Genome-Wide Association Analysis of the Anthocyanin and Carotenoid Contents of Rose Petals

Dietmar F. Schulz¹, Rena T. Schott†¹, Roeland E. Voorrips², Marinus J. M. Smulders², Marcus Linde¹ and Thomas Debener*¹

¹Abteilung Molekulare Pflanzenzüchtung, Institute for Plant Genetics, Leibnitz University Hannover, Hannover, Germany,
²Wageningen University and Research Plant Breeding, Wageningen University and Research Centre, Wageningen, Netherlands

Petal color is one of the key characteristics determining the attractiveness and therefore the commercial value of an ornamental crop. Here, we present the first genome-wide association study for the important ornamental crop rose, focusing on the anthocyanin and carotenoid contents in petals of 96 diverse tetraploid garden rose genotypes. Cultivated roses display a vast phenotypic and genetic diversity and are therefore ideal targets for association genetics. For marker analysis, we used a recently designed Axiom SNP chip comprising 68,000 SNPs with additionally 281 SSRs, 400 AFLPs and 246 markers from candidate genes. An analysis of the structure of the rose population revealed three subpopulations with most of the genetic variation between individual genotypes rather than between clusters and with a high average proportion of heterozygous loci. The mapping of markers significantly associated with anthocyanin and carotenoid content to the related Fragaria and Prunus genomes revealed clusters of associated markers indicating five genomic regions associated with the total anthocyanin content and two large clusters associated with the carotenoid content. Among the marker clusters associated with the phenotypes, we found several candidate genes with known functions in either the anthocyanin or the carotenoid biosynthesis pathways. Among others, we identified a glutathione-S-transferase, 4CL, an auxin response factor and F3’H as candidate genes affecting anthocyanin concentration, and CCD4 and Zeaxanthine epoxidase as candidates affecting the concentration of carotenoids. These markers are starting points for future validation experiments in independent populations as well as for functional genomic studies to identify the causal factors for the observed color phenotypes. Furthermore, validated markers may be interesting tools for marker-assisted selection in commercial breeding programmes in that they provide the tools to identify superior parental combinations that combine several associated markers in higher dosages.

Keywords: petal color, anthocyanin, carotenoid, genome wide association study, tetraploid roses
INTRODUCTION

Rose is one of the most economically important ornamental crops and is sold as cut flowers, pot roses and garden roses. The genus *Rosa* comprises a vast amount of genetic resources represented by more than 100 wild species as well as more than 30,000 mostly tetraploid varieties bred for ornamental purposes (Gudin, 2000; Wissemann, 2003). Cultivated tetraploid rose genomes are complex mixtures of at least 10 species that have been used in ornamental rose breeding for more than two centuries (Gudin, 2000; Zhang et al., 2013). As a result, rose is highly diverse in many morphological and physiological characteristics. Despite its commercial importance as an ornamental plant, genomic resources for rose research and breeding remain scarce, and to date, no genome sequence is available. At the diploid level, genetic maps have been constructed, and a number of monogenic and quantitative traits have been localized on these maps (Debener and Linde, 2009; Spiller et al., 2011). However, as most diploid populations have derived from a few diploid genotypes, genetic variability is low for most horticultural traits. Therefore, these traits can only be analyzed at the tetraploid level. In tetraploid varieties, several monogenic traits have been analyzed, but only few QTL have been described, mostly by analyses of biparental populations (Debener and Linde, 2009; Spiller et al., 2011).

The esthetic features of the rose flower are of central importance for the ornamental quality of rose cultivars; therefore, commercial breeding pays special attention to floral characteristics. Flower traits, e.g., the number and color of petals, were among the first traits investigated in genetic studies (De Vries and Dubois, 1978; Debener, 1999, 2003).

The anthocyanin concentration in cells of rose petals is a major determinant of red and pink color variants, although the final hues are influenced by several other factors as e.g., pH, copigments, metal ions, types of glycosylation, etc. (Jay et al., 2003; Grotewold, 2006; Tanaka et al., 2008). Although single loci influencing color variation have been identified, other researchers have described a quantitative inheritance of the anthocyanin content (Cardoso et al., 2012; Cericola et al., 2014; Henz et al., 2015).

The levels of carotenoids, which produce yellow colors, are influenced by biotic and abiotic factors, including the developmental stage, the environment and stress (Eugster and Märki-Fischer, 1991; Deli et al., 1998; Kishimoto et al., 2004). In fully opened flowers of Ipomoea, the chloroplast-type carotenoids are β-cryptoxanthin, zeaxanthin, and β-carotene, whereas lutein, violaxanthin and β-carotene are predominant in the early stage of petal development, and the same compounds were found in the leaves (Yamamizo et al., 2010).

QTL Mapping and GWAS

All of the QTLs studied to date in rose have been mapped in biparental populations (Crespel et al., 2002; Linde et al., 2006; Spiller et al., 2011; Moghaddam et al., 2012; Roman et al., 2015) using AFLP and microsatellite markers. Tetraploid populations derived from crosses between ornamental varieties display complex patterns of inheritance that complicate not only genetic analysis but also map construction and require many more markers (Bourke et al., 2015).

Association studies offer two main advantages over QTL studies based on biparental populations: a larger number of alleles per locus and a higher resolution of trait-marker associations due to a higher rate of recombination. In association genetics, genotyping can be restricted to candidate genes likely involved in the expression of the traits under study or markers covering the whole genome in genome-wide association studies (GWAS). Few studies cover polyploids, and very few GWAS have been performed on highly heterozygous polyploids, such as potato (D’hoop et al., 2010; Lindqvist-Kreuze et al., 2014), switchgrass (Lu et al., 2013) and cotton (Abdurakhmonov et al., 2009). Rose is an interesting ornamental crop for association studies because its cultivars are extremely polymorphic, and many traits can be studied simultaneously in populations of moderate size.

Recently, an analysis of a large collection of rose ESTs and the development of an Axiom SNP array was described (Koning-Boucoiran et al., 2015). These resources are a significant extension of the genomic resources available for roses because they now permit the highly reproducible genotyping of rose genomes with approximately 68,000 SNP markers each represented by two probes. Hence, sufficient numbers of markers are now available for GWAS in tetraploid rose.

Aims of the Present Study

The aim of the present study was to exploit the enormous biodiversity of cultivated roses in flower-related traits for an analysis of the underlying genetic factors, focusing on the contents of anthocyanins and carotenoids, which are the main components of the rose petal color. This was accomplished using a combination of association genetics methods and markers on the SNP array and additional markers derived from candidate genes, SSRs and AFLPs. In addition, we tried to gain information about the variability of the genetic diversity and heterozygosity within our association panel.

