc09639.1) mapping to chromosomes 3 and 8 respectively. The differentially expressed transcripts identified on chromosome 26 include both structural genes (Acc29585.1, 4-coumarate CoA ligase ; Acc29568.1, Shikimate O-hydroxycinnamoyltransferase;) and transcriptional regulators of polyphenol metabolism (Acc18102.1, AtMyb4 homolog). Collectively these observations suggest that polymorphism at qAsA26.1 could exert a direct or indirect influence on polyphenol metabolism. Over-expression of *GGP* in tomato and strawberry not only increased ascorbate but also increased flavonoids and phenylpropanoids [8] . Further evidence for cross-talk comes from studies of *Arabidopsis vtc* mutants, which have shown that these are also impaired in transcriptional regulation of anthocyanin synthesis [23].

2.4.2 Untargetted Metabolomics

Liquid chromatography-MS analysis of leaf extracts from the kiwifruit revealed some evidence for more frequent occurrence of elevated levels of flavonoids and phenylpropanoids in those carrying the *eriantha* marker allele (Table 2). This data is from a single time point in an orchard environment and we expect it would be highly influenced by local variability in infection by the pandemic *Pseudomonas syringae* var. Actinidiae [16,24]. Targetted metabolomic analyses of fruit and vine tissues with standards, especially for antioxidant and key carbohydrates is desirable to better characterise the phenotype of qAsA26.1 alleles.

205 **Table 2** Putative identities and molecular formulae of metabolites exhibiting difference at $p < 0.05$ based on untargetted metabolomics. The samples comprised a total of
206 N=130 leaf samples (no technical replication) from diploid backcross AI247 and AJ247 vines with (N=65) or without (N=65) the *A. eriantha* allele for marker KCH00062.

Column	RT [min]	Putative Candidate(s)	Molecular Weight	Formula	Group area : eriantha allele (+)	Group area : eriantha allele (-)	Ratio: + / -	Log2 Fold Change	P-value
C18	3.37		312.09	C14 H16 O8	7738.38	3601.05	2.15	1.1	0.0
C18	3.1	caffeoyl quinide	336.08	C16 H16 O8	3969.41	1585.49	2.5	1.32	0.01
C18	1.09		338.06	C15 H14 O9	4282.86	2079.64	2.06	1.04	0.0
C18	2.85		366.13	C13 H22 N2 O10	3257.78	1545.54	2.11	1.08	0.02
C18	4.59	carbohydrate derivative	416.21	C21 H28 N4 O5	90832.77	16046.98	5.66	2.5	0.01
C18	4.48	carbohydrate derivative	417.09	C13 H21 N7 O3 P2 S	3259.7	781.94	4.17	2.06	0.0
C18	4.76	carbohydrate derivative	430.22	C22 H30 N4 O5	121431.57	44149.79	2.75	1.46	0.01
C18	4.59	carbohydrate derivative	430.22	C22 H30 N4 O5	13720.41	5974.26	2.3	1.2	0.04
C18	4.08		436.19	C19 H28 N6 O4 S	4271.37	1198.71	3.56	1.83	0.01
C18	4.93	Kaempferol-3-O-glucoside	448.1	C21 H20 O11	52765.53	8300.52	6.36	2.67	0.0
C18	4.35	carbohydrate derivative	456.19	C19 H28 N4 O9	2884.91	1385.24	2.08	1.06	0.03
C18	3.49	spermidine derivative	456.2	C17 H29 N8 O5 P	7872.29	2542.55	3.1	1.63	0.01
C18	4.59	carbohydrate derivative	476.23	C19 H28 N10 O5	9879.76	4216.6	2.34	1.23	0.02
C18	4.76	carbohydrate derivative	476.23	C16 H37 N4 O10 P	100274.28	36739.65	2.73	1.45	0.01
C18	4.99	Isorhamnetin 3-galactoside	478.11	C22 H22 O12	46274.3	2049.91	22.57	4.5	0.04
C18	3.49	Fatty acid like	491.24	C23 H33 N5 O7	12980.96	3197.7	4.06	2.02	0.02
C18	3.55	organic acid	498.21	C23 H36 N2 O6 P2	11612.98	4222.33	2.75	1.46	0.0
C18	4.37	carbohydrate derivative	516.15	C23 H24 N4 O10	13026.27	3798.38	3.43	1.78	0.02
C18	3.85		531.22	C25 H44 N O3 P3 S	4708.15	1379.95	3.41	1.77	0.0
C18	3.82	Quercetin-carbohydrate derivative	549.23	C21 H32 N11 O5 P	51922.7	13083.19	3.97	1.99	0.01
C18	5.09	glutathione derivative	549.24	C24 H37 N7 O4 P2	22422.9	6422.03	3.49	1.8	0.02
C18	4.12		561.23	C20 H37 N9 O6 P2	8968.49	2888.64	3.11	1.63	0.03
C18	3.91		565.13	C28 H30 N3 O2 P3 S	3647.66	1762.66	2.07	1.05	0.0
C18	3.06		581.17	C27 H27 N5 O10	3067.62	774.45	3.96	1.99	0.01
C18	3.96	organic acid	586.23	C27 H35 N6 O7 P	6628.24	1273.95	5.2	2.38	0.03
C18	4.65	Luteolin-like	742.38	C36 H57 N8 O3 P3	40140.01	18953.97	2.12	1.08	0.03
C18	4.33	glucose derivative	760.39	C38 H56 N4 O12	11611.09	4014.93	2.89	1.53	0.05
C18	3.16		771.31	C28 H60 N3 O13 P3 S	20076.74	9413.9	2.13	1.09	0.01
Helic	1.09	Organic acid derivative	145.95		303336.81	106554.16	2.85	1.51	0.02

Helic	1.64	carbohydrate derivative	192.08	C11 H12 O3	47978.16	20483.51	2.34	1.23	5.34e-06
Helic	1.34	glycosylated phenylpropanoid	222.09	C12 H14 O4	33395.63	11436.95	2.92	1.55	0.02
Helic	3.54		241.98	C4 H9 N2 O4 P3	51090.99	21734.49	2.35	1.23	0.04
Helic	3.53		247.97	C5 H12 O3 S4	29426.88	4357.26	6.75	2.76	0.01
Helic	1.34		266.08	C13 H14 O6	26764.49	10639.77	2.52	1.33	0.03
Helic	5.9	coumaric acid deriv	282.07	C13 H14 O7	9722985.89	4295374.34	2.26	1.18	0.05
Helic	2.0	coumarin glycoside	324.08	C16 H12 N4 O4	37094.1	12640.85	2.93	1.55	1.97e-05
Helic	6.99		449.04	C21 H12 N3 O7 P	8334381.62	2709273.63	3.08	1.62	0.05
Helic	6.21		534.16	C21 H31 N2 O12 P	3488.47	1650.29	2.11	1.08	0.04
Helic	3.35		549.13	C24 H28 N3 O8 P S	13522.12	4106.46	3.29	1.72	0.0

207

3. Discussion

This study confirms the findings in other horticultural crops such as *Cucumis* [25-26] that pooled sequencing offers a cost-effective and practical means to conduct genome scans in segregating plant populations. Since restricted recombination will preclude further genetic dissection of the locus the application of more sophisticated RNASEQ strategies for analysis of differential transcript usage and QTL [27] would enable a more detailed dissection of allelic and splice variation in future studies. Since our preliminary evidence suggests the potential for complex pleiotropic effects, more detailed metabolic profiling would be desirable.

