Heliyon 6 (2020) e04518

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Transcriptome profiling of raspberry (*Rubus idaeus* Var. Amira) in response to infection by tomato ringspot virus (ToRSV)

Gloria González[a,*,1], Felipe Aguilera[b,1], Vívian D'Afonseca[c]

[a] Center of Biotechnology for Natural Resources (CenBIO), Faculty of Agricultural Sciences and Forestry, Universidad Católica del Maule, Talca, Chile
[b] Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Biológicas, Universidad de Concepción, Concepción, Chile
[c] Vice Rectory of Research and Post-Graduation (VRIP), Universidad Católica del Maule, Talca, Chile

ARTICLE INFO

Keywords:
Agricultural science
Plant biology
Bioinformatics
Microbiology
Genetics
Virology
Immunology
Rubus idaeus
Tomato ringspot virus (ToRSV)
RNA-Seq
De novo assembly
Differential expression
Immune response

ABSTRACT

Raspberry (*Rubus* sp.) is a berries fruit with an ongoing agricultural and commercial interest due to its high contents of flavonoids and nutrients beneficial for human health. The growing demand for raspberries is facing great challenges associated mainly with the dispersal of diseases, which produces a decrease in productivity and fruit quality. A broad range of genomic resources is available for other Rosaceae species; however, genomic resources for species of the *Rubus* genus are still limited. Here, we characterize the transcriptome of the *Rubus idaeus* (Var. Amira) in order to 1) provide clues in the transcriptional changes of *R. idaeus* against tomato ringspot virus (ToRSV); and 2) generate genomic resources for this economically important species. We generate more than 200 million sequencing reads from two mRNA samples of raspberry, infected and not infected by ToRSV, using Illumina technology. After *de novo* assembly, we obtained 68,853 predicted protein-coding sequences of which 71.3% and 61.3% were annotated using Gene Ontology and Pfam databases, respectively. Moreover, we find 2,340 genes with differential expression between raspberries infected and not infected by ToRSV. Analysis of these genes shows functional enrichments of the oxidation-reduction process, cell wall biogenesis, terpene synthase activity, and lyase activity. These genes could be involved in the raspberry immune response through the interaction of different metabolic pathways; however, this statement needs further investigations. Up-regulation of genes encoding terpene synthases, multicopper oxidases, laccases, and beta-glucosidases might suggest that these enzymes appear to be the predominant transcriptome immune response of *R. idaeus* against ToRSV. Furthermore, we identify thousands of molecular markers (i.e., SSRs and SNPs), increasing considerably the genomic resources currently available for raspberries. This study is the first report on investigating the transcriptional changes of *R. idaeus* against ToRSV.

1. Introduction

The *Rosaceae* is a moderately large family, with approximately 2,000 sexual species described (Kalkman, 2004). This family comprises several economically and nutritionally important crops cultivated worldwide, such as apples (*Malus domestica*), plums (*Prunus domestica*), pears (*Pyrus communis*), peaches (*Prunus persica*) and raspberries (*Rubus idaeus*). Of the worldwide cultivated fruit crops, raspberry is one of the most diverse genera comprising 12 subgenera (Jennings, 1988; Alice and Campbell, 1999). However, most of the berries cultivated in the world come from just two subgenera, red raspberry (*Rubus idaeus*) and black raspberry (*R. occidentalis*) (Deighton et al., 2000; Yousefi et al., 2013).

Recently, there has been a growing interest in understanding the physiological, biochemical, and molecular characteristics of raspberries. A number of studies have revealed that the fruits of raspberry species have essential nutrients, micronutrients, and phytochemicals with beneficial effects on the human diet and health (Hummer and Janick, 2009; Jean-Gilles et al., 2012). These benefits can be attributed to the medicinally active properties of phytochemicals such as polyphenols, flavonoids (e.g., anthocyanins and flavonols), condensed and hydrolysable tannins and phenolic acid derivatives (Kähkönen et al., 2001).