MATERIALS AND METHODS

Plant Material

An association panel of 96 rose cultivars with code numbers from 1 to 141 (87 tetraploid, 8 triploid, and 1 diploid) was used for the present study. Most of the cultivars were commercially available or provided by German rose-breeding companies (Table S1). Based on known pedigrees, we attempted to minimize genotypic relatedness, which can result in spurious associations, while capturing the vast diversity of phenotypic traits, including different flower colors, plant architectures, etc. Clones of each cultivar, grafted on *R. corymbifera* “Laxa,” were planted in three randomized blocks in a field at Hannover-Herrenhausen (Germany) in the spring of 2012. A second collection of cultivars was maintained under semi-controlled conditions as potted plants in three randomized blocks in a greenhouse (Federal Plant Variety Office, Hannover). The plants were initially cultivated in 3-l pots and then transferred to 7-l pots with the fertilized substrate Einheitserde T (Einheitserdewerke...
Gebr. Patzer, Sinntal-Altengronau, D) under natural light with a day and night temperature of 22°C ± 5°C.

**Anthocyanin Content of Petals**

Flowers were always sampled from 8 to 12 a.m. Opened buds at flower development stage 3 (Picone et al., 2004) were selected from each genotype and kept on ice until sample preparation on the same day. The anthocyanin content of petals was estimated according to Henz et al. (2015) with minor modifications. Three replicates (each 50 mg in fresh weight) from petals of each clone (3 biological replicates) were placed in 2 ml test tubes and extracted in 1 ml of methanol/HCl (99:1 v/v) (Figure S7). Following an overnight incubation (16 h) in the dark at 18°C, the total anthocyanin content in the solvent was determined based on the absorbance at 525 nm using a UV-Vis-Photometer UV mcSAFAS (Deelux Labortechnik GmbH, Germany). If necessary (ΔE525nm > 1.0), the anthocyanin extracts were diluted with the extraction solvent. Each clone was measured three times, and the overall mean was calculated for each cultivar. The anthocyanin content was recorded and evaluated in two environments: (i) in the field at Herrenhausen and (ii) in the greenhouse at the Federal Plant Variety Office, Hannover. The absorbance values were not used to calculate the levels of compounds, as this measurement involves variable mixtures of anthocyanidins.

**Carotenoid Content of Petals**

The content of carotenoids was evaluated from all cultivars cultivated in the greenhouse at the Federal Plant Variety Office Hannover and from 20 of the cultivars in the field at Herrenhausen. The accumulation of total carotenoids in rose petals was estimated according to de Vries et al. (1974) with modification. Petals (50 mg each) were extracted with 1 ml of a mixture of petroleum ether:acetone (3:2 v/v) for 4 h at 18°C in the dark, and the carotenoids in the samples were measured spectrophotometrically at a wavelength of 442 nm. The extracts showed three characteristic absorption maxima for carotenoids at 419, 442, and 471 nm (Figure S1). These maxima suggest Violaxanthin (420, 443, and 471 nm) or Neoxanthin (420, 442, and 473 nm) as components of the extract in addition to other possible compounds, respectively (Wellburn, 1994; Tinoi et al., 2006). The overlap in absorption at 442 nm is the reason that we did not determine the levels of a particular compound.

**DNA Extraction**

DNA was extracted from young rose leaves as described by Klie et al. (2013). The quality of DNA was checked on agarose gels, and quantification was performed using a Nanodrop 2000c spectrophotometer (PeQLab Biotecnologie GmbH, Erlangen, Germany).

**Microsatellites (SSRs)**

PCR was performed in 15 µl of Williams buffer (Williams et al., 1990) containing 0.2 mM dNTP, 0.5 µM forward and reverse primers, and 1 U of DCS Taq DNA polymerase (Enzymatics, Beverly, USA). PCR conditions were as follows: initial denaturation for 5 min. at 94°C, 28 cycles of 45 s at 94°C, 45 s at 50, 55 or 60°C (Table S2), 60 s at 72°C. SSR marker bands were visually inspected and dominantly scored, and the data were transferred to a 0/1 matrix.

**AFLP**

AFLP markers were generated according to the protocol of Klie et al. (2013) with 250 ng of genomic DNA. We tested 21 HindIII and MseI primer pair combinations in the end reaction. Bands were scored dominantly and recorded in a 0/1 matrix for absence/presence of marker fragments.

**SNPs**

SNPs were analyzed using the Axiom WagRhSNP array, which contains 89,893 SNPs derived from cut roses and from garden roses (Koning-Boucoiran et al., 2015). The hybridisation intensities were interpreted as tetraploid SNP dosage scores (AAAA, AAAB, AABB, ABBB, and BBBB) using fitTetra (Voortrips et al., 2011) and were used to calculate the statistics required for the association study. SNP markers that were polymorphic and scorable were used for GWAS after filtering for minor allele frequency (MAF > 0.1) and missing data (<10%). Heterozygosity was calculated as the percentage of heterozygous loci (ABBB, AABB, and ABAB) compared to the total number of loci.

**Population Structure**

The population structure was modeled in STRUCTURE 2.3.4 (Pritchard et al., 2000; Falush et al., 2007) with a burn-in of 10,000 cycles and varied in the number of following Markov Chain Monte Carlo (MCMC) iterations (50,000 and 100,000) and the number of AFLP (400) and microsatellite and candidate gene markers (175 and 527). The SNP markers were not used for determining the population structure. We used the implemented admixture model with correlated allele frequencies and an initial alpha of 1 and accomplished three independent runs based on 10 repeats of the simulations for each K, from K = 1 to 10. Then, the most likely number of subpopulations was estimated based on the method of Evanno et al. (2005) using the InP (D) value (estimated likelihood) with the software StructureHarvester (http://taylor0.biology.ucla.edu/structureHarvester/) (Earl and von Holdt, 2012).

Furthermore, a PCoA (Principal Coordinate Analysis) was performed using DARwin 5.0.158 (Perrier and Jacquemoud-Collet, 2006) with the same subset of 927 AFLP, SSR and candidate gene markers.

**Genetic Diversity Analysis**

The genetic distance among the collection of rose genotypes was calculated with DARwin using a subset of 16,040 SNPs at the tetraploid dosage state from the filtered SNP set (MAF > 0.1 and missing data < 10%). An unweighted neighbor-joining (Saitou and Nei, 1987) dendrogram was constructed based on a distance dissimilarity matrix using a bootstrap analysis with 100 repetitions.

