frontiers
in Plant Science

ORIGINAL RESEARCH
published: 21 April 2015
doi: 10.3389/fpls.2015.00249

Using RNA-Seq to assemble a rose transcriptome with more than 13,000 full-length expressed genes and to develop the WagRhSNP 68k Axiom SNP array for rose (*Rosa* L.)

Carole F. S. Koning-Boucoiran*[1†], *G. Danny Esselink*[1], *Mirjana Vukosavljev*[1†], *Wendy P. C. van 't Westende*[1], *Virginia W. Gitonga*[1†], *Frans A. Krens*[1], *Roeland E. Voorrips*[1], *W. Eric van de Weg*[1], *Dietmar Schulz*[2], *Thomas Debener*[2], *Chris Maliepaard*[1], *Paul Arens*[1] and *Marinus J. M. Smulders*[1]

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

In order to develop a versatile and large SNP array for rose, we set out to mine ESTs from diverse sets of rose germplasm. For this RNA-Seq libraries containing about 700 million reads were generated from tetraploid cut and garden roses using Illumina paired-end sequencing, and from diploid *Rosa multiflora* using 454 sequencing. Separate *de novo* assemblies were performed in order to identify single nucleotide polymorphisms (SNPs) within and between rose varieties. SNPs among tetraploid roses were selected for constructing a genotyping array that can be employed for genetic mapping and marker-trait association discovery in breeding programs based on tetraploid germplasm, both from cut roses and from garden roses. In total 68,893 SNPs were included on the WagRhSNP Axiom array. Next, an orthology-guided assembly was performed for the construction of a non-redundant rose transcriptome database. A total of 21,740 transcripts had significant hits with orthologous genes in the strawberry (*Fragaria vesca* L.) genome. Of these 13,390 appeared to contain the full-length coding regions. This newly established transcriptome resource adds considerably to the currently available sequence resources for the *Rosaceae* family in general and the genus *Rosa* in particular.

Keywords: *Rosa*, transcriptomics, EST, SNP array, genotyping, assembly

OPEN ACCESS

Edited by:
Soren K. Rasmussen,
University of Copenhagen, Denmark

Reviewed by:
Fabrice Foucher,
Institut Nationale de la Recherche Agronomique, France
Kathryn Kamo,
United States Department of Agriculture, USA

****Correspondence:***
Marinus J. M. Smulders,
Wageningen UR Plant Breeding,
Wageningen University and Research Centre, PO Box 386, NL-6708 PB Wageningen, Netherlands
rene.smulders@wur.nl

† Present Address:
Carole F. S. Koning-Boucoiran,
HAN University of Applied Sciences, Institute of Applied Sciences, Nijmegen, Netherlands;
Mirjana Vukosavljev,
Pheno Geno Roses D.O.O., Novi Sad, Serbia;
Virginia W. Gitonga,
Lex+ East Africa, Naivasha, Kenya

Specialty section:
This article was submitted to Crop Science and Horticulture, a section of the journal Frontiers in Plant Science

Received: *28 January 2015*
Accepted: *27 March 2015*
Published: *21 April 2015*

Citation:
Koning-Boucoiran CFS, Esselink GD, Vukosavljev M, van 't Westende WPC, Gitonga VW, Krens FA, Voorrips RE, van de Weg WE, Schulz D, Debener T, Maliepaard C, Arens P and Smulders MJM (2015) Using RNA-Seq to assemble a rose transcriptome with more than 13,000 full-length expressed genes and to develop the WagRhSNP 68k Axiom SNP array for rose (Rosa L.). Front. Plant Sci. 6:249. doi: 10.3389/fpls.2015.00249

Introduction

Whereas, cut rose is economically the most important ornamental crop worldwide (761 million euros in The Netherlands in 2011), the rose genome sequence has not been completed yet. In fact, a long history of interspecific hybridization and selection (Debener and Linde, 2009; Smulders et al., 2011; Vukosavljev et al., 2013; Zhang et al., 2013) has led to a complicated taxonomy that is not fully resolved (Koopman et al., 2008; Fougère-Danezan et al., 2015). Commercial rose cultivars are mostly tetraploid and highly heterozygous. Therefore, inheritance patterns of quantitative traits may be complex. Phenotypic traits such as flower stem production, flower shape, flower color or disease resistance are of economic importance for breeders and growers (Debener and Linde, 2009; Smulders et al., 2011), and need to be better understood genetically in order to be able to apply marker-assisted selection in breeding programs. There is a consensus genetic map for rose (Spiller et al., 2011).

EST sequences are an efficient source of various markers for the construction of dense genetic linkage maps and the identification of QTLs (Vukosavljev et al., 2015). Based on next-generation sequencing technologies, four EST studies on roses have been published so far. Dubois et al. (2012) produced ESTs from 13 rose tissues and studied the expression of genes involved in flowering and scent biosynthesis. Yan et al. (2014) analyzed gene expression during flower blooming, whereas Kim et al. (2012) focused on miRNAs related to color genes in four rose cultivars. Yan et al. (2015) analyzed ascorbate biosynthesis genes and transcription factors in *Rosa roxburghii* fruits. These studies used a *de novo* assembly pipeline. An orthology-guided reference transcriptome assembly was developed by Ruttink et al. (2013) in *Lolium*, in order to be able to improve the assembly sequences from a highly heterozygous species in the absence of a reference genome sequence. The Genome Database for Rosaceae (GDR, Jung et al., 2014) contains, besides about 500,000 rose ESTs (2013), data of the fully sequenced woodland strawberry (*Fragaria vesca* L.) genome (Shulaev et al., 2011) that may be used as a reference genome for comparative genomics in rose. Strawberry is the closest related species for which a genome sequence is available.

SNP array platforms have been developed for various *Rosaceae* crops including apple (Chagné et al., 2012; Bianco et al., 2014), cherry (Peace et al., 2012), and peach (Verde et al., 2012). They have been shown to be an important resource facilitating the production of dense genetic maps and subsequent QTL mapping of important traits, genome-wide association analysis, pedigree-based analysis, and genomic selection (Bianco et al., 2014). The required density depends on the type of application and on the degree of LD on the germplasm—which in outcrossing species is often small (a few centimorgan). Dense maps are necessary to be able to identify haplotypes and study patterns of introgression with sufficient resolution power (Zhang et al., 2013). Tightly linked markers improve the efficiency of marker-assisted selection in breeding (Jänsch et al., 2015). In order to generate a versatile and large SNP array for rose, we set out to mine ESTs from diverse sets of germplasm comprising tetraploid cut and garden roses as well as a diploid garden rose, so that the SNPs on the array would be polymorphic in a wide range of rose genetic backgrounds.

Here, we present the results of (i) three transcriptome *de novo* assemblies based on tetraploid cut and garden rose genotypes and a diploid rose, (ii) the development of a 68K genotyping SNP marker array on the Axiom platform, and the (iii) construction and (iv) annotation of a non-redundant rose transcriptome for tetraploid roses using the genome sequence of diploid strawberry. This study generates valuable genomic resources (i.e., EST library with annotations and genetic diversity between different genotypes) and resulted in the construction of a large SNP array that can serve as a genotyping platform for future studies in diploid and tetraploid roses.