The qAsA26.1 QTL is notable for its large effect, size and simple dominant inheritance. Although large effect QTL (> 20%) for AsA levels have been reported in other fruits such as apple [28] and tomato [21,22,29], this QTL leads to AsA levels an order of magnitude higher. The structural, linkage and expression data presented here suggest that this QTL constitutes a supergene- a group of tightly linked loci inherited as a single Mendelian locus [30] . Supergenes commonly exert multiple pleiotropic effects and may be key to preserving adaptive variation through protecting a haplotype comprising multiple genes [31]. The qAsA26.1 region bears many similarities to the partially differentiated *Actinidia* sex chromosome (chromosome 25; [32]). Whereas *A. chinensis* is widely distributed in eastern lowland China, *A. eriantha* is restricted to southeastern China [33]. Because AsA can play multiple functional roles in higher plants including as a key redox signal in responses to biotic and abiotic stresses [34], it may be speculated that this extended haplotype has been preserved due to its benefits on adaptive fitness. More detailed functional analysis of the genes lying within qAsA26.1 may permit testing whether the locus action is due to a single as opposed to multiple linked regulators [35]

The simple inheritance and large effect of this QTL offer some interesting opportunities not only for plant breeding but also for studies of AsA in human and plant physiology. Our findings suggest that practical genetic markers may be easily obtained and applied due to limited recombination and that these could be used to develop breeding lines fixed for high AsA alleles of qAsA26.1 . The availability of the 'White' genome assembly will greatly simplify design of allele-specific markers that can be applied in highly heterozygous and polyploid backgrounds. Selecting lines with comparable eating qualities expressing 'normal' or 'super-high' AsA could provide unique materials for human dietary studies. Similarly, the ability to obtain both male and female vines with significantly different AsA content in vegetative tissues would allow replicated testing of hypotheses concerning the role of AsA in plant adaptation and fitness. In addition to marker-based methods, we hope that use of such materials may facilitate discovery of new targets for improvement of AsA levels that are transferable to other crops [36,37].

4. Materials and Methods

4.1 Plant Materials and Phenotyping

4.1.1 Tetraploid Populations

Pool-GWAS was performed on a set of 80 individuals from 11 (*A. chinensis* var. *deliciosa* x *A. eriantha*) x (*A. chinensis* var. *deliciosa* x *A. chinensis* var *chinensis*) families. The common maternal parent was a tetraploid high AsA hybrid vine generated by sib mating F_1 progeny from an *A. chinensis* var. *deliciosa* x *A. eriantha* cross. The *A. eriantha* plants were seedlings originating from seed gifted by the Guangxi Institute of Botany, Guilin, China in 1988. A series of hybrid populations was generated by pollinating this with F_1 male progeny from an *A. chinensis* var. *deliciosa* x *A. chinensis* var *chinensis* cross (Figure 2) . Seedlings were planted at the Plant & Food Research Centre in Kerikeri New Zealand (Lat 35.2 deg S) in 2010 and analysis of fruit AsA was performed in 2013.

4.1.2 Diploid Populations

Marker validation was performed in three (*A. chinensis* var. *chinensis* x *A. eriantha*) x *A. chinensis* var. *chinensis* (CKEA x CK) and three (*A. eriantha* x *A. chinensis* var. *chinensis*) x *A. chinensis* var. *chinensis* (EACK x CK) backcross families. Four of these families were previously used to map petal colour [39]. The two additional additional families, which were also employed for RNASEQ, AJ247 (EACK x CK) and AI247 (CKEA x CK) had the same A. *eriantha* parentage respectively as populations EACK2 and CKEA3 and CKEA4 reported by Fraser *et al.* [15].

4.1.3 Ascorbate Analyses

Three whole fruit per seedling were analysed. Each fruit was cut equatorially as a 1 mm slice using a double bladed knife. The three fruit slices were immediately placed in a plastic 15 mL tube and frozen in liquid nitrogen, then stored at -80C until analysis. Fruit were then thawed and centrifuged at 4000g to separate solid material from the juice. It was critical to freeze the fruit before analysis as directly centrifuged fruit juice gave a much lower ascorbate reading. A 0.1 mL aliquot of the juice was then transferred to a microtube containing 0.9 mL of 0.8% w/v metaphosphoric acid, 2 mM EDTA and 2 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP HCL). These samples were then centrifuged at 14,000g for 15 minutes to clarify the juice and then analysed by HPLC using a rocket column (Altima C18 3 micron from Phenomenex Ltd (Auckland New Zealand) at 35 C. Ascorbate was quantified by injecting 5 µL into a Dionex Ultimate® 3000 Rapid Separation LC system (Thermo Scientifc). Instrument control and data analysis was performed using Chromeleon v7.2 (Thermo Scientific). Solvent A was 5 mL methanol, 1mL 0.2M EDTA pH 8.0 and 0.8 mL o-phosphoric acid in 2 L. Solvent B was 100% acetonitrile. The flow as 1.0 mL/min and the linear gradient started with 100% A and B was increased to 30% at 4.5 min, then to 90% B at 6 min. The column was then washed with 100% B and then returned to 100% A. The column was monitored at 245 nm and ascorbate quantified by use of authentic standards. Ascorbate was verified by its UV spectrum. This method gave the sum of oxidized and reduced ascorbate, namely total ascorbate. Ascorbate concentration in the juice was calculated directly and in preliminary assays compared to ascorbate extracted from powdered flesh. The juice method gave about a 5% higher result than the powdered whole fruit method.

4.3 Pooled DNA Sequencing

4.3.1 Library Preparation

DNA was isolated from leaf bud tissue collected in spring 2015 using a CTAB extraction method [38] followed by purification with Qiagen columns and quantitated using the 3500 Genetic Analyzer (Applied Biosystems™). Four normalised DNA pools were created of 20 individuals each as shown in Table 3.

Table 3 Summaries of sequencing pool phenotypes

Pool ID	Description	Mean AsA mg/100g FW	SD	Mean fruit weight g	SD
1	High AsA/High Fruit weight	385.9	59.5	96.6	10.85
2	High AsA/Low Fruit weight	433.55	57.1	56.6	13.51
3	Low AsA/HighFruit weight	100.15	41.12	95.65	15.13
4	Low AsA/LowFruit weight	87.25	24.56	59.85	15.04

Small-insert Thruplex DNA-seq libraries (Rubicon Genomics Ptd) were synthesised at NZ Genomics Ltd and sequenced on two lanes of Illumina Hi-Seq 2500 yielding 965 million reads

totaling 120 Gbp with 92.8% > Q30. Quality control using FastQC Screen (http://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/) revealed that 85-88% of reads mapped to the Red 5 *A. chinensis*_ var *chinensis* reference version PS1 1.68.5 and 6% mapping to *Actinidia* chloroplast reference [39].

4.3.2 Sequencing Data Processing

Bam alignment files for variant calling were generated following GATK best-practice approaches [41]. Reads were aligned using BWA-MEM v0.7.12[42] to pseudomolecules of draft assembly version PS1_1.68.5 of *A. chinensis var. chinensis* Red 5 [4,43], an inbred female genotype related to 'Hong Yang' [12]. This draft assembly has been deposited at https://doi.org/10.5281/zenodo.1297303.

Bam files were merged using Samtools 1.3.1 [47] and read groups were added using Picard Tools (http://broadinstitute.github.io/picard/) AddOrReplaceReadGroups. Duplicates were marked with Picard MarkDuplicates and indel realignment was performed using GATK RealignerTargetCreator and IndelRealigner. Depth of coverage in regions of interest was calculated using GATK depthofcoverage.