**Kinship Matrix Calculation**

A kinship matrix was used to establish and describe the relationship between the genotypes. Pairwise kinship coefficients were estimated using the programme SPAGeDi (Hardy and
Vekemans, 2002) based on the method of Hardy (2003) using 10,000 random SNP markers at the tetraploid dosage state from the 16,040 SNPs above. The diagonal of the matrix from SPAGEDi was set to two, and negative values were set to zero (Yu et al., 2006).

**Trait-Marker Associations**

Trait-marker association analysis was performed using the mixed linear model (MLM, \( K + Q \)) in TASSEL 3.0 (Bradbury et al., 2007). SNP markers with a minor allele frequency of less than 0.1 and with more than 10% missing data were excluded from further analysis. In TASSEL 3.0, marker allele configurations can only be used in a diploid configuration (e.g., AA, AB, or BB). Therefore, bi-allelic SNPs of the tetraploid rose cultivars were coded as diploids. For this, all possible heterozygous genotypes (AAAB, AABB, and ABBB) were coded as AB, similar to how Li et al. (2014) analyzed diploid and tetraploid Alfalfa genotypes and how Lindqvist-Kreuze et al. (2014) analyzed potato genotypes. Associations were estimated including the Q-matrix for population effects based on the output from STRUCTURE 2.3.4 (based on AFLP, microsatellite and candidate gene data) and the kinship matrix (K) calculated with SPAGEDi (based on SNP data). Bonferroni adjustments of the \( p \)-values were made to correct for the number of independent tests and to establish a threshold (Johnson et al., 2010). For this, a total of 19,074 independent tests (number of contigs) were assumed because a precise estimation of the real number of independent tests could not be made due to unknown linkages between most of the markers. An SNP marker was considered associated if its \(-\log_{10} p\)-value was greater than 5.58.

**Statistical Analysis**

The nonparametric Kruskal–Wallis rank-sum test was used to identify significant differences in the mean SNP effect in groups of cultivars. Spearman rank correlation was used to test the association between the anthocyanin content in petals from greenhouse and field-grown roses. Significant differences in the means of heterozygosity in different growth types of roses were calculated using the Wilcoxon signed rank test. The data were tested for normal distribution using the Shapiro–Wilk test (\( \alpha = 0.05 \)). The data that were not normally distributed were transformed using log- and Box-Cox transformation (Wessa, 2016). The statistical calculations were performed in Excel 2007, MYSTAT 12 (Systat Software, Inc.) and QtiPlot 0.9.9 (Vasilief, 2015).

**RESULTS**

**Population Structure**

The population structure was analyzed based on three independent runs of STRUCTURE that varied in the number of MCMC iterations (50,000 and 100,000) and the number of AFLP, microsatellite and candidate gene markers (575 and 927) for \( K = 1 \) to \( K = 10 \). The optimum number of \( K \) can be identified according to the maximum value of \( \text{LnP}(D) \) (Pritchard et al., 2000). In our data, the likelihood distribution increased slightly, leveled off and then decreased with a clear plateau from \( K = 3 \) to \( K = 5 \) (Figure S2A). Using the method of Evanno et al. (2005) in two independent runs with a total of 575 AFLP and SSR markers (burn-in 10,000; MCMC 50,000, and 100,000) and one run with 627 markers (burn-in 10,000; MCMC 50,000), the maximum for \( \Delta K \) was estimated at \( K = 3 \) (Figure S2B). The result was confirmed by additional independent runs (data not shown).

The structure of the population at \( K = 3 \) is visualized in Figure 1A. Below it are depicted the concordant results from the cluster analysis of the SNP data (Figure 1B). When using a threshold of 0.7 to assign individuals to a subpopulation or to classify them as a mix or as hybrid individuals (as (D’hoop et al., 2010) used in highly heterozygous tetraploid potato), subpopulation I, the largest group, consisted of 44 cultivars, which clustered according to their type or habit, particularly hybrid tea and floribunda roses. Subpopulation II contained 17 recently bred (1985–2011) cultivars, except for New Dawn (1930), which all have a groundcover habit. Subpopulation III, the smallest group, comprised only five cultivars belonging to the old garden type of roses: Damask (‘Rose de Resht’, before 1900), Alba (‘Small Maidens Blush’, 1797), Bourbon (‘Louise Odier’, 1851), Hybrid Perpetual (‘Mrs. John Laing’, 1885) and Portland (‘Mme Knorr’, 1855 und ‘Mme Boll’, 1858) roses. The positioning of the cultivars was supported by high bootstrap values, except for some in the first subpopulation. However, 30 cultivars could not be assigned to any of the three subpopulations using the threshold of 0.7 for the classification. The results of a principal coordinate analysis (PCoA; Figure S3) based on genetic distances agreed with this division in three subpopulations. The hybrid cultivars that shared part of subpopulations I and II can be observed as an intersection (black dots) between these subpopulations.

**Genetic Diversity Analysis**

The kinship estimates based on SNPs (Figure S4) indicated no familial relationship between most of the rose genotypes. Approximately 59% of the pairwise kinship coefficients had values near zero (\(<0.005\)). Higher kinship estimates (0.10–0.20) were found between climber and ground cover roses. The highest values (0.26–0.39) were found within the group of old garden roses (population III).

The heterozygosity was determined based on the SNP data without considering the dosage of the markers (i.e., AAAB, AABB, and ABBB are all classified as heterozygote). When defined this way, the percentage of heterozygous loci is identical to the percentage polymorphic loci. On average, varieties displayed 55.2% of heterozygous loci ranging from 27% for variety No. 105 and 66.9% for variety No. 2. No correlation between heterozygosity and the age of the variety was observed (Figure S5).

On the other hand, there were significant differences between ground cover, climber, bedding roses, Hybrid teas and shrub roses, both in the average level of heterozygosity and in the variation in heterozygosity within groups (Figure 2). Groundcover roses had the lowest heterozygosity (44.4% heterozygous loci). Hybrid teas were significantly higher in heterozygosity (60.1%) compared to climbers and ground cover roses.
Phenotypic Characterization of morphological Traits
Total Anthocyanins
In many cultivars, the anthocyanin content in the petals was low or not detectable, which was not unexpected because 23 cultivars had a white and yellow flower color. The measured anthocyanins were in the range from $E_{525} = 0.35$ to $E_{525} = 33.22$ in the greenhouse and from $E_{525} = 0.33$ to $E_{525} = 38.39$ in the field. The distribution was skewed to the left (Figure S6 and Table 1) and very similar for both environments ($r = 0.942$, Figure 3).