Materials and Methods

Plant Material

Three sources of material were used. From the parents of the K5 segregating population of tetraploid cut roses (*Rosa hybrida*, Yan et al., 2005; Koning-Boucoiran et al., 2012), P540 (mother) and P867 (father), petals were harvested in three stages of flower development (S1, S2, S3, **Figure 1A**). This material is designated as K5. For garden roses (designated as GR) whole flowers at three flowering stages (closed, half-opened, and fully-opened) as well as young leaves were harvested of 12 European and Canadian garden rose cultivars: Morden Fireglow, Adelaide Hoodless, Prairie Joy, Morden Blush, Diamond Border, Nipper, J.P. Connell, Princess of Wales, Heritage, Graham Thomas, Morden Centennial (MC), and Red New Dawn (RND) (**Figure 1A**; Vukosavljev et al., 2013). For the diploid *Rosa multiflora* hybrid 88/124-46 (Biber et al., 2010; designated as Rh88) leaves were harvested from plants grown under optimal growing conditions (control) and from plants 1–6 days after inoculation with black spot, *Diplocarpon rosae* (Debener et al., 1998), powdery mildew, *Podosphaera pannosa* (Linde and Debener, 2003), or downy mildew, *Peronospora sparsa* (Schulz et al., 2009), 1 h after wounding or after 1 h of 40°C heat stress. Sampled tissues were immediately frozen in liquid nitrogen and stored at −80°C until RNA extraction.

RNA Preparation

Total RNA was isolated from P540 and P867 by using the RNeasy Plant Mini Kit (QIAGEN, Westburg, The Netherlands) according to the manufacturer's instructions. The K5 RNA samples were prepared by pooling equal amounts of total RNA isolated from petals of the three flower stages as mentioned above. For GR total RNA was isolated from the 12 garden roses according to Chang et al. (1993) with modifications described in **Supplementary Table ESM 1**. Total RNA samples were prepared by pooling equal amounts of RNA from the four samples (three flower stages and leaves). For Rh88 total RNA from treated and untreated leaves was isolated using the Invitek RNA extraction kit (STRATEC Molecular GmbH, Berlin, Germany) according to the manufacturer's instructions. Remaining DNA was removed by digestion with RNAse-free DNAse as specified in the extraction kit.

Sequencing of cDNA

RNA samples of the K5 parents were sequenced by ServiceXS (Leiden, The Netherlands) using Illumina's standard operation protocols (2 × 75 bp paired end) on a Genome Analyser II. The RNA samples from the 12 garden rose cultivars (GR) were sent to GATC Biotech (Constance, Germany) where 12 cDNA libraries were sequenced using 2 × 100 bp paired end sequencing on a HiSeq 2000. For the pools of stressed and unstressed Rh88 leaves two random primed normalized cDNA libraries were constructed and sequenced in two 454 FLX Titanium runs at the Roche 454 sequencing center in Branford (USA). The first pool consisted of eight independent RNA isolations of untreated leaves and the second pool consisted of equal amounts of RNA from four independent extractions of all stress-treated leaves.

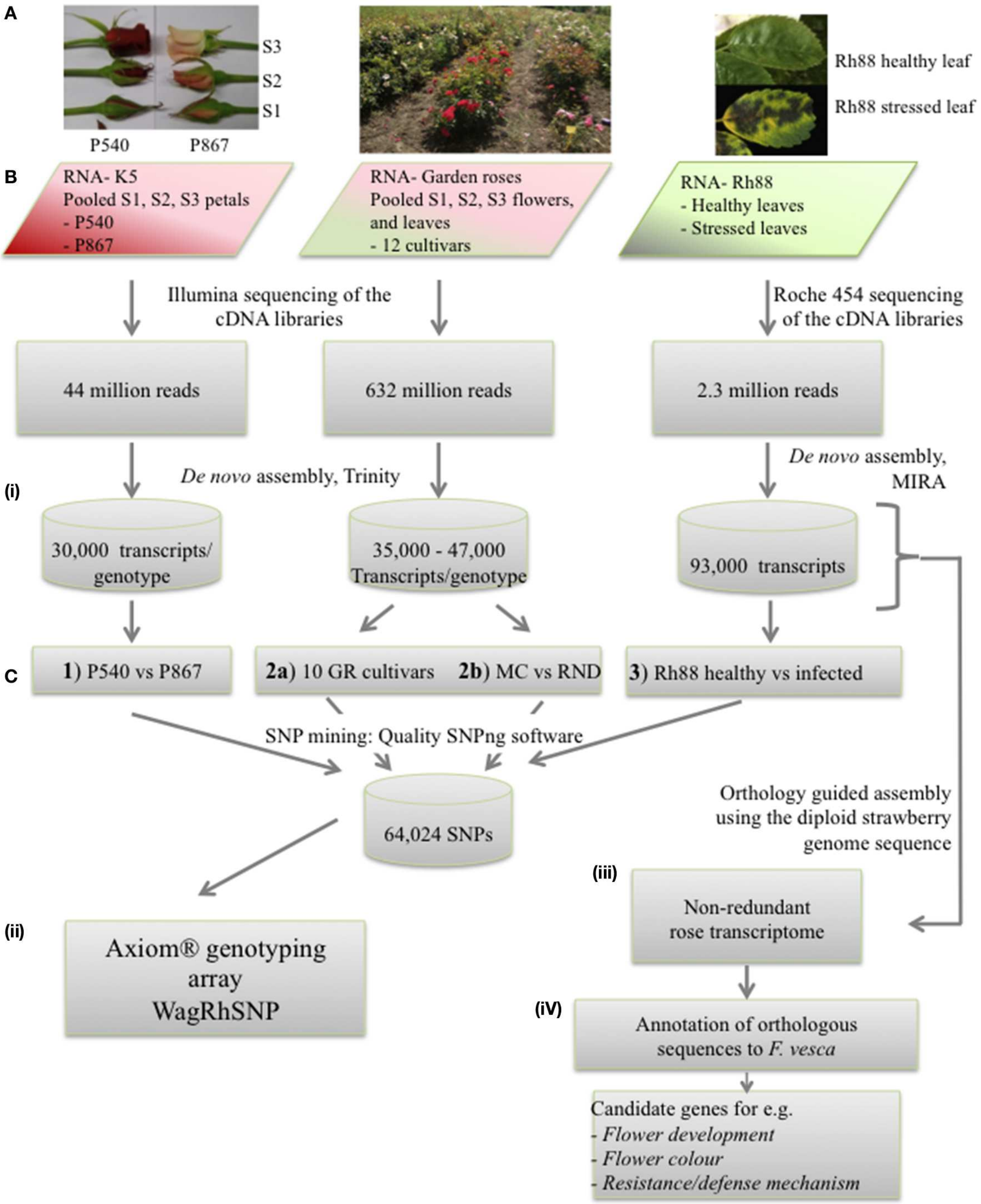

FIGURE 1 | Overview of the strategy used to assemble the rose EST data, to mine SNPs, and to develop the WagRhSNP Axion SNP array. **(A)** Pictures of the plant material used to isolate RNA. **(B)** Sequencing of the three sets. **(C)** Data analysis: (i) SNP mining in the three sample sets (GR, garden roses; MC, Morden Centennial; RND, Red New Dawn), (ii) development of the WagRhSNP Axiom® genotyping array, (iii) transcriptome assembly, and (iv) identification and annotation of sequences orthologous with *F. vesca*. All SNPs are in Supplementary file ESM2. The SNPs identified in set 1) are coded with a 'K', those in 2a) with a 'G', in 2b) with a 'M' and in 3) with a 'D'.