4.3.3 Pooled GWAS and Variant Analysis

Pool-GWAS scans for association of individual SNPs with AsA and fruit weight were performed using Popoolation2[13]. Variants were summarised using samtools pileup (flags -B -Q 0) and called using popoolation mpileup2sync.jar with option –min-qual 20. Replicated contingency tests were initially performed on non-normalised data over AsA concentration and fruit weight strata using the cmh-test.pl script (flags –min-count 6 –mincoverage 4 –max-coverage 120). To facilitate comparison over sites with varying coverage, common odd ratios were calculated for significant SNPs using R mantelhaen.test.

Subsequent analyses focused on the genic regions using data resampled with replacement to a read depth of 40 using subsample-synchronized.pl. CMH tests p-values were adjusted for multiple testing using R p.adjust with the Benjamini & Hochberg correction. Output files for CMH tests are available as supplementary material at 10.5281/zenodo.1309045

To complement the SNP-based analysis, windowed scans for AsA QTL were performed by Next Generation Sequencing Bulked Segregant Analysis (NGS-BSA) [49] using the R package QTLseqr [14] . Input files were generated from VCF files separately for high and low fruit weight samples using samtools bcftools (http://www.htslib.org/doc/bcftools.html), filtering on a set of fixed polymorphisms (file PS1_EA_specific_SNPs.csv.gz in 10.5281/zenodo.3257749) identified between a set of *A. chinensis* genotypes [33] and *A. eriantha* using Bambam intersnp [50]. Two pairs of bulks were compared: High AsA/High Fruit Weight versus Low AsA/High Fruit Weight (pools 1 and 3) and High AsA/Low Fruit Weight and Low AsA/Low Fruit Weight (pools 2 and 4)(Table 2). QTLseqr accepts two population types, F_2 and RIL. However, the lines used for constructing these pools were the result of backcrosses with the SNP data filtering to collect only alleles segregating in EA, therefore the function simulateAlleleFreq (https://rdrr.io/github/bmansfeld/QTLseqR/man/simulateAlleleFreq.html) was modified to permit analysis of a backcross population type (called BC4x) where the expected allele frequency was 0:0.25 and the expected segregation ratio was 1:1. NGS-BSA analysis was conducted to estimate QTL locations based on allele frequency differences among the pairs of pools. SNPs from all 29 chromosomes were analysed in each single analysis. The population type was set to BC4x, the window size was 500Kb, and the simulations were bootstrapped 10,000 times. The FDR was set to $P < 0.001$ based on adjustment by the method of Benjamini and Hochberg [51].

For downstream analysis, SNPs and indels were called with the frequentist variant caller Varscan2 v2.4.2[48], using a hard filter for MAF > 0.1, minimum coverage 20 and only reporting sites called in all 4 pools. VCF files are available at 10.5281/zenodo.1309045.

4.3.4 Chromosomal Analyses

Alignment of Red5 and *A. eriantha* pseudomolecules were performed using Last [46] following repeat masking with Windowmasker [45]. LTR elements were annotated using LTRHarvest [46]. Recombination distances for chromosome 26 were determined using Joinmap3 (https://www.kyazma.nl/index.php/JoinMap/) from GBS genotyping data used to construct a genetic linkage map in the 'Hort16A x P1' family (N=236) [16].

4.3.5 PCR Marker Design for Validation

Filtered SNP loci detected by Popoolation2 cmh_test.pl which were homozygous in low AsA pools were used as targets for HRM primer design (Table 3) using the script https://github.com/PlantandFoodResearch/pcr_marker_design/blob/master/design_primers.py [52] . PCR amplification and HRM analysis on a on a Roche LightCycler 480 were performed as described previously [52]

Table 3 Primer sets used for high-resolution melting (HRM) in this study and their target intervals on genome references.

Primer set name	Forward primer	Reverse Primer	Target Interval (*A. eriantha* 'White')	Target Interval (*A. chinensis* Red5)
KCH00062	GTGGCATTACTTTCCATATTGGG	TGGGCATTGAGTTGTAACCC	CHR26:8460956-8461055	CHR26:7836781-7836880
CHR26:8193148	AGGATAGTTGGCAATTTCCAGG	TGGTAAGCCCAATAGACTATACCC	CHR26:8898197-8898278	CHR26:8206006-820608
CHR26:8874229	ACATACCATTCGGAAGCGTG	ACTGTAGGAACTGAATAGTGATCG	CHR26:9597461-9597577	CHR26:8887032-8887148
CHR26:8453577	GATAATGCGCCCACAGTTCC	GTTGAACTTTGAAGGAAACCTGC	*Not determined*	CHR26:8466420-8466503

4.4 RNASEQ and Untargetted Metabolomic Analysis

4.4.1 Sample Collection and Processing

Tissue was sampled in October 2016 between 11 am and 1 pm from young leaves (3-5cm) of AI47 and AJ47 families used for marker validation and placed in RNAlater (SigmaAldrich.com) for shipping at 4 deg C. The first fully expanded leaf from the same vine was also sampled for metabolomic analysis by taking ten 2mm discs with a biopsy punch and placing into 50 % v/v/ methanol. Metabolomic analysis is described in Appendix B. RNA was prepared using the Spectrum Plant Total RNA Kit (Sigma-Aldrich Co. LLC), and purified with the RNeasy Plant Mini Kit (Qiagen N.V.). Poly(A) RNA was isolated from 1.5ug total RNA using NEXTflex Poly(A) Beads (PerkinElmer, Inc.) Six libraries (3 high AsA, 3 low AsA) were made using the NEXTflex Rapid Directional qRNA-Seq Kit (PerkinElmer, Inc.). Samples were pooled by family and by whether they had AsA phenotype and were heterozygous for KCH0062 marker. Pools were formed as follows: Pool 1 N=3 AI247 high AsA; Pool 2 AI247 N=3 low AsA; Pool 3 AJ247 N=8 high AsA; Pool4 AJ247 N=13 low AsA; Pool5 AJ247 N=7 high AsA; Pool6 AJ247 N=10 low AsA.

Synthesis of cDNA and quantitative PCR for genes GGP, GMD , T2 and DHAR2 were performed on individual samples from pools 1,2,5 and 6 against PP2A catalyst control as reported previously [1].

RNA pools were sequenced on the Illumina HiSeq 2500 platform by Otago Genomics Facility (Dunedin, New Zealand) yielding 2.7-3.1 Mbp per library with Q30 < 89%. Merged reads were filtered for ribosomal RNA content using SortMeRNA [55] , de-interleaved and then trimmed using Trimmomatic[58] with options ILLUMINACLIP:2:30:10 SLIDINGWINDOW:5:20 MINLEN:40 HEADCROP:9.

4.4.2 RNASEQ Read Assignment

Unguided alignment to the Red5 version PS1_1.68.5 reference genome was performed using HiSat2 [57]. Alignments were split into species-specific bam files by read assignment with PolyCat [19] based on a homeo-SNP index built from the set of fixed *A.chinensis-A.eriantha* polymorphisms (Supplementary File XXX) using the script snpMerge.pl. (https://gist.github.com/jaudall/de14e367b208ccbe3b3be1465167b39b). Bambam counter [50] was used to count reads from the split bam files in Red 5 gene models.