Carotenoids
Carotenoids in the rose petal extracts were measured by their characteristic absorbance at 442 nm (Figure S1) from all cultivars cultivated in the greenhouse at the Federal Plant Variety Office Hannover and from 20 of the cultivars in the field at Herrenhausen. Because the Spearman rank correlation between the measured values was very high (Spearman’s $\rho = 0.939$) the carotenoid contents of the additional 76 cultivars grown in the field were not estimated to avoid redundant data (Figure S8). In white flowers and in many red flowers, the yellowish to orange pigments were not present or only found as minor pigments in the petals. The maximum amount of carotenoids was detected in the yellow flowering cultivar “China Girl” ($E_{442} = 0.8854$). In roses classified as orange, a balanced occurrence of anthocyanins and carotenoids was always measured. However, there was no overall correlation between the anthocyanin and carotenoid content in the rose petals ($r = -0.1803$, $p = 0.0836$).

Marker Trait Associations
Anthocyanins
Marker-trait associations were calculated in Tassel 3.0 with 39,831 SNPs based on their diploid allelic state (A:A, A:B,
The anthocyanin content in greenhouse-grown roses was significantly associated with 17 SNP markers, five of which were also associated with the anthocyanin content from field-grown roses (Table 2). These SNPs were Rh12GR_283_1910Q (in the Auxin response factor 8 gene), RhK5_1258_2078P (3 β-OH-steroid-dehydrogenase/decarboxylase isoform 2), RhK5_7371_202Q (Glutathione S-transferase), Rh12GR_20064_1031P and RhMCRND_20203_163Q (both are in the Medium-chain-fatty-acid-CoA ligase). We estimated the effects of the SNPs on the anthocyanin content in greenhouse-grown roses from 3.985 to 7.589 (Table 2). Under both conditions, the largest effect was found for the marker RhK5_1258_2078P. In Figure 4, two boxplots of anthocyanin content showed the direct effects of the markers. For the SNP in the auxin response factor 8 gene, the mean for the heterozygous genotypes A:B was 8.63 (E525nm) and was significantly higher (p = 5.03E-8) than 1.93 (E525nm) for the homozygous B:B genotypes. For the 3-β-OH-steroid-dehydrogenase SNP, the difference between the mean of the two groups (A:A = 10.26; A:B = 1.91) was also significant (p = 8.56E-7).

Because the genome sequence of Rosa sp. is not complete, the contigs of 133 SNPs with hits just below the Bonferroni threshold plus the 17 significant SNPs were blasted against the closely related genomes of Fragaria vesca and Prunus persica and mapped on these genomes (Figures 6A,B, Table S6). The assumption is that both genomes display sufficient microsynteny to the rose genome. This assumption is supported by the fact that these rose SNPs clustered in distinct regions of the Fragaria genome, particularly in linkage groups Fvb1, Fvb2, Fvb4, Fvb5, and Fvb6, and in the partly homologous linkage groups Pp01, Pp03, Pp05, Pp06, and Pp08 of Prunus persica (https://www.rosaceae.org/gb/gbrowse_syn/peach_apple_strawberry/). A blast of the contigs located three of the SNP markers in the coding region of anthocyanin biosynthesis genes, 4-coumarate-ligase (4_CL), flavonoid 3′-hydroxylase (F3′H) and glutathione-S-transferase (GST). The positions of these genes of the anthocyanin biosynthesis pathway and of further transcription factors are shown in the genome plots as green dots (Figures 6A,B). The putative transcription factors that are associated with anthocyanin accumulation include a ubiquitin-like protein SMT3 (SUMO1), WRKY transcription factor 17 and UTP4/Cirhin, a WD40 repeat protein (Freed et al., 2012).

### Carotenoids

Because as many as 351 SNPs were significantly associated with the accumulation of carotenoids in rose petals and surpassed the
### TABLE 2 | Significant SNPs for anthocyanin content in rose petals from greenhouse- and field-grown roses.