Pre-Processing of the Sequences

Illumina reads were pre-processed using Prinseq-lite (vs. 0.20.3) which included the trimming of nucleotides having a phred score lower than 25, the trimming of 10 nucleotides from the 5′ end to remove the bias of the nucleotide content of the reads due to the random hexamer priming (Hansen et al., 2010), the trimming of poly A/T tails, the removal of duplicate reads, of low complexity reads (DUST approach), of reads shorter than 50 nucleotides and of reads with more than one ambiguous nucleotide. Next, the paired-end reads were processed for overlapping sequences using COPE (Liu et al., 2012). All unconnected reads were used as normal paired-end reads, all connected read pairs (i.e., merged read pairs) and single reads were used as single-end reads.

The 454 reads were pre-processed using the FASTX toolkit (v0.0.13) with the same trimming and filtering as the Illumina reads except that reads smaller than 100 bp were discarded. Duplicate reads were removed using USEARCH v5.2.32 (Edgar, 2010). The reads were corrected for homopolymer nucleotide tracks using Acacia v1.52 (Bragg et al., 2012).

Transcriptome *De novo* Assembly

De novo transcriptomes were assembled per sample set using Trinity (min_kmer_cov 2, Grabherr et al., 2011) for the Illumina datasets for transcriptome assembly and SNP calling. For the 454 dataset the MIRA/CAP3 assemblers in the iAssembler (v1.3.2) pipeline (Zheng et al., 2011) were used for the transcriptome assembly, and the CLC assembler for SNP detection, in which sequences of the control and stressed leaves were combined to increase the number of reads. To select for relevant biological transcripts within each data set, RSEM (RNA-Seq by Expectation-Maximization) was used for transcript abundance estimation. If less than 1% of the total reads of a component (IsoPct) matched with a specific transcript, the transcript was not included in the subsequent steps of the analysis. Of the others the most abundant transcript/isoform was selected. To this end, all reads were mapped against all transcripts. Rh88 transcripts were also filtered for fungal sequences by blasting against available fungal sequences of *Marssonina brunnea* (Zhu et al., 2012).

SNP Mining

SNPs were identified within subsets of the sequences: (1) between K5 parents P540 and P867, (2a) among 12 garden rose cultivars, (2b) in a subset of those, namely between the two garden roses MC and RND (Vukosavljev et al., in preparation), and (3) within Rh88 (**Figure 1C**). For each of the sets the individual transcriptomes were assembled with CAP3 (default settings with−p 97) to generate a reference transcriptome per dataset. All reads of each dataset were mapped to their specific reference transcriptome with Bowtie 2 (Langmead and Salzberg, 2012) with modified settings (—very-sensitive—rfg 5, 10) and filtered for map quality (>25) using SAMtools.

The resulting SAM file was used for SNP calling using QualitySNPng (Nijveen et al., 2013) with modified settings (the minimal similarity score per polymorphic site, similarityAllPolymorphicSites: 0.8, the minimal number of reads per allele set at 5). Using the filtering options of QualitySNPng, SNPs found in transcripts displaying a larger number of haplotypes than theoretically expected were discarded. For the tetraploid parents of each mapping population (P540 × P867 and MC × RND), a maximum of 8 haplotypes (four haplotypes per parent) per transcript was expected. The resulting SNP markers with 35 flanking nucleotides on both sites without additional SNPs or InDels were marked as potential markers. To filter against paralogous markers, transcripts containing selected SNP markers were searched against their own reference transcripts (BLASTn *e*-value 1-30). Transcripts with two or more hits were discarded since this indicates that they may be present several times in the genome.

To prevent interference with chloroplast DNA during the array hybridization all sequences around SNP markers were screened against the chloroplast genome of *Fragaria vesca* (http://www.rosaceae.org). The genome sequence of *Fragaria vesca* was used to identify and remove sequences with potential splice junction sites. For this, markers were searched against the Fragaria sequences (BLASTn *e*-value 1-5) and discarded using custom perl scripts if their sequences matched with fewer than 68 bp (95%). All A/T and C/G polymorphisms were also excluded since genotyping these SNPs requires twice the number of probes using the Axiom platform. As a last step the mined SNP markers of the three sets were compared to remove redundancy across sets.

Axiom Genotyping Array: WagRhSNP

The selected SNPs were submitted to Affymetrix (Santa Clara, CA, USA) for a final analysis to determine whether the probes could be synthesized. Affymetrix discarded SNPs that shared similar sequences, as this could interfere with hybridization. We decided to include two probes for each SNP on the array, each probe targeting one of the strands (coded as AX_set_ID), as this allows an additional quality check during genotype calling with dosage scoring (Smulders et al., 2015). The array also includes 3000 non-polymorphic control probes (coded as DQC-sample name, DQC is a measure of the extent to which the distribution of signal values is separated from background values). The SNPs on the array are listed in **Supplementary Table ESM 2**, mentioning for each SNP its SNP_ID, the flanking sequences, the alleles of the SNP, the Rh-Fv ortholog transcript code, the *Fragaria vesca* protein code, and the annotation in *Fragaria* whenever available.

Orthology-Guided Reference Transcriptome Assembly

An orthology-guided assembly procedure according to Ruttink et al. (2013) was followed for the construction of a non-redundant rose transcriptome sequence. The set of non-redundant proteins of *Fragaria vesca* was downloaded from PLAZA2.5 (Van Bel et al., 2012) to guide the assembly. tBLASTn, carried out in 2013, (*e*-value cut-off 1e-5 and up to 250 hits allowed) was used to search for protein hits (*e*-value 1e-10) with all retained rose transcripts of this study. The transcripts with a significant tBLASTn hit with a *F. vesca* protein were grouped and assembled using CAP3 with default settings. The assembled transcripts were compared to the *F. vesca* proteins using BLASTx, and if the highest-scoring protein returned the *Fragaria* gene originally used as the highest scorer, then the two genes were considered as putative orthologs, and the transcript was selected and tentatively named after the most likely orthologous gene in *F. vesca* (Rh-Fv transcripts). Next, the longest ORF of each selected rose sequence was determined (Trinity package, Grabherr et al., 2011), the 3′ and 5′ UTR sequences trimmed off and the remaining sequences were reassembled using CAP3 with default settings to select the final set of orthologous sequences of rose.