4.4.2 RNASEQ Transcript Analysis

A *de novo* assembly was performed on trimmed reads using Trinity v 2.32 [60] yielding an assembly of 345495 transcripts in 213327 genes with contig N50 of 798 bp. Transcript abundance was estimated using RSEM [59] and differential expression analysis was performed using DESeq2 Release 3.9 [60]. Transcripts exhibiting differential expression at FDR < 0.01 were aligned to the Red 5 genome assembly using gmap [61] and intersection with annotated gene models was performed using bedtools [62]. Putative open reading frames and deduced peptides were identified with Transdecoder (https://github.com/TransDecoder) and annotated using GhostKoala [63]

Supplementary Materials:

The following are available online at https://zenodo.org/: normalized read counts, variant data files and coordinates of homeo-SNPs between *A. chinensis* and *A. eriantha* on the draft 'Red5' *A, chinensis* 10.5281/zenodo.3257749; pseudomolecules and annotations for the draft 'Red5' *A, chinensis* assembly doi: 10.5281/zenodo.1297304; scaffold files of the low-coverage *A. eriantha* assembly of genotype EA01_01 doi:10.5281/zenodo.1309031;

The following are available online at www.mdpi.com/xxx/s1 :

Table S1 Genome locations, read counts and associated annotations of *de novo* assembled transcripts mapping to kiwifruit genome assembly Red5_PS1_1.69.0 (https://www.ncbi.nlm.nih.gov/assembly/GCA_003024255.1/) which exhibited differential expression between high and low AsA pools at FDR $p < 0.05$.

Table S2 Trinity *de novo* assembly transcripts related to mannose metabolism that did not exhibit significant expression between AsA pools. Transcipts annotated as GDP-L-fucose synthase include the differentially expressed TRINITY_DN123292_c0_g2_i2.

Author Contributions: Population development and phenotyping A.S., W.L.; Study design J.M.,A.S,M.C.; Genome sequencing and analysis R.C.,E.H., S.T., S.D.,M.K.,A.C.,J.M.; Marker Design and analysis M.K.,J.M.,J.T.; RNASEQ and analysis M.C., R.M.,R.L.,J.M.; Metabolomics M.S.; Manuscript Draft J.M.; Manuscript Review and Editing S.B.,W.L.,R.M.,A.S.

Funding: Funding for this research was provided by Zespri International Ltd. and Plant and Food Research.

Acknowledgments: We thank Joshua Udall (BYU) for assistance with read assignment methods. We acknowledge technical support from the Plant and Food Research Kiwifruit Breeding Team.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data or in the writing of the manuscript

Appendix A

Table A1. Positions on the 'Red5' version PS1.68.5 assembly and allele counts of SNP loci exhibiting significant association with fruit AsA level by CMH tests with adjusted P-value < 10^{-6}. Allele counts at SNP loci were normalised by resampling to 40 reads and are in the format A:T:C:G:N:del. B denotes Ewens homozygosity information

Chr	Pos	Ref	Pool1 Counts	Pool1 B	Pool2 Counts	Pool2 B	Pool3 Counts	Pool3 B	Pool4 Counts	Pool4 B	Padj	odds ratio
CHR26	703957	A	45:59:0:0:0:0	0.3	45:52:0:0:0:0	0.3	70:14:0:0:0:0	0.2	59:0:0:0:0:0	0	1.77e-10	12.4
CHR26	703959	T	0:47:59:0:0:0	0.3	0:49:52:0:0:0	0.3	0:70:14:0:0:0	0.2	0:59:0:0:0:0	0	6.89e-10	11.8
CHR26	703960	T	54:52:0:0:0:0	0.3	49:50:0:0:0:0	0.3	12:72:0:0:0:0	0.18	0:59:0:0:0:0	0	6.07e-09	11.8
CHR26	703962	T	0:46:63:0:0:0	0.3	0:49:53:0:0:0	0.3	0:71:14:0:0:0	0.19	0:59:0:0:0:0	0	1.63e-10	12.8
CHR26	704022	G	64:0:0:58:0:0	0.3	42:0:0:45:0:0	0.3	11:0:0:65:0:0	0.18	0:0:0:38:0:0	0	8.99e-07	10.5
CHR26	704029	T	64:54:0:0:0:0	0.3	40:42:0:0:0:0	0.3	11:65:0:0:0:0	0.18	0:43:0:0:0:0	0	1.33e-07	11.5
CHR26	955933	T	0:19:22:0:0:0	0.3	0:17:18:0:0:0	0.3	0:40:2:0:0:0	0.08	0:57:0:0:0:0	0	3.44e-07	47.5
CHR26	2453803	C	0:14:20:0:0:0	0.29	0:20:19:0:0:0	0.3	0:0:50:0:0:0	0	0:0:44:0:0:0	0	2.13e-06	Inf
CHR26	2453811	A	13:0:14:0:0:0	0.3	21:0:18:0:0:0	0.3	49:0:0:0:0:0	0	45:0:0:0:0:0	0	8.99e-07	Inf
CHR26	3001514	G	19:0:0:12:0:0	0.29	15:0:0:10:0:0	0.29	0:0:0:34:0:0	0	0:0:0:27:0:0	0	5.0e-06	Inf
CHR26	5096282	A	38:24:0:0:0:0	0.29	45:22:0:0:0:0	0.28	68:2:0:0:0:0	0.06	76:0:0:0:0:0	0	8.99e-07	41.8
CHR26	5136827	A	80:0:0:35:0:0	0.27	69:0:0:36:0:0	0.28	82:0:0:8:0:0	0.13	104:0:0:0:0:0	0	2.72e-06	10.2
CHR26	6431391	A	18:0:0:15:0:0	0.3	16:0:0:23:0:0	0.29	39:0:0:0:0:0	0	37:0:0:0:0:0	0	2.74e-06	Inf
CHR26	6431400	A	17:0:0:13:0:0	0.3	12:0:0:23:0:0	0.28	37:0:0:0:0:0	0	41:0:0:0:0:0	0	5.39e-07	Inf
CHR26	6432735	G	15:0:0:12:0:0	0.3	21:0:0:15:0:0	0.3	2:0:0:40:0:0	0.08	0:0:0:37:0:0	0	3.28e-06	55.6
CHR26	6433469	G	33:0:0:32:0:0	0.3	21:0:0:49:0:0	0.27	2:0:0:48:0:0	0.07	0:0:0:74:0:0	0	1.19e-06	44.1
CHR26	6516587	C	0:22:19:0:0:0	0.3	0:29:15:0:0:0	0.28	0:4:44:0:0:0	0.13	0:0:39:0:0:0	0	2.03e-07	28.7
CHR26	7647158	T	0:12:13:0:0:0	0.3	0:19:18:0:0:0	0.3	0:43:0:0:0:0	0	0:45:0:0:0:0	0	1.69e-06	Inf
CHR26	7647167	T	0:12:13:0:0:0	0.3	0:22:20:0:0:0	0.3	0:38:0:0:0:0	0	0:45:0:0:0:0	0	4.35e-06	Inf
CHR26	8214875	T	0:39:0:0:0:0	0	0:37:6:0:0:0	0.18	0:16:21:0:0:0	0.3	0:18:30:0:0:0	0.29	4.92e-06	19.4
CHR26	8414723	C	11:0:16:0:0:0	0.29	20:0:14:0:0:0	0.29	0:0:40:0:0:0	0	0:0:40:0:0:0	0	5.77e-06	Inf
CHR26	8571426	G	20:0:0:12:0:0	0.29	17:0:0:14:0:0	0.3	0:0:0:34:0:0	0	0:0:0:33:0:0	0	1.73e-06	Inf
CHR26	9081448	G	0:16:0:13:0:0	0.3	0:16:0:22:0:0	0.3	0:1:0:46:0:0	0.05	0:0:0:47:0:0	0	3.28e-06	108.3
CHR26	9263130	G	18:0:0:18:0:0	0.3	14:0:0:18:0:0	0.3	39:0:0:0:0:0	0	35:0:0:0:0:0	0	4.2e-06	Inf
CHR26	9436328	A	18:0:0:18:0:0	0.3	12:0:0:28:0:0	0.27	32:0:0:2:0:0	0.1	43:0:0:0:0:0	0	2.18e-07	44.2