<table>
<thead>
<tr>
<th>Greenhouse</th>
<th>p-value</th>
<th>Effect (E_{525\text{nm}})</th>
<th>Allele A:A</th>
<th>Allele A:B</th>
<th>Allele B:B</th>
<th>Function (identified gene)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Combined</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rh12GR_283_1910Q</td>
<td>2.35E-09</td>
<td>−5.412</td>
<td>−</td>
<td>77</td>
<td>104</td>
<td>gene31631-v1.0-hybrid_Auxin_response_factor_B_(putative)</td>
</tr>
<tr>
<td>RhMCRND_20203_163Q</td>
<td>7.20E-09</td>
<td>−4.091</td>
<td>−</td>
<td>81</td>
<td>102</td>
<td>gene31669-v1.0-hybrid_Medium-chain-fatty-acid-CoA_igase_(probable)</td>
</tr>
<tr>
<td>Rh12GR_20064_1031P</td>
<td>1.86E-07</td>
<td>−4.521</td>
<td>102</td>
<td>75</td>
<td>−</td>
<td>gene31669-v1.0-hybrid_Medium-chain-fatty-CoA_igase_(probable)</td>
</tr>
<tr>
<td>RhKS_7371_202Q</td>
<td>2.02E-07</td>
<td>−4.492</td>
<td>−</td>
<td>79</td>
<td>104</td>
<td>gene31672-v1.0-hybrid_Glutathione_S transferase_(similar_to)</td>
</tr>
<tr>
<td>RhKS_1258_2078P</td>
<td>3.75E-07</td>
<td>−6.764</td>
<td>66</td>
<td>117</td>
<td>−</td>
<td>gene08692-v1.0-hybrid_3beta-hydroxysteroid-dehydrogenase/decarboxylase_isofrm_2_(At3BETAHSD/D2)_ (similar_to)</td>
</tr>
<tr>
<td>RhMCRND_7128_1021Q</td>
<td>4.51E-07</td>
<td>−1.935</td>
<td>−</td>
<td>83</td>
<td>92</td>
<td>gene31668-v1.0-hybrid_Medium-chain-fatty-acid-CoA_igase_(probable)</td>
</tr>
<tr>
<td>Rh12GR_64257_531Q</td>
<td>4.62E-07</td>
<td>−5.776</td>
<td>−</td>
<td>61</td>
<td>106</td>
<td>gene16750-v1.0-hybrid_Hypothetical_protein</td>
</tr>
<tr>
<td>RhKS_2615_1401Q</td>
<td>5.58E-07</td>
<td>−5.549</td>
<td>−</td>
<td>76</td>
<td>105</td>
<td>gene09148-v1.0-hybrid_Cysteine_proteinase_RD21a_(putative)</td>
</tr>
<tr>
<td>RhMCRND_20203_163Q</td>
<td>5.58E-07</td>
<td>−5.549</td>
<td>−</td>
<td>76</td>
<td>105</td>
<td>gene09148-v1.0-hybrid_Cysteine_proteinase_RD21a_(putative)</td>
</tr>
<tr>
<td>Rh12GR_77973_217Q</td>
<td>1.17E-06</td>
<td>−4.794</td>
<td>108</td>
<td>75</td>
<td>−</td>
<td>—</td>
</tr>
<tr>
<td>RhKS_12663_103QQ</td>
<td>1.32E-06</td>
<td>−3.628</td>
<td>−</td>
<td>69</td>
<td>114</td>
<td>gene24536-v1.0-hybrid_Possible_hemolysin_C_(probable)</td>
</tr>
<tr>
<td>Rh12GR_17814_425Q</td>
<td>1.62E-06</td>
<td>−3.136</td>
<td>104</td>
<td>81</td>
<td>−</td>
<td>gene31679-v1.0-hybrid_Ubiquitin-like_protein_SMT3_(putative)</td>
</tr>
<tr>
<td>RhKS_15799_993Q</td>
<td>1.80E-06</td>
<td>−5.472</td>
<td>−</td>
<td>60</td>
<td>119</td>
<td>gene31669-v1.0-hybrid_Medium-chain-fatty-acid-CoA_igase_(probable)</td>
</tr>
<tr>
<td>RhMCRND_1368_1182Q</td>
<td>2.00E-06</td>
<td>5.215</td>
<td>−</td>
<td>62</td>
<td>121</td>
<td>gene31631-v1.0-hybrid_Auxin_response_factor_B_(putative)</td>
</tr>
<tr>
<td>Rh12GR_283_1910P</td>
<td>2.38E-06</td>
<td>3.516</td>
<td>−</td>
<td>81</td>
<td>94</td>
<td>gene31631-v1.0-hybrid_Auxin_response_factor_B_(putative)</td>
</tr>
<tr>
<td>RhKS_19460_153P</td>
<td>2.45E-06</td>
<td>4.855</td>
<td>123</td>
<td>61</td>
<td>−</td>
<td>gene04292-v1.0-hybrid_Cytokinin-O-glucosyltransferase_2_(AtZOG2)_ (probable)</td>
</tr>
<tr>
<td>Rh12GR_38264_410P</td>
<td>2.58E-06</td>
<td>−4.584</td>
<td>88</td>
<td>79</td>
<td>−</td>
<td>—</td>
</tr>
<tr>
<td><strong>Greenhouse</strong></td>
<td><strong>p-value</strong></td>
<td><strong>Effect</strong> (E_{525\text{nm}})</td>
<td><strong>Allele A:A</strong></td>
<td><strong>Allele A:B</strong></td>
<td><strong>Allele B:B</strong></td>
<td><strong>Function (identified gene)</strong></td>
</tr>
<tr>
<td>Rh12GR_283_1910Q</td>
<td>9.73E-11</td>
<td>−5.740</td>
<td>−</td>
<td>40</td>
<td>52</td>
<td>gene31631-v1.0-hybrid_Auxin_response_factor_B_(putative)</td>
</tr>
<tr>
<td>RhKS_1439_606P</td>
<td>8.10E-09</td>
<td>−4.927</td>
<td>68</td>
<td>19</td>
<td>−</td>
<td>U3_small_nuclear_RNA-associated_protein_4/UTP4</td>
</tr>
<tr>
<td>RhMCRND_20203_163Q</td>
<td>1.17E-08</td>
<td>−5.135</td>
<td>−</td>
<td>42</td>
<td>51</td>
<td>gene31669-v1.0-hybrid_Medium-chain-fatty-acid-CoA_igase_(probable)</td>
</tr>
<tr>
<td>RhKS_1258_2078P</td>
<td>9.44E-08</td>
<td>−7.589</td>
<td>33</td>
<td>60</td>
<td>−</td>
<td>gene08692-v1.0-hybrid_3beta-hydroxysteroid-dehydrogenase/decarboxylase_isofrm_2_(At3BETAHSD/D2)_ (similar_to)</td>
</tr>
<tr>
<td>Rh12GR_20064_1031P</td>
<td>1.21E-07</td>
<td>−5.112</td>
<td>51</td>
<td>39</td>
<td>−</td>
<td>gene31669-v1.0-hybrid_Medium-chain-fatty-acid-CoA_igase_(probable)</td>
</tr>
<tr>
<td>RhKS_7371_202Q</td>
<td>1.23E-07</td>
<td>−5.426</td>
<td>−</td>
<td>41</td>
<td>52</td>
<td>gene31672-v1.0-hybrid_Glutathione_S transferase_(similar_to)</td>
</tr>
<tr>
<td>RhMCRND_319_1197P</td>
<td>7.49E-07</td>
<td>4.398</td>
<td>−</td>
<td>60</td>
<td>26</td>
<td>Urease_(similar_to)</td>
</tr>
<tr>
<td>RhKS_12076_566Q</td>
<td>7.74E-07</td>
<td>−5.193</td>
<td>−</td>
<td>42</td>
<td>52</td>
<td>Photosystem_I_reaction_center_subunit_XI_,<em>chloroplast</em>(PSI-L),<em>Precursor</em></td>
</tr>
<tr>
<td>RhMCRND_982_2342Q</td>
<td>8.43E-07</td>
<td>−6.928</td>
<td>−</td>
<td>30</td>
<td>60</td>
<td>gene30626-v1.0-hybrid_Formin-like_protein_20_(AtFh20)_(similar_to)</td>
</tr>
<tr>
<td>Rh12GR_17814_425Q</td>
<td>8.45E-07</td>
<td>−4.391</td>
<td>52</td>
<td>42</td>
<td>−</td>
<td>gene31679-v1.0-hybrid_Ubiquitin-like_protein_SMT3_(probable)</td>
</tr>
</tbody>
</table>