Functional domains predicted in *Fragaria* and available in GDR (Jung et al., 2014; http://www.rosaceae.org) were mined for our rose Rh-Fv transcripts. They were also scanned for protein signatures from superfamilies reported in various databases such as Smart, Tigr, Panther, Pfam, FPrint, Profilescan, ProDom, and Gene3D (Zdobnov and Apweiler, 2001).

Results

Individual *De novo* Transcriptome Assemblies

RNA-Seq results of K5 (two cut rose genotypes) and GR (12 garden rose cultivars) were obtained by using Illumina paired-end sequencing (**Figure 1A**, **Table 1**). To generate the RNA-Seq data of Rh88 (garden rose Rh88 hybrid, **Figure 1A**) 454 sequencing was used.

Sequencing of cDNAs from petals of the tetraploid cut rose K5 parents gave 44 million 75 bp paired-end reads. After quality trimming and merging, 81% of the reads remained either as paired-end (12 million) or as single end reads (23 million). In total ∼78,000 transcripts were obtained, but ∼18,000 (24%) of these were subsequently discarded since they were less abundant and possibly splice variants. For SNP calling it is better to avoid these. On one hand they could represent paralogous genes, which would lead to nucleotide differences between paralogs rather than between alleles of the same locus. On the other hand, if they are from splice variants they may map in multiple contigs, which could mean that all of these would unnecessarily be excluded from the SNP calling. Ultimately, approximately 30,000 transcripts per genotype (**Table 1**) were identified.

Similarly, the sequencing of cDNAs from petals and leaves of GR resulted in 632 million 100 bp paired-end reads. After quality trimming and merging, 71% of the reads remained either as paired-end or single end reads. After filtering of the most abundant transcripts during the assembly, 48% were discarded as possible splice variants. Ultimately, between 35,000 and 47,000 transcripts per cultivar were identified (**Table 1**).

The sequences of Rh88 of healthy and stressed leaves were combined and resulted in 2.3 million reads with an average length of 360 bp. After quality trimming, duplicate removal, and homopolymer correction 38% remained. Filtering of the most abundant transcripts yielded 93,974 transcripts (**Table 1**). This large number was partly due to the presence, in the black spot and powdery mildew infected leaves, of fungal genes. The sequences were therefore blasted (blastx) against available fungal sequences of *M. brunnea* (Zhu et al., 2012), based on which 12,705 sequences with an average homology of 53.4% were discarded. The remainder was used in further analysis.

Development of the WagRhSNP Array

SNPs were mined in transcripts containing at least one reliable SNP (Tang et al., 2006) for the three rose datasets separately (**Table 2**). The smallest transcript was 135 bp long, assembled from 10 reads and showed 1 reliable SNP while the largest transcript was 14,270 bp long, from 15,286 reads, showing 77 reliable SNPs. Some transcripts contained up to 123 reliable SNPs. A small majority of the SNPs (62.3% for K5, 59.8% for GR, and 57.7% for Rh88) were transitions (C/T or G/A), as is common

TABLE 1 | Sequencing results and analysis.

		K5		GR												Rh88
		P540	**P867**	**Morden fireglow**	**Adelaide hoodless**	**Prairy joy**	**Morden blush**	**Diamond border**	**Nipper**	**J.P. Connell**	**Princess of wales**	**Heritage**	**Graham thomas**	**Morden centennial**	**Red new dawn**	
Pre-processing	Raw sequences	21,618,027	22,429,015	60,647,695	41,913,639	48,707,935	41,848,533	43,835,093	50,571,712	75,422,211	51,936,149	45,693,660	64,714,441	56,337,877	51,161,500	2,321,272
	Nb of paired end reads*	5,007,927	7,373,422	26,291,879	16,666,189	21,243,155	18,504,871	19,233,108	17,583,539	29,474,176	22,940,332	19,677,516	27,523,415	23,823,735	22,280,757	
	Nb of single end reads*	10,947,580	12,280,512	17,443,525	12,039,348	14,115,627	12,061,017	12,681,671	14,427,176	21,993,145	14,860,946	13,137,663	18,819,850	16,129,980	14,498,963	885,661
De novo assembly	Transcripts	39,354	38,905	77,119	71,751	77,995	69,404	70,791	74,434	110,332	79,095	66,307	82,446	84,115	82,244	112,023
	Discarded variants	9331	8894	30,196	39,677	32,298	29,675	35,294	37,457	67,667	41,332	31,281	42,848	45,570	43,066	0
	Remaining transcripts	30,023	30,011	41,555	38,318	37,102	39,725	35,497	36,976	42,664	37,763	35,025	39,597	38,545	39,176	93,974

*Number of paired-end reads and single reads per sample after quality trimming and merging during pre-processing of the data.

TABLE 2 | Number of SNPs mined in the three rose datasets.

Sample set	Number of transcripts[a]	Average transcript length (bp)	STD	Average transcript coverage[b]	STD	Number of reliable SNPs per transcript	Average density SNP/100 bp	STD
K5	19080	1134.5	952.2	643.0	2017.6	1–123	0.6	0.41
GR[c]	51106	1342.9	994.0	787.9	2700.3	1–96	0.5	0.40
Rh88	5493	1296.2	740.2	48.5	34.7	1–57	0.4	0.53

STD = standard deviation.

[a] *Transcripts containing at least one reliable SNP.*

[b] *Average of the number of reads per transcript.*

[c] *GR: All 12 garden rose cultivars included.*

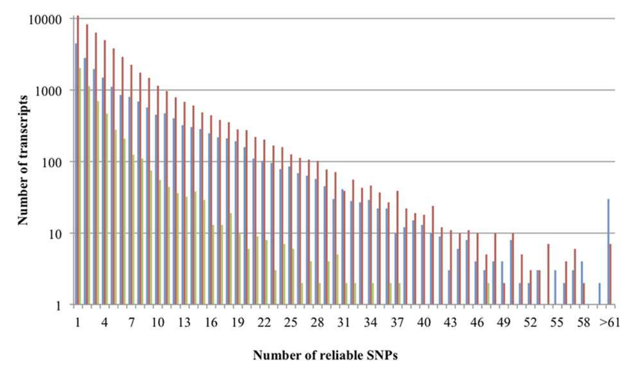

FIGURE 2 | Distribution of reliable SNPs per transcript of K5 (blue bars), GR (red bars), and Rh88 (green bars).

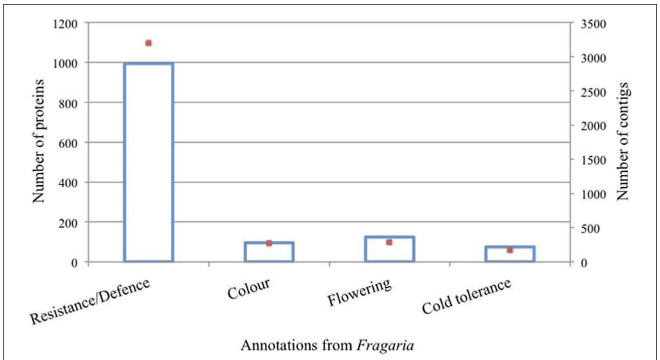

FIGURE 3 | Number of proteins (blue bars) and their respective transcripts (red dots) in which they were identified for relevant protein functions in the sampled tissues.