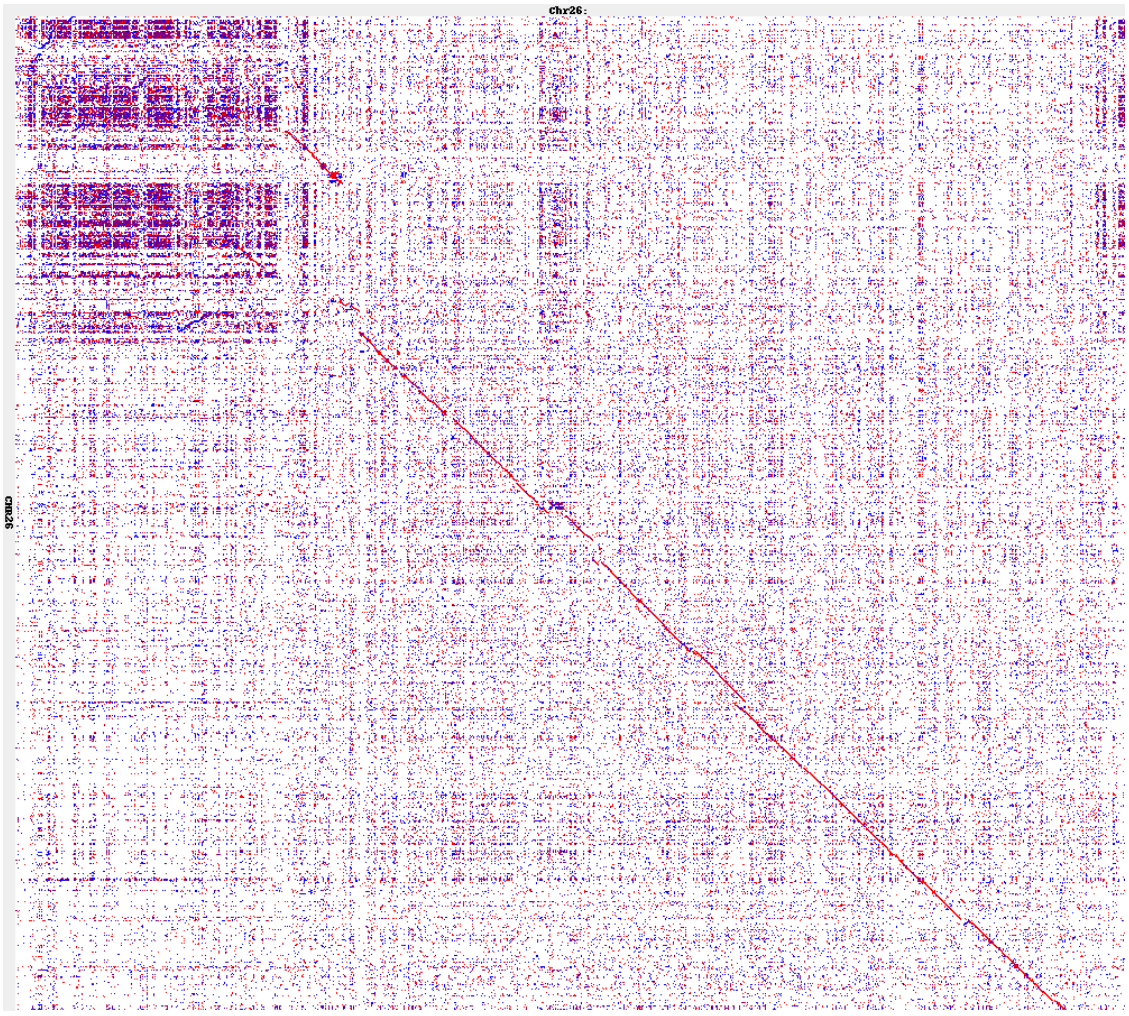

Figure A1. LAST dotplot alignment of chromosome 26 pseudomolecules of *A. chinensis var chinensis* 'Red5' (y-axis) with *A. eriantha* 'White' (x-axis)

Appendix B Untargetted Metabolomics Analysis Methodology

LC-MS Data Acquisition

LCMS system

The system consisted of a Thermo Scientific™ (San Jose, CA, USA) Q Exactive™ Plus Orbitrap coupled with a Vanquish™ UHPLC system (Binary Pump H, Split Sampler HT, Dual Oven). Calibrations were performed immediately prior to sample analysis batch with Thermo™ premixed solutions (Pierce™ LTQ ESI Positive and negative ion calibration solutions, catalogue numbers: 88322 and 88324 respectively).

Aqueous normal phase conditions

A 2 µL aliquot of each prepared extract was separated with a mobile phase consisting of 0.1 % formic acid in acetonitrile (A) and 5mM ammonium acetate in water (B) by normal phase chromatography (hypersil Gold HILIC 1.9µm, 100mm x2.1mm, P/N:26502-102130) maintained at 55 °C with a flow rate of 400 µl/min. A gradient was applied: 0-1 min/5%B, linear increase to 12 min/98%B, isocratic 16min/98%B, equilibration 16-17 min/5%B, isocratic to end 20min/5%B.

Reverse phase conditions

A 2 µL aliquot of each prepared extract was separated with a mobile phase consisting of 0.1 % formic acid in type 1 water (A) and 0.1 % formic acid in acetonitrile (B) by reverse phase chromatography (Accucore Vanquish C18 1.5µm, 100mm x2.1mm, P/N: 27101-102130, Thermo Scientific) maintained at 40°C with a flow rate of 400µl/min. A gradient was applied: 0-1 min/0%B, linear increase to 7 min/50%B, linear increase to 8min/98%B, isocratic to 11 min/98%B, equilibration 11-12min/0%B, isocratic to end 17 min/0%B.

The eluent from (H) and (C18) chromatography was scanned from 0.5-16 and 0.4-11.5 minutes respectively by API-MS (Orbitrap) with heated electrospray ionisation (HESI) at 350°C in the negative and positive mode with capillary temperature of 320°C. Data were acquired for precursor masses from m/z 80–1200 amu (H) and m/z 100-1500(C18) at 70K resolution (AGC target 3e6, maximum IT 100ms, profile mode) with data dependent ms/ms for product ions generated by normalised collision energy (NCE:35, 45, 65) at 17.5K resolution (TopN 10, AGC target 2e5, Maximum IT 50ms, Isolation 1.4 m/z).

Samples were grouped based on family and KHC00062 marker genotype, and additionally each was subsampled to create a sample mix of each group and solution blanks. Samples were analysed by four analytical methods (2 columns, aqueous reverse phase(C18) and aqueous normal phase (Helic) with two ionisation modes -, + (n or p)) creating four datasets (Cn, Cp, Hn, Hp).