(Continued)
TABLE 2 | Continued

<table>
<thead>
<tr>
<th>Greenhouse</th>
<th>p-value</th>
<th>Effect&lt;sup&gt;b&lt;/sup&gt; (E&lt;sub&gt;525 nm&lt;/sub&gt;)</th>
<th>Allele A:A</th>
<th>Allele A:B</th>
<th>Allele B:B</th>
<th>Function (identified gene)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rh12GR_92431_414Q</td>
<td>8.48E-07</td>
<td>−3.985</td>
<td>−</td>
<td>35</td>
<td>58</td>
<td>gene31646-v1.0-hybrid_Serine/threonine-protein_kinase_PBS1_(probable)</td>
</tr>
<tr>
<td>RhK5_2615_1401Q</td>
<td>8.64E-07</td>
<td>−5.525</td>
<td>−</td>
<td>39</td>
<td>53</td>
<td>gene09148-v1.0-hybrid_Cysteine_proteinase_RD21a_(RD21)<em>Precursor</em>(putative)</td>
</tr>
<tr>
<td>RhK5_15799_993P</td>
<td>1.00E-06</td>
<td>−6.814</td>
<td>−</td>
<td>30</td>
<td>61</td>
<td>gene26062-v1.0-hybrid_Formin-like_protein_20_(AtFH20)_(similar_to)</td>
</tr>
<tr>
<td>RhK5_5774_854P</td>
<td>1.11E-06</td>
<td>4.321</td>
<td>62</td>
<td>26</td>
<td></td>
<td>Translation_initiation_factor_IF-2_</td>
</tr>
<tr>
<td>RhK5_12663_103Q</td>
<td>1.37E-06</td>
<td>4.258</td>
<td>33</td>
<td>57</td>
<td></td>
<td>gene24536-v1.0-hybrid_Possible_hemolysin_C_(probable)</td>
</tr>
<tr>
<td>RhK5_21626_409P</td>
<td>1.66E-06</td>
<td>4.064</td>
<td>61</td>
<td>25</td>
<td></td>
<td>gene07023-v1.0-hybrid_Formin-like_protein_20_(AtFH20)_(similar_to)</td>
</tr>
<tr>
<td>Rh12GR_3292_1365Q</td>
<td>2.20E-06</td>
<td>4.122</td>
<td>18</td>
<td>68</td>
<td></td>
<td>gene10566-v1.0-hybrid_Putative_indole-3-acetic_acid-amido_synthetase_GH3.9_(AtGH3-9)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>p-value</th>
<th>Effect&lt;sup&gt;b&lt;/sup&gt; (E&lt;sub&gt;525 nm&lt;/sub&gt;)</th>
<th>Allele A:A</th>
<th>Allele A:B</th>
<th>Allele B:B</th>
<th>Function (identified gene)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rh12GR_283_1910Q</td>
<td>4.51E-08</td>
<td>−5.451</td>
<td>−</td>
<td>37</td>
<td>52</td>
<td>gene31631-v1.0-hybrid_Auxin_response_factor_8_(putative)</td>
</tr>
<tr>
<td>RHMCGRND_20203_163Q</td>
<td>5.29E-08</td>
<td>−2.631</td>
<td>−</td>
<td>40</td>
<td>51</td>
<td>gene31669-v1.0-hybrid_Medium-chain-fatty-acid-CoA_ligase_(probable)</td>
</tr>
<tr>
<td>RhK5_1258_2078P</td>
<td>1.30E-07</td>
<td>−6.511</td>
<td>33</td>
<td>57</td>
<td></td>
<td>gene06692-v1.0-hybrid_3beta-hydroxysteroid-dehydrogenase/decarboxylase_isoform_2_(At3BETAHSD/D2)_similar_to</td>
</tr>
<tr>
<td>Rh12GR_20064_10231P</td>
<td>1.82E-06</td>
<td>5.051</td>
<td>51</td>
<td>36</td>
<td></td>
<td>gene31669-v1.0-hybrid_Medium-chain-fatty-acid-CoA_ligase_(probable)</td>
</tr>
<tr>
<td>RhK5_7371_202Q</td>
<td>1.94E-06</td>
<td>5.100</td>
<td>38</td>
<td>52</td>
<td></td>
<td>gene31672-v1.0-hybrid_Glutathione_S-transferase_(similar_to)</td>
</tr>
<tr>
<td>RHMCGRND_7128_1021Q</td>
<td>2.37E-06</td>
<td>−1.901</td>
<td>39</td>
<td>48</td>
<td></td>
<td>gene31668-v1.0-hybrid_Medium-chain-fatty-acid-CoA_ligase_(probable)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Total number of A:A, A:B, and B:B genotypes was counted in two environments and exceeded the number of 96 cultivars.

<sup>b</sup>The effect was presented as untransformed values.

![Box plot of the effect of the SNPs in the auxin response factor 8 and 3-ß-OH-steroid-dehydrogenase genes on the anthocyanin content in rose petals](image)

Bonferroni threshold of α = 2.62e-6 (Table S7), the effects of the significant SNPs on carotenoid content ranged from 0.00015 to 0.259 (E<sub>442 nm</sub>). Most of the significant SNPs formed two large clusters in linkage group 5 of *F. vesca* and *P. persica* with more than 250 SNPs (Figures 7A,B). They may be located on a part of the chromosome with low recombination. Two of
the SNPs were located on contigs encoding genes of the MEP (methylerythritol 4-phosphate) pathway, CMS (2-c-methyl-d-erythritol-cyclodiphosphatase) and DXR (1-deoxy-d-xylulose 5-phosphate reducto-isomerase). A third enzyme, more upstream in the carotenoid biosynthesis, is Zeaxanthin epoxidase (ZEP, \( p = 8.77e^{-6} \)). It was located in linkage groups Fvb1 and Pp07. ZEP is a part of the branch of the \( \beta \)-carotenoid biosynthesis that catalyzes the step from zeaxanthin to violaxanthin. Additionally, several significant SNPs were mapped to linkage group four of \( F. vesca \) and to linkage group one of \( P. persica \). The positions of these SNPs were close to the assumed position of a carotenoid cleavage dioxygenase gene, CCD4. Other significant SNPs were located in two cytochrome P450 monoxygenases (CP450): cytochrome P450_71A24 (Pp04, Fvb5) and cytochrome CP450_CYP749A22 (Pp01, Fvb4).

The effects of SNPs on the carotenoid biosynthesis genes CMS and DXR are shown as box plots in Figure 5. For the DXR-SNP, the mean of the homozygous A:A genotypes was 0.074 (E_442 nm), whereas the mean for the A:B genotypes was 0.325 (E_442 nm). The effect of the SNP for CMS, the subsequent gene after DXR in the carotenoid pathway, was 0.206 (E_442 nm), and the mean was also obviously higher in the A:B group [A:A = 0.0475 (E_442 nm); A:B = 0.253 (E_442 nm)].

**DISCUSSION**

Floral traits in ornamental roses are determined by a number of quantitative traits, e.g., petal number, flower size and a large number of secondary metabolites that constitute flower color and flower fragrance. Here, we applied GWAS based on the rose WagRhSNP array to analyse factors influencing the amount of anthocyanins and carotenoids in petals. In this study, we tried for the first time to utilize association genetics to exploit the vast phenotypic variation in mainly tetraploid cultivated roses for an analysis of quantitative traits.