(in almond: 51%, Wu et al., 2008; in wheat and maize: 45% and 55% respectively, Edward et al., 2008; in cassava, up to 65%, Lopez et al., 2005). The average SNP density varied between 0.4 and 0.6 per 100 bp among the three sample sets but the variation between transcripts within each sample was large, as pointed out by the standard deviations (**Table 2**). The distribution of reliable SNPs per transcript is given in **Figure 2** (note the log scale). Half of the transcripts had 1-3 SNPs per transcript.

The WagRhSNP array includes a total of 68,893 reliable SNPs. Of these, 26,354 SNPs were identified between the cut rose parents (labeled: K_transcript number_SNP number), 26,364 SNPs among the 12 garden rose cultivars (named: G_transcript number_SNP number), 14,293 SNPs between MC and RND (named: M_transcript number_SNP number) and 1882 SNPs between alleles of Rh88 (named D_transcript number_SNP number). Probes for both strands are on a genotyping array which we named the WagRhSNP rose array totaling 137,786 probes (described in **Supplementary Table ESM 2**). This Axiom® array is commercially available for genetic studies in rose. The availability of a signal from two independent probes for each SNP enables additional quality control during scoring of the signal dosage, which is important for accurate genotyping in tetraploids (Vukosavljev et al., in preparation; Arens et al., in preparation).

Orthology-Guided Assembly of the Rose Transcriptome

The orthology-guided assembly procedure, developed for *Lolium* by Ruttink et al. (2013), was applied. The non-redundant protein coding sequences from strawberry (34,748 unigenes) were mapped against the 628,240 rose transcripts identified during the *de novo* assembly (**Figure 1**). In total 381,621 (60.7%) transcripts were mapped against 28437 strawberry unigenes. They could be assembled into 21740 orthologous sequences, whereby singletons (i.e., strawberry unigenes with only a hit against single rose transcript) were discarded. Of these 13,390 sequences (61.6%) corresponded to complete unique ORFs (**Supplementary Table ESM 3**).

A single *Fragaria* protein could map against up to 25 rose transcripts, which partly overlapped. For example, when looking at FV0G46670, the orthology-guided assembly identified 25 rose transcripts out of 250 transcripts that had been identified during the *de novo* assembly.

Annotation of the Rose Transcripts

The annotated rose transcripts with orthologs in *Fragaria* (named Rh-Fv transcripts) were investigated for functions and involvement in processes such as disease resistance and defense mechanisms, flower development and flower color (**Supplementary Table ESM 4**). The transcripts were further grouped into GO classes and functional domains, and were mined based on the InterPro Scan prediction of *Fragaria* (Zdobnov and Apweiler, 2001). A total of 2498 different protein domains were predicted to be present (**Supplementary Table ESM 5**). The putative annotations of the *Fragaria vesca* genome were linked to the protein domains identified in our dataset. Within those domains, 8090 proteins were identified based on 17,726 transcripts (**Supplementary Table ESM 5**).

TABLE 3 | List of the top 20 gene ontology (GO) terms the most represented among the annotated transcripts from cut and garden roses, for each of three main GO categories.

GO terms		Number of transcripts	GO categories
GO:0005515	Protein binding	1871	Molecular function 38.2%
GO:0003677	DNA binding	1077	
GO:0004674	Protein serine/threonine kinase activity	978	
GO:0003676	Nucleic acid binding	881	
GO:0003700	Transcription factor activity	828	
GO:0005488	Binding	613	
GO:0003824	Catalytic activity	588	
GO:0016491	Oxidoreductase activity	553	
GO:0003723	RNA binding	532	
GO:0005524	ATP binding	464	
GO:0000166	Nucleotide binding	410	
GO:0009055	RNA binding Electron carrier activity	376	
GO:0043565	Sequence-specific DNA binding	316	
GO:0004888	Transmembrane receptor activity	265	
GO:0016301	Kinase activity	262	
GO:0004497	Monooxygenase activity	251	
GO:0016758	Transferase activity	242	
GO:0020037	Heme binding	241	
GO:0004553	Hydrolase activity	218	
GO:0005215	Transporter activity	205	
GO:0006468	Protein amino acid phosphorylation	1027	Biological process 38.3%
GO:0055114	Oxidation reduction	1012	
GO:0006355	Regulation of transcription, DNA-dependent	644	
GO:0008152	Metabolic process	553	
GO:0055085	Transmembrane transport	534	
GO:0006508	Proteolysis	466	
GO:0006915	Apoptosis	423	
GO:0007165	Signal transduction	396	
GO:0006952	Defense response	340	
GO:0005975	Carbohydrate metabolic process	336	
GO:0006412	Translation	329	
GO:0045087	Innate immune response	260	
GO:0006457	Protein folding	258	
GO:0009651	Response to salt stress	244	
GO:0046686	Response to cadmium ion	232	
GO:0006629	Lipid metabolic process	186	
GO:0009793	Embryo development ending in seed dormancy	176	
GO:0009737	Response to abscisic acid stimulus	170	
GO:0006886	Intracellular protein transport	165	
GO:0006810	Transport	162	
GO:0005634	Nucleus	1401	Cellular compound 20.8%
GO:0016020	Membrane	1175	
GO:0005886	Plasma membrane	1172	
GO:0016021	Integral to membrane	634	
GO:0009507	Chloroplast	566	
GO:0005737	Cytoplasm	462	
GO:0005622	Intracellular	452	
GO:0005739	Mitochondrion	333	
GO:0005773	Vacuole	306	
GO:0005840	Ribosome	281	
GO:0031224	Intrinsic to membrane	259	
GO:0005829	Cytosol	243	
GO:0005783	Endoplasmic reticulum	205	
GO:0009941	Chloroplast envelope	199	
GO:0009570	Chloroplast stroma	188	
GO:0005618	Cell wall	177	
GO:0009505	Plant-type cell wall	173	
GO:0005730	Nucleolus	112	
GO:0005794	Golgi apparatus	98	
GO:0009535	Chloroplast thylakoid membrane	96	
			Others 3%

Figure 3 illustrates the distribution of four protein functions (resistance/defense, flower color, flowering, and cold tolerance) and the number of rose candidate genes identified within each functional class. For instance, 300,000 transcripts were annotated as involved in resistance/defense mechanisms, with homology to 1000 Fragaria proteins. Among them six different putative mlo-like genes (up to 25 transcripts) were identified out of the eight mlo-like genes present in the *Fragaria* genome. These included the full sequence of RhMLO1, RhMLO3 and RhMLO4, and a partial sequence of RhMLO2 (Kaufmann et al., 2012). The unique putative TMV resistance protein N matched with 270 rose genes.