Data processing

Data were processed with the aid of Xcalibur®4.1 and Compound Discoverer 3.0 (Thermo Electron Corporation).Calculated exact molecular weights generated from m/z ions and spectra features (isotope ratios, precursor and product fragment ions) were utilised to predict compound formula, targeted search lists, internal and published library spectra and known compounds/metabolites from selected parameters. Differential analysis was applied based on grouping sets to filter compounds of interest. Significant features were manually interpreted or confirmed with reference to theoretical spectra features and or literature from SciFinder™ with chemistry associated with keywords or in combination with chemical classes/structure search of interest.

References

1. Bulley, S. M.; Rassam, M.; Hoser, D.; Otto, W.; Schünemann, N.; Wright, M.; MacRae, E.; Gleave, A.; Laing, W. Gene expression studies in kiwifruit and gene over-expression in Arabidopsis indicates that GDP-L-galactose guanyltransferase is a major control point of vitamin C biosynthesis. *Journal of Experimental Botany* **2009**, *60*, 765–778, doi:10.1093/jxb/ern327.
2. Jiang, Z.-Y.; Zhong, Y.; Zheng, J.; Ali, M.; Liu, G.-D.; Zheng, X.-L. L-ascorbic acid metabolism in an ascorbate-rich kiwifruit (Actinidia. Eriantha Benth.) cv. 'White' during postharvest. *Plant Physiology and Biochemistry* **2018**, *124*, 20–28, doi:10.1016/j.plaphy.2018.01.005.
3. Tang, W.; Sun, X.; Yue, J.; Tang, X.; Jiao, C.; Yang, Y.; Niu, X.; Miao, M.; Zhang, D.; Huang, S.; Shi, W.; Li, M.; Fang, C.; Fei, Z.; Liu, Y. Chromosome-scale genome assembly of kiwifruit Actinidia eriantha with single-molecule sequencing and chromatin interaction mapping.. *Gigascience* **2019**, *8*.
4. Pilkington, S. M.; Crowhurst, R.; Hilario, E.; Nardozza, S.; Fraser, L.; Peng, Y.; Gunaseelan, K.; Simpson, R.; Tahir, J.; Deroles, S. C.; Templeton, K.; Luo, Z.; Davy, M.; Cheng, C.; McNeilage, M.; Scaglione, D.; Liu, Y.; Zhang, Q.; Datson, P.; De Silva, N.; Gardiner, S. E.; Bassett, H.; Chagné, D.; McCallum, J.; Dzierzon, H.; Deng, C.; Wang, Y.-Y.; Barron, L.; Manako, K.; Bowen, J.; Foster, T. M.; Erridge, Z. A.; Tiffin, H.; Waite, C. N.; Davies, K. M.; Grierson, E. P.; Laing, W. A.; Kirk, R.; Chen, X.; Wood, M.; Montefiori, M.; Brummell, D. A.; Schwinn, K. E.; Catanach, A.; Fullerton, C.; Li, D.; Meiyalaghan, S.; Nieuwenhuizen, N.; Read, N.; Prakash, R.; Hunter, D.; Zhang, H.; McKenzie, M.; Knäbel, M.; Harris, A.; Allan, A. C.; Gleave, A.; Chen, A.; Janssen, B. J.; Plunkett, B.; Ampomah-Dwamena, C.; Voogd, C.; Leif, D.; Lafferty, D.; Souleyre, E. J. F.; Varkonyi-Gasic, E.; Gambi, F.; Hanley, J.; Yao, J.-L.; Cheung, J.; David, K. M.; Warren, B.; Marsh, K.;

Snowden, K. C.; Lin-Wang, K.; Brian, L.; Martinez-Sanchez, M.; Wang, M.; Ileperuma, N.; Macnee, N.; Campin, R.; McAtee, P.; Drummond, R. S. M.; Espley, R. V.; Ireland, H. S.; Wu, R.; Atkinson, R. G.; Karunairetnam, S.; Bulley, S.; Chunkath, S.; Hanley, Z.; Storey, R.; Thrimawithana, A. H.; Thomson, S.; David, C.; Testolin, R.; Huang, H.; Hellens, R. P.; Schaffer, R. J. A manually annotated Actinidia chinensis var. chinensis (kiwifruit) genome highlights the challenges associated with draft genomes and gene prediction in plants. *BMC Genomics* **2018**, *19*, 257, doi:10.1186/s12864-018-4656-3.