**Heterozygosity is not Influenced by Cultivar Age**

The large number of markers used to genotype the association panel revealed a high average heterozygosity of 55.2% in the population, which is in agreement with previous studies on marker diversity in cultivated roses (Debener et al., 1996; Esselink et al., 2003). The value of heterozygosity is in fact the percentage of polymorphic SNPs; thus, it is not surprising that the value is somewhat lower than that found with 24 SSR markers in garden roses (Vukosavljev et al., 2013). Unlike many other cultivated plant species (Kilian et al., 2007; Gil-Ariza et al., 2009; Wang et al., 2010; Gross et al., 2014), we did not see a reduction over time when we considered the year of release of the variety. This may corroborate observations on the sensitivity of rose for inbreeding depression, which makes selfing unsuitable as a breeding strategy (Pipino et al., 2011). It might also be the result of the large range of ornamental traits that breeders have selected for, causing selection to never be consistently in the same direction. We found a lower percentage of polymorphic loci in the climber (40.9%) and ground cover genotypes (50.2%). This could indicate some degree of ascertainment bias if the genetic background of these groups of roses is partly different from those of the cut and garden roses from which the SNPs derived (Koning-Boucoiran et al., 2015; Smulders et al., 2015). This would require further studies, for instance, a study that searches for homologous regions in the genomes of diploid species that have contributed to the tetraploid groups.

**Population Structure**

Population structure and relatedness between genotypes can be confounding factors in association mapping (Nordborg and Weigel, 2008). Minimal population structure or relatedness will...
result in high statistical power, but larger collections offer more power, and a collection of 100–500 individuals is recommended (Hirschhorn and Daly, 2005; Rafalski, 2010). The STRUCTURE software with the implemented Bayesian clustering approach is a common tool to assess population structure with a moderate number of markers. Together with the estimation of kinship, this tool can reduce the rate of false positives in association mapping (Pritchard et al., 2000; Rafalski, 2010). With 400 AFLP and 175 SSR markers, we identified three subpopulations in our rose association panel of 96 different cultivars. This is comparable to the sample sizes used in various association mapping studies with other crop species. For instance, the association panel of Lindqvist-Kreuze et al. (2014) comprised 103 potato genotypes, and population structure was estimated with 120 SNP markers. Seventy-one almond cultivars were the basis of an association study on kernel phytosterol content (Font i Forcada et al., 2015) in which population structure was corrected using 40 SSRs. Simko et al. (2009) used 68 lettuce cultivars for association with disease resistance and validated the detected marker trait association in a second set of 132 cultivars.

Using Synteny
A major problem of association studies is filtering true marker-trait associations out of a large number of false-positive associations. Next to p-values and the extent of the observed effects, the clustering of significant markers in particular genomic regions is a criterion for true marker trait associations. However, for rose, no completed genome sequence is available and relative positions are known only for a small subset of the markers that we applied in this study. Therefore, we used two related rosaceous genomes assuming sufficient microsynteny to the rose genome: the very closely related genome of strawberry and the peach genome, which represents the next-closest relative (Shulaev et al., 2008). In several conventional marker mapping studies, strawberry was shown to be highly similar in its genome structure and marker order with only minor differences to roses (Gar et al., 2011; Spiller et al., 2011; Terefe-Ayana et al., 2012). This strategy proved informative, as we found a couple of clear clusters of significantly associated SNPs in both genomes. Our attempt to locate candidate markers to a genomic region failed for some markers (e.g., SUMO1) in one of the two genomes, but many others were located in both genomes (Figures 6A,B).

SNPs Associated With Anthocyanin Accumulation in Rose Petals
As a simple measure for red and pink flower colors, we used the total amounts of anthocyanins determined by spectrophotometry in extracts prepared at defined flower stages. This has the advantage that other factors, e.g., cellular pH, cofactors, or flower age, which influence the visual characteristics of the anthocyanins in the natural context, are excluded; therefore, the phenotypic complexity can be partially reduced. This strategy was successful, as evidenced by the high correlation between the greenhouse and the field environment, which differ significantly in terms of temperature profiles and UV radiation.

Our study indicates that at least five genomic regions contain factors influencing anthocyanin concentration. Interestingly, all of these regions contained either SNP markers from genes with known functions in anthocyanin metabolism or candidate genes mapping to these regions (Table S3). The cluster in linkage group Fvb1 of Fragaria comprised the marker with the lowest p-value, a homolog to an auxin response factor known to influence anthocyanin concentration by regulating auxin expression in apple (Ji et al., 2015), Arabidopsis (Liu et al., 2014), tobacco (Zhu et al., 2013) and cabbage (Kang and Burg, 1973), and a family member of GSTs, which have important functions in anthocyanin transport from the cytosol to the vacuole. GSTs are responsible for color variation in a number of ornamental species, including petunia and carnation (Zhao, 2015). Homologues of the SUMO-1 transcription factor on Fragaria chromosome 1 are regulators of signal transduction in auxin signaling in plants (del Pozo et al., 1998; Vierstra and Callis, 1999). Because chromosome 1 of Fragaria is mostly collinear with the ICM linkage group 2 of Rosa (Gar et al., 2011), this region is likely to be close to the QTL for anthocyanin content in a diploid biparental rose population (Henz et al., 2015).

Several transcription factors on Fragaria chromosome 2 were associated with anthocyanin accumulation: WRKY transcription factor 17, a ubiquitin-like protein_SMT3 (SUMO1), and UTP4/Girih, a WD40 repeat protein (Freed et al., 2012). Zorrilla-Fontanesi et al. (2011) detected in strawberry associations between a putative R2R3 Myb transcription factor and QTLs for anthocyanin accumulation in linkage group 2. The Fragaria chromosome 2 is largely syntenic with rose the ICM linkage group 6 harboring major QTLs for the anthocyanin content, which were also stable across several environments (Henz et al., 2015). The cluster of SNPs mapped on Fragaria chromosome 4 included SNPs in the rose 4_CL gene and various Myb transcription factors, which are known to regulate anthocyanin biosynthesis. The transcription factor myb90-like (myb90), also known as “Production of Anthocyanin Pigment 2” (PAP2), was identified on Fragaria chromosome 6 and is a member of the MBW-complex (Maier and Hoecker, 2015). The MBW-complex activates anthocyanin biosynthetic genes and is a complex of the transcription factors R2R3-Myb, basic Helix-Loop-Helix (bHLH), and WD40 proteins (Petroni and Tonelli, 2011; Maier and Hoecker, 2015).