GO terms were assigned to the annotated transcripts (**Supplementary Table ESM 4**). Around 20,000 genes (38%) were assigned to both molecular functions (such as transport, signal transduction, and structural molecules) and to biological processes (such as flower development, protein metabolism and response to stress). Around 10,000 genes (21%) were assigned to the category cellular compounds (e.g., organelle synthesis/regulation). **Table 3** shows, for instance, that 978 transcripts belong to the GO class (GO:0004674) protein serine/threonine kinase activity, which includes not only the above identified protein (LRR receptor-like serine/threonine-protein kinase) but also other types of kinases involved in other processes (**Supplementary Table ESM 4**).

The Rh-Fv transcripts (i.e., those with orthologous sequences in the *Fragaria* genome) of our dataset were compared to the

ROSAseq database (Dubois et al., 2012; http://iant.toulouse.inra.fr/R.chinensis). This database contains 80,714 rose EST clusters (based on 454 sequencing) longer than 100 nucleotides (average length of 444 ± 209.4 bp), annotated with the *Fragaria vesca* genome. Of these, 56,899 EST clusters had a BLASTn hit to 14,302 Rh-Fv transcripts of our study with a mean nucleotide identity of 96.2%. In general multiple ROSAseq EST clusters mapped to a single Rh-Fv transcript: 95% of these Rh-Fv transcripts matched with up to 10 EST clusters from the ROSAseq database, while ca. 85% of the ROSAseq EST clusters matched a single Rh-Fv transcript (**Supplementary Table ESM 6**). For instance, three not annotated ROSAseq EST clusters (RC013751, RC050162, and RC061808) were similar to Rh-Fv transcript FV1G02570.m1, which was annotated as a putative mlo-like protein 1. Furthermore, five EST clusters from the ROSAseq database (RC016326, RC022993, RC028093, RC040307, RC072319), four of which not annotated, showed similarity to one Rh-Fv transcript (FV1G02570.m1) annotated as putative mlo-like protein 6. On the other hand, ROSAseq clusters annotated as putative mlo-like proteins 2, 11, and 14 did not have any similarities with the Rh-Fv transcripts.

Discussion

In this study, three sets of transcript sequence data were combined and analyzed in order to develop a SNP genotyping array. The generated markers may be used to produce dense genetic linkage maps at tetraploid and diploid level, improve QTL and gene function analyses, and the study of synteny with *Fragaria*. The genotyping array with 68,893 SNPs is commercially available for the *Rosaceae* community.

The WagRhSNP Axiom Array

We chose to use the Axiom® array system of Affymetrix (Santa Clara, CA, USA). Advantages of the Axiom array system for SNP detection include the large number of probes that fit on the array, the small size of conserved probe sequences (so that additional SNPs do not interfere so often in sequences with a high density of SNPs), and that the array can be produced for 480 samples (5 microtiter plates) onwards using photolithographic templates (so that arrays ordered later will be identical to the original ones). Axiom arrays are being developed for other rosaceous crops as well, notably a 90K array for octoploid strawberry within the RosBREED project (Smulders et al., 2015).

By analyzing three data sets with two different sequencing platforms, more than 68k SNPs were identified and included on the WagRhSNP Axiom array. Per transcript the SNP frequencies of the K5 and the GR samples were 1 SNP every 167 bp and 200 bp, respectively, which is higher than the SNP frequency of 1 SNP/288 bp found in the highly heterozygous genome of apple (Chagné et al., 2012). SNPs on this array originate for approximately 40% from tetraploid cut roses and 60% from tetraploid garden roses, but it can be expected that they will be useful in all tetraploid germplasm, as cut roses represent a subset of the germplasm of garden roses (Vukosavljev et al., 2013). We included around 1000 SNPs identified in diploid *R. multiflora*. As tetraploid roses are the result of extensive hybridisation between (diploid) species and are probably segmental allopolyploids, many of the SNPs on the array that have been identified as polymorphism within and between tetraploid cultivars, may be polymorphic in diploid germplasm as well. Zhang et al. (2013) observed that SNP haplotypes at a SNPSTR locus were shared between tetraploid and diploid *Rosa* species.

The array will also be very useful as the SNPs reside in coding regions of genes that are being expressed. Genetic map positions of the SNP markers can thus be linked to transcript sequences and, if available, gene annotation. The Supplementary files contain the keys for the connections to the genes predicted in the *Fragaria vesca* genome sequence. Similarly, gene annotations can be screened for candidate genes, which then can be examined for the presence of SNPs.

Assembly Issues

The individual assemblies were performed without a reference genome and are therefore difficult to validate but they highlighted the diversity among the samples and they were used to identify reliable SNPs. Ruttink et al. (2013) indicated that *de novo* assemblies in highly heterozygous species typically yield more transcripts than the actual number of genes expressed, and that was the case here as well (**Table 1**). We proceeded to construct a common transcriptome from these samples by using orthology-guided assembly from Ruttink et al. (2013). The three different rose datasets had been produced with two different sequencing platforms (Illumina paired-end on a GAII and a HiSeq, and 454). The Illumina paired-end reads were short (110–200 bp) but we obtained up to 15,000 reads per transcript. The 454 reads of the diploid *R. multiflora* cultivar were longer (350 bp) but transcript depth was limited (max. 400 reads/transcript). Finseth and Harrison (2014) concluded that using Illumina reads alone one can produce a high quality transcriptome appropriate for RNA-Seq gene expression analyses, but that utilizing both 454 and Illumina is preferred. Hodgins et al. (2014) came to a similar conclusion.

Annotation

The orthology-guided assembly based on the *Fragaria* genome (**Supplementary Table ESM 3**) produced 21,740 orthologs, of which 13,390 appeared to be full-length coding regions with an average length of 1089 bp. In this way, we could identify over 1/3 of all the genes estimated to be present in the rose genome, a good result as tissues such as root, stem, fruit, and seed were not included in this study. They provide an additional resource to the 14,252 peptides (EST clusters with an average length of 444 bp) orthologous to *Fragaria* identified by Dubois et al. (2012) in their ROSAseq database, which was produced from diploid roses.

Conclusion

Our data provides the most comprehensive transcriptome resource currently available for rose with 13,390 expressed full-length genes identified. This resource adds significantly to the currently available genomics and bioinformatics resources for the genus *Rosa*. SNPs in many of these genes are present on the 68k WagRhSNP Axiom array, which will support candidate gene identification. The dense SNP array with 68,893 SNPs will enable producing dense genetic maps that are useful in genetic research and marker-assisted breeding.

Author and Contributions

Conception of the study: CK-B, VG, FK, WW, TD, CM, PA, MS; Collection of material: CK-B, MV, WW, VG, DS, PA; Production of the data: CK-B, MV, WW, DS, DT, MS; Assembly: DE; SNP selection and array design: GE, RV, WW, CM, PA, MS; Annotation of the rose transcripts: CK-B, GE; Writing of the manuscript: CK-B, PA and MS. All authors read and approved the final version of the manuscript.