5. Huang, S.; Ding, J.; Deng, D.; Tang, W.; Sun, H.; Liu, D.; Zhang, L.; Niu, X.; Zhang, X.; Meng, M.; Yu, J.; Liu, J.; Han, Y.; Shi, W.; Zhang, D.; Cao, S.; Wei, Z.; Cui, Y.; Xia, Y.; Zeng, H.; Bao, K.; Lin, L.; Min, Y.; Zhang, H.; Miao, M.; Tang, X.; Zhu, Y.; Sui, Y.; Li, G.; Sun, H.; Yue, J.; Sun, J.; Liu, F.; Zhou, L.; Lei, L.; Zheng, X.; Liu, M.; Huang, L.; Song, J.; Xu, C.; Li, J.; Ye, K.; Zhong, S.; Lu, B.-R.; He, G.; Xiao, F.; Wang, H.-L.; Zheng, H.; Fei, Z.; Liu, Y. Draft genome of the kiwifruit Actinidia chinensis. *Nature Communications* **2013**, *4*, doi:10.1038/ncomms3640.
6. Bulley, S. M.; Laing, W. Ascorbic Acid-Related Genes. In *Compendium of Plant Genomes*; Springer International Publishing, 2016; pp. 163–177.
7. Bulley, S.; Laing, W. The regulation of ascorbate biosynthesis. *Current Opinion in Plant Biology* **2016**, *33*, 15–22, doi:10.1016/j.pbi.2016.04.010.
8. Bulley, S.; Wright, M.; Rommens, C.; Yan, H.; Rassam, M.; Lin-Wang, K.; Andre, C.; Brewster, D.; Karunairetnam, S.; Allan, A. C.; Laing, W. A. Enhancing ascorbate in fruits and tubers through over-expression of the l-galactose pathway gene GDP-l-galactose phosphorylase. *Plant Biotechnology Journal* **2012**, *10*, 390–397, doi:10.1111/j.1467-7652.2011.00668.x.
9. Laing, W. A.; Martínez-Sánchez, M.; Wright, M. A.; Bulley, S. M.; Brewster, D.; Dare, A. P.; Rassam, M.; Wang, D.; Storey, R.; Macknight, R. C.; Hellens, R. P. An Upstream Open Reading Frame Is Essential for Feedback Regulation of Ascorbate Biosynthesis in Arabidopsis. *The Plant Cell* **2015**, *27*, 772–786, doi:10.1105/tpc.114.133777.
10. Mellidou, I.; Chagne, D.; Laing, W. A.; Keulemans, J.; Davey, M. W. Allelic Variation in Paralogs of GDP-l-Galactose Phosphorylase Is a Major Determinant of Vitamin C Concentrations in Apple Fruit. *PLANT PHYSIOLOGY* **2012**, *160*, 1613–1629, doi:10.1104/pp.112.203786.
11. Truffault, V.; Gest, N.; Garchery, C.; Causse, M.; Duboscq, R.; Riqueau, G.; Sauvage, C.; Gautier, H.; Baldet, P.; Stevens, R. VARIATION IN TOMATO FRUIT ASCORBATE LEVELS AND CONSEQUENCES OF MANIPULATION OF ASCORBATE METABOLISM ON DROUGHT STRESS TOLERANCE. *Acta Horticulturae* **2014**, 75–84, doi:10.17660/actahortic.2014.1048.8.
12. Huang, S.; Ding, J.; Deng, D.; Tang, W.; Sun, H.; Liu, D.; Zhang, L.; Niu, X.; Zhang, X.; Meng, M.; Yu, J.; Liu, J.; Han, Y.; Shi, W.; Zhang, D.; Cao, S.; Wei, Z.; Cui, Y.; Xia, Y.; Zeng, H.; Bao, K.; Lin, L.; Min, Y.; Zhang, H.; Miao, M.; Tang, X.; Zhu, Y.; Sui, Y.; Li, G.; Sun, H.; Yue, J.; Sun, J.; Liu, F.; Zhou, L.; Lei, L.; Zheng, X.; Liu, M.; Huang, L.; Song, J.; Xu, C.; Li, J.; Ye, K.; Zhong, S.; Lu, B.-R.; He, G.; Xiao, F.; Wang, H.-L.; Zheng, H.; Fei, Z.; Liu, Y. Draft genome of the kiwifruit Actinidia chinensis. *Nature Communications* **2013**, *4*, doi:10.1038/ncomms3640.
13. Kofler, R.; Pandey, R. V.; Schlotterer, C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **2011**, *27*, 3435–3436, doi:10.1093/bioinformatics/btr589.
14. Mansfeld, B. N.; Grumet, R. QTLseqr: An R Package for Bulk Segregant Analysis with Next-Generation Sequencing. *The Plant Genome* **2018**, *11*, 0, doi:10.3835/plantgenome2018.01.0006.
15. Fraser, L. G.; Seal, A. G.; Montefiori, M.; McGhie, T. K.; Tsang, G. K.; Datson, P. M.; Hilario, E.; Marsh, H. E.; Dunn, J. K.; Hellens, R. P.; Davies, K. M.; McNeilage, M. A.; Silva, H. N. D.; Allan, A. C. An R2R3 MYB transcription factor determines red petal colour in an Actinidia (kiwifruit) hybrid population. *BMC Genomics* **2013**, *14*, 28, doi:10.1186/1471-2164-14-28.
16. Tahir, J.; Gardiner, S. E.; Hoyte, S.; Bassett, H.; Brendolise, C.; Chatterjee, A.; Templeton, K.; Deng, C.; Crowhurst, R.; Montefiori, M.; Morgan, E.; Wotton, A.; Funnell, K.; Wiedow, C.; Knaebel, M.; Hedderley, D.; Vanneste, J.; McCallum, J.; Hoeata, K.; Chagne, D.; Gea, L. Multiple quantitative trait loci contribute tolerance to bacterial canker incited by Pseudomonas syringae pv. actinidiae in kiwifruit (Actinidia chinensis). **2019**, doi:10.1101/526798.
17. Zhang, L.; Li, Z.; Wang, Y.; Jiang, Z.; Wang, S.; Huang, H. Vitamin C flower color and ploidy variation of hybrids from a ploidy-unbalanced Actinidia interspecific cross and SSR characterization. *Euphytica* **2010**, *175*, 133–143, doi:10.1007/s10681-010-0194-z.

18. Page, J. T.; Gingle, A. R.; Udall, J. A. PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms.. *G3 (Bethesda)* **2013**, *3*, 517–25.
19. Bonin, C. P.; Reiter, W.-D. A bifunctional epimerase-reductase acts downstream of the MUR1 gene product and completes the de novo synthesis of GDP-L-fucose in Arabidopsis. *The Plant Journal* **2000**, *21*, 445–454, doi:10.1046/j.1365-313x.2000.00698.x.
20. Jadid, N.; Mialoundama, A. S.; Heintz, D.; Ayoub, D.; Erhardt, M.; Mutterer, J.; Meyer, D.; Alioua, A.; Dorsselaer, A. V.; Rahier, A.; Camara, B.; Bouvier, F. DOLICHOL PHOSPHATE MANNOSE SYNTHASE1 Mediates the Biogenesis of Isoprenyl-Linked Glycans and Influences Development Stress Response, and Ammonium Hypersensitivity inArabidopsis. *The Plant Cell* **2011**, *23*, 1985–2005, doi:10.1105/tpc.111.083634.
21. Rigano, M. M.; Lionetti, V.; Raiola, A.; Bellincampi, D.; Barone, A. Pectic enzymes as potential enhancers of ascorbic acid production through the D -galacturonate pathway in Solanaceae. *Plant Science* **2018**, *266*, 55–63, doi:10.1016/j.plantsci.2017.10.013.
22. Gao, C.; Ju, Z.; Li, S.; Zuo, J.; Fu, D.; Tian, H.; Luo, Y.; Zhu, B. Deciphering Ascorbic Acid Regulatory Pathways in Ripening Tomato Fruit Using a Weighted Gene Correlation Network Analysis Approach. *Journal of Integrative Plant Biology* **2013**, *55*, 1080–1091, doi:10.1111/jipb.12079.
23. Page, M.; Sultana, N.; Paszkiewicz, K.; Florance, H.; Smirnoff, N. The influence of ascorbate on anthocyanin accumulation during high light acclimation in Arabidopsis thaliana: further evidence for redox control of anthocyanin synthesis.. *Plant Cell Environ* **2012**, *35*, 388–404.
24. Scortichini, M.; Marcelletti, S.; Ferrante, P.; Petriccione, M.; Firrao, G. Pseudomonas syringae pv. actinidiae: a re-emerging, multi-faceted, pandemic pathogen.. *Mol Plant Pathol* **2012**, *13*, 631–40.
25. Zhang, H.; Yi, H.; Wu, M.; Zhang, Y.; Zhang, X.; Li, M.; Wang, G. Mapping the Flavor Contributing Traits on Fengwei Melon (Cucumis melo L.) Chromosomes Using Parent Resequencing and Super Bulked-Segregant Analysis. *PLOS ONE* **2016**, *11*, e0148150, doi:10.1371/journal.pone.0148150.
26. Wei, Q.-zhen; Fu, W.-yuan; Wang, Y.-zhu; Qin, X.-dong; Wang, J.; Li, J.; Lou, Q.-feng; Chen, J.-feng Rapid identification of fruit length loci in cucumber (Cucumis sativus L.) using next-generation sequencing (NGS)-based QTL analysis. *Scientific Reports* **2016**, *6*, doi:10.1038/srep27496.
27. Love, M. I.; Soneson, C.; Patro, R. Swimming downstream: statistical analysis of differential transcript usage following Salmon quantification. *F1000Research* **2018**, *7*, 952, doi:10.12688/f1000research.15398.2.
28. Davey, M. W. Genetic Control of Fruit Vitamin C Contents. *PLANT PHYSIOLOGY* **2006**, *142*, 343–351, doi:10.1104/pp.106.083279.
29. Stevens, R.; Buret, M.; Duffe, P.; Garchery, C.; Baldet, P.; Rothan, C.; Causse, M. Candidate Genes and Quantitative Trait Loci Affecting Fruit Ascorbic Acid Content in Three Tomato Populations. *PLANT PHYSIOLOGY* **2007**, *143*, 1943–1953, doi:10.1104/pp.106.091413.
30. 31. Thompson, M. J.; Jiggins, C. D. Supergenes and their role in evolution. *Heredity* **2014**, *113*, 1–8, doi:10.1038/hdy.2014.20.
31. 32. Schwander, T.; Libbrecht, R.; Keller, L. Supergenes and Complex Phenotypes. *Current Biology* **2014**, *24*, R288–R294, doi:10.1016/j.cub.2014.01.056.
32. 33. Pilkington, S. M.; Tahir, J.; Hilario, E.; Gardiner, S. E.; Chagné, D.; Catanach, A.; McCallum, J.; Jesson, L.; Fraser, L. G.; McNeilage, M. A.; Deng, C.; Crowhurst, R. N.; Datson, P. M.; Zhang, Q. Genetic and cytological analyses reveal the recombination landscape of a partially differentiated plant sex chromosome in kiwifruit.. *BMC Plant Biol* **2019**, *19*, 172.
33. Ferguson, A. R. Botanical Description. In *Compendium of Plant Genomes*; Springer International Publishing, 2016; pp. 1–13.
34. Gest, N.; Gautier, H.; Stevens, R. Ascorbate as seen through plant evolution: the rise of a successful molecule?. *Journal of Experimental Botany* **2012**, *64*, 33–53, doi:10.1093/jxb/ers297.
35. Nijhout, H. F. Developmental Perspectives on Evolution of Butterfly Mimicry. *BioScience* **1994**, *44*, 148–157, doi:10.2307/1312251.
36. Locato, V.; Cimini, S.; Gara, L. D. Strategies to increase vitamin C in plants: from plant defense perspective to food biofortification. *Frontiers in Plant Science* **2013**, *4*, doi:10.3389/fpls.2013.00152.
37. Macknight, R. C.; Laing, W. A.; Bulley, S. M.; Broad, R. C.; Johnson, A. A. T.; Hellens, R. P. Increasing ascorbate levels in crops to enhance human nutrition and plant abiotic stress tolerance. *Current Opinion in Biotechnology* **2017**, *44*, 153–160, doi:10.1016/j.copbio.2017.01.011.
38. Dellaporta, S. Plant DNA Miniprep and Microprep: Versions 2.12.3. In *The Maize Handbook*; Springer New York, 1994; pp. 522–525.