SNPs Associated with Carotenoid Accumulation in Rose Petals
Eugster and Märki-Fischer (1991) could identify nearly 40 different carotenoids in the extract of rose petals, most prominently Violaxanthin together with Auroxanthin, Luteoxanthin and ß-Carotene (Ohmiya, 2011). Glick (2009) identified as main components Violaxanthin and Neoxanthin, which comprised 85% of the total carotenoid content in the petals of the rose cultivar “Frisco.” The biosynthetic pathway to Violaxanthin occurs via Zeaxanthin, and the modification is catalyzed by the enzyme Zeaxanthin epoxidase (ZEP). The carotenoids in the rose petals in our association study panel...
Schulz et al. Marker Association in Roses

FIGURE 6 | (A) GWAS for anthocyanin content. SNPs are mapped to the genome of homologous sequences in Fragaria vesca, including those in annotated genes (green dots). On top of the graph, the positions of various known candidate genes in the F. vesca genome sequence are shown as red triangles. The purple dotted line represents the Bonferroni adjusted significance level. For abbreviations of genes, including functions, see Table S8.

(B) GWAS for anthocyanin content. SNPs are mapped to the genome of homologous sequences in Prunus persica, including those in annotated genes (green dots). On top of the graph, the positions of various known candidate genes in the P. persica genome sequence are shown as red triangles. The purple dotted line represents the Bonferroni adjusted significance level. For abbreviations of genes, including functions, see Table S8.

were characterized spectroscopically at 442 nm, which does not distinguish between different carotenoids. We detected as many as 303 significant SNPs associated with carotenoid content, 250 of which clustered in two positions on chromosome 5 of both Prunus and Fragaria. The majority of these 250 significant SNPs are not located in the genes causing the effect themselves, but they are closely linked to one or few of such genes located on the same chromosomal regions with low recombination rates. This extreme clustering was most probably due to the high linkage disequilibrium around the two clusters. The causes of this LD are not clear yet. Possible reasons might be either linkage to factors suppressing recombination in roses or the presence of genes under high selection pressure in cultivated roses. However, the fact that these two large clusters of significant SNPs were detected on both the Fragaria and the Prunus genome independently indicates that this is not due to a computational artifact of the GWAS or potential assembly errors in the target genome regions, but a real effect of the chromosomal region. Both target sequences have been independently assembled by different research groups and our results on rose sequences that match the same region in both heterologous genomes indicate synteny for the location of these sequences in all three genomes. Due to the huge number of significant carotenoid SNPs, we further discuss only SNPs in potential candidate genes (Table S4).

Carotenoids are synthesized from isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) via the MEP pathway. Enzymes of the MEP pathway significantly influence carotenoid production (Lois, 2000; Moehs et al., 2001; Rodríguez-Concepción et al., 2001, 2003; Carretero-Paulet et al., 2002, 2006). We identified two SNPs for carotenoid accumulation in roses in the coding region of genes of the MEP pathway: CMS and DXR. The importance of DXR was shown with Arabidopsis, as down-regulation resulted in reduced pigmentation and defects in chloroplast development, whereas overexpression led to the accumulation of isoprenoids, such as chlorophylls and carotenoids (Carretero-Paulet et al., 2006).

Carotenoid accumulation is also influenced by degradation. Campbell et al. (2010) showed that a reduced expression of the
carotenoid cleavage dioxygenase 4 (CCD4) gene increased the carotenoid level in mature potato tubers 2- to 5-fold. Similarly, Glick (2009) found a high correlation between carotenoid degradation in the rose cultivars “Frisco” and “Golden gate” and the expression of RhCCD4. In chrysanthemum, the loss of the CmCCD4a gene caused a change in petal color from white to yellow; the level of color mutation from white to yellow may depend on the copy number of the CmCCD4a gene (Ohmiya et al., 2012; Yoshioka et al., 2012). We detected several significant SNPs for carotenoid accumulation on Fragaria chromosome 4 and Prunus chromosome 1, close to a CCD4 gene (Figures 7A,B, Table S5). It is obvious that the activity of the CCD4 genes affects the carotenoid content in different plants, but the role of the CCD4 genes in the degradation of carotenoids in roses needs more research.

Beside these SNPs located in the two regions on Fragaria chromosomes 4 and 5 and the corresponding Prunus chromosomes 1 and 5, we detected a SNP in the coding region of ZEP with a \( p \)-value of 8.77e-6 for the association located on Fragaria chromosome 1.

CONCLUSIONS

This is the first association mapping study in rose. We focussed on the anthocyanin and carotenoid contents, which largely determine petal color. The phenotype data were collected in the field and in the greenhouse, and the overall levels of these compounds were not influenced by the differences in environment. To analyse the GWAS-associated SNPs in the absence of a rose genome sequence, we mapped the underlying rose contigs to the genome sequence of the related species Fragaria vesca and Prunus persica. Clusters of hits on these sequenced genomes in regions with known candidate genes confirmed that these genomes are probably largely syntenic and suggested that we identified 17 (anthocyanins) to 351 (carotenoids) trait-marker associations. Some of these had large effect sizes: these QTLs may be useful in breeding for intense flower colors in that parental breeding lines with combinations of several markers with high SNP dosages (duplex to quadruplex) might now be selected using the validated SNPs.
AUTHOR CONTRIBUTIONS

DS conducted the experiments, did the statistical analysis and wrote most parts of the manuscript. RS conducted parts of the experiments. RV conducted part of the data analysis and contributed in writing the manuscript. ML contributed to the experimental setup, contributed data to some of the experiments and contributed in writing the manuscript. TD was involved in planning the experiments, in statistical analysis, and wrote parts of the manuscript.

FUNDING

The research was in part funded within the program “Zentrales Innovationsprgramm Mittelstand (ZIM) of the German Bundesministerium für Wirtschaft und Energie (BMWi).”

ACKNOWLEDGMENTS

This study was carried out in part with financial support from the AiF (ZIM: KF 2554802MD0). We are grateful to the German rose breeders W. Kordes'Söhne, Rosen Tantau and Noack Rosen for providing plant material and carrying out part of the field trials. We thank Klaus Dreier and the gardeners from the Department of Molecular Plant Breeding for excellent assistance, especially Burkhard Spellerberg and Jörn Klocke, and the Federal Plant Variety Office for providing greenhouse facilities in Hannover.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2016.01798/full#supplementary-material

REFERENCES


Schulz et al. Marker Association in Roses.


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Schulz, Schott, Vorrips, Smulders, Linde and Debener. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.