Acknowledgments

This research was partly supported by the TTI Green Genetics projects "Hyperrose" and "Polyploids" and by the TKI-U "Polyploids" project (BO-26.03-002-001). Terra Nigra provided the K5 parents. Peter Cox (Roath) provided the 12 garden roses. The Roche 454 sequencing center in Branford (USA) is acknowledged for sequencing the Rh88 samples. L. Bellon, F. Brew, M. Mittmann, A. Pirani and T. Webster of Affymetrix are thanked for constructive discussions on the design of the array.

Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2015.00249/abstract

Supplementary Table ESM 1 | Adjusted protocol for RNA extraction from rose flowers based on Chang et al. (1993). (docx)

Supplementary Table ESM 2 | The complete description of all SNPs on the WagRhSNP array. For each SNP the following is listed: SNP_ID, flanking sequences, and alleles of the SNP (A/B). In addition, the Rh-Fv ortholog transcript code, the *Fragaria vesca* protein code, and the annotation in *Fragaria* are given whenever available. This file provides all sequence information about the SNPs on the array and their flanking regions, but it can also be used as a key to the most similar sequence in the Fragaria vesca (Fv) genome, if available, and vice versa. (csv)

Supplementary Table ESM 3 | Orthologous ORF sequences of roses, constructed using the orthology-guided assembly procedure. The sequence header contains the Rh_Fv ortholog name and the related Fragaria (Fv) transcript annotation. (csv)
Supplementary Table 3 will be uploaded to GDR and ROSAseq.

Supplementary Table ESM 4 | InterPro predictions for the orthologous transcripts of cut and garden roses. The predictions originate from the related Fragaria (Fv) transcript. (csv)

Supplementary Table ESM 5 | GO classes and GO annotations of the Rh-Fv transcripts. The annotations originate from the related Fragaria (Fv) transcript. (csv)

Supplementary Table ESM 6 | Key from ROSASeq transcripts to Rh_Fv_orthologs. The file contains ROSASeq transcripts with high similarity with Rh_Fv orthologous ORF sequences. The annotations originate from the related Fragaria (Fv) transcript. (csv)

References

Bianco, L., Cestaro, A., Sargent, D. J., Banchi, E., Derdak, S., Di Guardo, M., et al. (2014). Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus × domestica* Borkh). *PLoS ONE* 9:e110377. doi: 10.1371/journal.pone.0110377

Biber, A., Kaufmann, H., Linde, M., Spiller, M., Terefe, D., and Debener, T. (2010). Molecular markers from a BAC transcript spanning the *Rdr1* locus: a tool for marker-assisted selection in roses. *Theor. Appl. Genet.* 120, 765–773. doi: 10.1007/s00122-009-1197-9

Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P., and Tyson, G. W. (2012). Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat. Methods* 9, 425–426. doi: 10.1038/nmeth.1990

Chagné, D., Crowhurst, R. N., Troggio, M., Davey, M. W., Gilmore, B., Lawley, C., et al. (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS ONE* 7:e31745. doi: 10.1371/journal.pone.0031745

Chang, S., Puryear, J., and Cairney, J. (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* 11, 113–116. doi: 10.1007/BF02670468

Debener, T., Drewes-Alvarez, R., and Rockstroh, K. (1998). Identification of five physiological races of blackspot, *Diplocarpon rosae*, Wolf on roses. *Plant Breed.* 117, 267–270. doi: 10.1111/j.1439-0523.1998.tb01937.x

Debener, T., and Linde, M. (2009). Exploring complex ornamental genomes: the rose as a model plant. *Crit. Rev. Plant Sci.* 28, 267–280. doi: 10.1080/07352680903035481

Dubois, A., Carrere, S., Raymond, O., Pouvreau, B., Cottret, L., Roccia, A., et al. (2012). Transcriptome database resource and gene expression atlas for the rose. *BMC Genomics* 13:638. doi: 10.1186/1471-2164-13-638

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edward, K. J., Poole, R. L., and Barker, G. L. (2008). "SNP Discovery in Plants," in *Plant Genotyping II: SNP Technology*, *2*, *1*, ed R. J. Henry (Wallingford: CABI), 272.

Finseth, F. R., and Harrison, R. G. (2014). A Comparison of next generation sequencing technologies for transcriptome assembly and utility for RNA-Seq in a non-model bird. *PLoS ONE* 9:e108550. doi: 10.1371/journal.pone.0108550

Fougère-Danezan, M., Joly, S., Bruneau, A., Gao, X. F., and Zhang, L. B. (2015). Phylogeny and biogeography of wild roses with specific attention to polyploids. *Ann. Bot.* 115, 275–291. doi: 10.1093/aob/mcu245

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Hansen, K. D., Brenner, S. E., and Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* 38, e131. doi: 10.1093/nar/gkq224

Hodgins, K. A., Lai, Z., Oliveira, L. O., Still, D. W., Scascitelli, M., Barker, M. S., et al. (2014). Genomics of Compositae crops: reference transcriptome assemblies and evidence of hybridization with wild relatives. *Mol. Ecol. Resour.* 14, 166–177. doi: 10.1111/1755-0998.12163

Jänsch, M., Broggini, G. A. L., Weger, J., Bus, V. G. M., Gardiner, S. E., Bassett, H., et al. (2015). Identification of SNPs linked to eight apple disease resistance loci. *Mol. Breed.* 35, 45. doi: 10.1007/s11032-015-0242-4

Jung, S., Ficklin, S. P., Lee, T., Cheng, C.-H., Blenda, A., Zheng, P., et al. (2014). The genome database for rosaceae (GDR): year 10 update. *Nucleic Acids Res.* 42, D1237–D1244. doi: 10.1093/nar/gkt1012

Kaufmann, H., Qiu, X., Wehmeyer, J., and Debener, T. (2012). Isolation, molecular characterization, and mapping of four rose MLO orthologs. *Front. Plant Sci.* 3:244. doi: 10.3389/fpls.2012.00244

Kim, J., Park, J. H., Lim, C. J., Lim, J. Y., Ryu, J. Y., Lee, B.-W., et al. (2012). Small RNA and transcriptome deep sequencing proffers insight into floral gene regulation in Rosa cultivars. *BMC Genomics* 13:657. doi: 10.1186/1471-2164-13-657

Koning-Boucoiran, C. F. S., Gitonga, V. W., Yan, Z., Dolstra, O., van der Linden, C. G., van der Schoot, J., et al. (2012). The mode of inheritance in tetraploid cut roses. *Theor. Appl. Genet.* 125, 591–607. doi: 10.1007/s00122-012-1855-1

Koopman, W. J. M., Wisseman, V., De Cock, K., Van Huylenbroeck, J., De Riek, J., Sabatino, G. J. H., et al. (2008). AFLP markers as a tool to reconstruct complex relationships: a case study in Rosa (Rosaceae). *Am. J. Bot.* 95, 353–366. doi: 10.3732/ajb.95.3.353