39. Yao, X.; Tang, P.; Li, Z.; Li, D.; Liu, Y.; Huang, H. The First Complete Chloroplast Genome Sequences in Actinidiaceae: Genome Structure and Comparative Analysis. *PLOS ONE* **2015**, *10*, e0129347, doi:10.1371/journal.pone.0129347.
40. Tahir, J.; Gardiner, S.; Hoyte, S.; Bassett, H.; Brendolise, C.; Chatterjee, A.; Templeton, K.; Deng, C.; Crowhurst, R.; Montefiori, M.; others Multiple quantitative trait loci contribute tolerance to bacterial canker incited by Pseudomonas syringae pv. actinidiae in kiwifruit (Actinidia chinensis). *BioRxiv* **2019**, 526798.
41. Auwera, G. A. Van der; Carneiro, M. O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; others From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics* **2013**, 11–10.
42. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760, doi:10.1093/bioinformatics/btp324.
43. Crowhurst, R.; Liu, Y.; Scaglione, D. The Kiwifruit Genome. In *The Kiwifruit Genome*; Springer, 2016; pp. 101–114.
44. Kielbasa, S. M.; Wan, R.; Sato, K.; Horton, P.; Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Research* **2011**, *21*, 487–493, doi:10.1101/gr.113985.110.
45. Morgulis, A.; Gertz, E. M.; Schaffer, A. A.; Agarwala, R. WindowMasker: window-based masker for sequenced genomes. *Bioinformatics* **2005**, *22*, 134–141, doi:10.1093/bioinformatics/bti774.
46. Ellinghaus, D.; Kurtz, S.; Willhoeft, U. LTRharvest an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **2008**, *9*, 18, doi:10.1186/1471-2105-9-18.
47. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; and, R. D. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079, doi:10.1093/bioinformatics/btp352.
48. Koboldt, D. C.; Zhang, Q.; Larson, D. E.; Shen, D.; McLellan, M. D.; Lin, L.; Miller, C. A.; Mardis, E. R.; Ding, L.; Wilson, R. K. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* **2012**, *22*, 568–576, doi:10.1101/gr.129684.111.
49. Takagi, H.; Abe, A.; Yoshida, K.; Kosugi, S.; Natsume, S.; Mitsuoka, C.; Uemura, A.; Utsushi, H.; Tamiru, M.; Takuno, S.; Innan, H.; Cano, L. M.; Kamoun, S.; Terauchi, R. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal* **2013**, *74*, 174–183, doi:10.1111/tpj.12105.
50. Page, J. T.; Liechty, Z. S.; Huynh, M. D.; Udall, J. A. BamBam: genome sequence analysis tools for biologists.. *BMC Res Notes* **2014**, *7*, 829.
51. Benjamini, Y.; Hochberg, Y. Multiple Hypotheses Testing with Weights. *Scandinavian Journal of Statistics* **1997**, *24*, 407–418, doi:10.1111/1467-9469.00072.
52. Baldwin, S.; Revanna, R.; Thomson, S.; Pither-Joyce, M.; Wright, K.; Crowhurst, R.; Fiers, M.; Chen, L.; Macknight, R.; McCallum, J. A. A Toolkit for bulk PCR-based marker design from next-generation sequence data: application for development of a framework linkage map in bulb onion (Allium cepa L.). *BMC Genomics* **2012**, *13*, 637, doi:10.1186/1471-2164-13-637.
53. Butler, J.; MacCallum, I.; Kleber, M.; Shlyakhter, I. A.; Belmonte, M. K.; Lander, E. S.; Nusbaum, C.; Jaffe, D. B. ALLPATHS: De novo assembly of whole-genome shotgun microreads. *Genome Research* **2008**, *18*, 810–820, doi:10.1101/gr.7337908.
54. Boetzer, M.; Henkel, C. V.; Jansen, H. J.; Butler, D.; Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **2010**, *27*, 578–579, doi:10.1093/bioinformatics/btq683.
55. Kopylova, E.; Noé, L.; Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **2012**, *28*, 3211–3217, doi:10.1093/bioinformatics/bts611.
56. Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120, doi:10.1093/bioinformatics/btu170.
57. Kim, D.; Langmead, B.; Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements.. *Nat Methods* **2015**, *12*, 357–60.
58. Haas, B. J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P. D.; Bowden, J.; Couger, M. B.; Eccles, D.; Li, B.; Lieber, M.; MacManes, M. D.; Ott, M.; Orvis, J.; Pochet, N.; Strozzi, F.; Weeks, N.; Westerman, R.; William, T.; Dewey, C. N.; Henschel, R.; LeDuc, R. D.; Friedman, N.; Regev, A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.. *Nat Protoc* **2013**, *8*, 1494–512.

59. Li, B.; Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **2011**, *12*, doi:10.1186/1471-2105-12-323.
60. Love, M. I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **2014**, *15*, doi:10.1186/s13059-014-0550-8.
61. Wu, T. D.; Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **2005**, *21*, 1859–1875, doi:10.1093/bioinformatics/bti310.
62. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current Protocols in Bioinformatics* **2014**, *47*, 11.12.1–11.12.34, doi:10.1002/0471250953.bi1112s47.
63. Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences.. *J Mol Biol* **2016**, *428*, 726–731.
64. Ewens, W. J. The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **1972**, *3*, 87–112, doi:10.1016/0040-5809(72)90035-4.