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Linde, M., and Debener, T. (2003). Isolation and identification of eight races of powdery mildew of roses (*Podosphaera pannosa*) (Wallr.: Fr.) de Bary and the genetic analysis of the resistance gene *Rpp1*. *Theor. Appl. Genet.* 107, 256–262. doi: 10.1007/s00122-003-1240-1

Liu, B., Yuan, J., Yiu, S.-M., Li, Z., Xie, Y., Chen, Y., et al. (2012). COPE: an accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. *Bioinformatics* 28, 2870–2874. doi: 10.1093/bioinformatics/bts563

Lopez, C., Piegu, B., Cooke, R., Delseny, M., Tohme, J., and Verdier, V. (2005). Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihot esculenta* Crantz). *Theor. Appl. Genet.* 110, 425–431. doi: 10.1007/s00122-004-1833-3

Nijveen, H., van Kaauwen, M., Esselink, D. G., Hoegen, B., and Vosman, B. (2013). QualitySNPng: a user-friendly SNP detection and visualization tool. *Nucleic Acids Res.* 41, W587–W590. doi: 10.1093/nar/gkt333

Peace, C., Bassil, N., Main, D., Ficklin, S., Rosyara, U. R., Stegmeir, T., et al. (2012). Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. *PLoS ONE* 7:e48305. doi: 10.1371/journal.pone.0048305

Ruttink, T., Sterck, L., Rohde, A., Bendixen, C., Rouzé, P., Asp, T., et al. (2013). Orthology guided assembly in highly heterozygous crops: creating a reference transcriptome to uncover genetic diversity in *Lolium perenne*. *Plant Biotechnol. J.* 11, 605–617. doi: 10.1111/pbi.12051

Schulz, D. F., Linde, M., Blechert, O., and Debener, T. (2009). Evaluation of genus *Rosa* germplasm for resistance to black spot, downy mildew and powdery mildew. *Eur. J. Hort. Sci.* 74, 1–9. Available online at: http://www.pubhort.org/ejhs/2009/file_902214.pdf

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., et al. (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43, 109–116. doi: 10.1038/ng.740

Smulders, M. J. M., Arens, P., Koning-Boucoiran, C. F. S., Gitonga, V. W., Krens, F., Atanassov, A., et al. (2011). "Rosa. Chapter 12," in *Wild Crop Relatives: Genomic and Breeding Resources Plantation and Ornamental Crops*, ed C. Kole (Berlin; Heidelberg: Springer-Verlag), 243–275.

Smulders, M. J. M., Voorrips, R. E., Esselink, G. D., Santos Leonardo, T. M., Van 't Westende, W. P. C., Vukosavljev, M., et al. (2015). Development of the WagRhSNP Axiom SNP array based on sequences from tetraploid cut roses and garden roses. *Acta Hort.* 1064, 177–184. Available online at: http://www.actahort.org/books/1064/1064_20.htm

Spiller, M., Linde, M., Hibrand-Saint Oyant, L., Tsai, C.-J., Byrne, D. H., Smulders, M. J. M., et al. (2011). Towards a unified genetic map of diploid rose. *Theor. Appl. Genet.* 122, 489–500. doi: 10.1007/s00122-010-1463-x

Tang, J., Vosman, B., Voorrips, R. E., van der Linden, C. G., and Leunissen, J. A. M. (2006). QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. *BMC Bioinformatics* 7:438. doi: 10.1186/1471-2105-7-438

Van Bel, M., Proost, S., Wischnitzki, E., Movahedi, S., Scheerlinck, C., Van de Peer, Y., et al. (2012). Dissecting plant genomes with the PLAZA comparative genomics platform. *Plant Physiol.* 158, 590–600. doi: 10.1104/pp.111.189514

Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C. T., Gasic, K., et al. (2012). Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS ONE* 7:e35668. doi: 10.1371/annotation/33f1ba92-c304-4757-91aa-555de64a0768

Vukosavljev, M., Esselink, G. D., Van' t Westende, W. P. C., Cox, P., Visser, R. G. F., Arens, P., et al. (2015). Efficient development of highly polymorphic microsatellite markers based on polymorphic repeats in transcriptome sequences of multiple individuals. *Mol. Ecol. Resour.* 15, 17–27. doi: 10.1111/1755-0998.12289

Vukosavljev, M., Zhang, J., Esselink, G. D., van 't Westende, W. P. C., Cox, P., Visser, R. G. F., et al. (2013). Genetic diversity and differentiation in roses: a garden rose perspective. *Sci. Hortic.* 162, 320–332. doi: 10.1016/j.scienta.2013.08.015

Wu, S.-B., Wirthensohn, M. G., Hunt, P., Gibson, J. P., and Sedgley, M. (2008). High resolution melting analysis of almond SNPs derived from ESTs. *Theor. Appl. Genet.* 118, 1–14. doi: 10.1007/s00122-008-0870-8

Yan, H., Zhang, H., Chen, M., Jian, H., Baudino, S., Caissard, J.-C., et al. (2014). Transcriptome and gene expression analysis during flower blooming in *Rosa chinensis* 'Pallida'. *Gene* 540, 96–103. doi: 10.1016/j.gene.2014.02.008

Yan, X., Zhang, X., Lu, M., He, Y., and An, H. (2015). *De novo* sequencing analysis of the *Rosa roxburghii* fruit transcriptome reveals putative ascorbate biosynthetic genes and EST-SSR markers. *Gene* 561, 54–62. doi: 10.1016/j.gene.2015.02.054

Yan, Z., Denneboom, C., Hattendorf, A., Dolstra, O., Debener, T., Stam, P., et al. (2005). Construction of an integrated map of rose with AFLP, SSR, PK, RGA, RFLP, SCAR and morphological markers. *Theor. Appl. Genet.* 110, 766–777. doi: 10.1007/s00122-004-1903-6

Zdobnov, E. M., and Apweiler, R. (2001). InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848. doi: 10.1093/bioinformatics/17.9.847

Zhang, J., Esselink, G. D., Che, D., Fougère-Danezan, M., Arens, P., and Smulders, M. J. M. (2013). The diploid origins of allopolyploid rose species studied using single nucleotide polymorphism haplotypes flanking a microsatellite repeat. *J. Hort. Sci. Biotechnol.* 88, 85–92. Available online at: http://www.jhortscib.org/Vol88/88_1/11.htm

Zheng, Y., Zhao, L., Gao, J., and Fei, Z. (2011). iAssembler: a package for *de novo* assembly of Roche-454/Sanger transcriptome sequences. *BMC Bioinformatics* 12:453. doi: 10.1186/1471-2105-12-453

Zhu, S., Cao, Y.-Z., Jiang, C., Tan, B.-Y., Wang, Z., Feng, S., et al. (2012). Sequencing the genome of *Marssonina brunnea* reveals fungus-poplar co-evolution. *BMC Genomics* 13:382. doi: 10.1186/1471-2164-13-382

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.