The red raspberry (*Rubus idaeus*) is recognized as an agriculturally and economically important species and is widely cultivated around the globe (Hummer and Janick, 2009). In South America, Chile is the main

[*] Corresponding author.
E-mail address: ggonzalez@ucm.cl (G. González).
[1] These authors contributed equally.

https://doi.org/10.1016/j.heliyon.2020.e04518
Received 14 August 2019; Received in revised form 28 October 2019; Accepted 17 July 2020

producing country of *R. idaeus* (Var. Amira), with 9,000 ha of the total cultivated area and 36,000 tons of raspberry production (http://www.internationalraspberry.net/home, last accessed 21 October 2019). Most of the Chilean raspberry production is as a frozen product and is destined mainly for exportation to the United States (43%), followed by Canada (16%) and France (9%) (http://www.odepa.cl/articulo/balance-general-de-la-industria-de-frambuesas-congeladas-febrero-de-2014/, last accessed 21 October 2019).

Chilean raspberry - *R. idaeus* - plantations are made on the basis of material spread through the etiolated shoot, which facilitates the spread of diseases and decreases significantly the yield and quality of fruit produced. Some pathogens such as viruses, viroids, and phytoplasmas can spread very easily through this plantation technique (Martin et al., 2013). The main threat affecting Chilean raspberry production is the tomato ringspot virus (ToRSV), causing infectious symptoms such as chlorotic rings and designs (Medina et al., 2006; Morales et al., 2009). This virus also produces a whitening of veins that usually appears during spring and tends to disappear during summer with plants without apparent symptoms of infection, but affecting clearly the quality of the fruit (Medina et al., 2006; Morales et al., 2009). Infections by fungi and bacteria can also cause harvesting problems associated mainly with distortions of the normal development of the plant, however, these infections can be controlled relatively easy with proper harvesting management. Conversely, viral infections are much more difficult to control and can cause a permanent decrease in raspberry production (Requena et al., 2007).

Although *R. idaeus* (Var. Amira) is an economically important agricultural species in Chile, there are currently no transcriptomic resources and molecular markers suitable for this species, limiting the use of genetic approaches to control ToRSV and other infectious diseases. This is not an exception, similar issues occur with other economically important raspberry species, with only genomic resources for two varieties: *Rubus* sp. (Var. Lochness) (Garcia-Seco et al., 2015) and *Rubus* ideaeus (Var. Nova) (Hyung Hyun et al., 2014).

Thanks to the advent of next-generation sequencing technologies, we can now generate a large amount of sequence data in a relatively faster and cheaper fashion. Transcriptome sequencing (RNA-Seq) has opened up many doors for high-throughput discovery of genes and genetic markers, as well as for quantification of gene expression, especially when no genome sequence is available. Most importantly, RNA-Seq has also become a standard technique in non-model plant species, producing transcriptome-wide maps that consist of transcript discovery and gene expression levels at any particular developmental and physiological condition (O'Rourke et al., 2014; Kakumanu et al., 2012; Cabeza et al., 2014; Kamenetsky et al., 2015).

Here, we performed high-throughput transcriptome sequencing using Illumina RNA-Seq technology to characterize the transcriptome of *Rubus idaeus* (Var. Amira) in order to profile gene expression levels in raspberries infected and not infected by ToRSV. Differential expression analysis revealed thousands of differentially expressed genes (DEGs) between raspberries infected and not infected by ToRSV. Transcriptome analysis also revealed thousands of microsatellites (SSRs) and single-nucleotide polymorphisms (SNPs). This study offers valuable resources for the development of molecular markers that can be used for further genetic research in raspberries.

2. Materials and methods

2.1. Plant material

Rubus idaeus (Var. Amira) plants were collected from Curicó province, Buenas Paz, Molina, Chile. The sampling was performed on raspberries that had actively growing sprouts, as well as typical signs and symptoms of tomato ringspot virus infection on leafs, including chlorosis, leaf curling and yellow ring spotting. The number of samples collected was 30 (4 replicates per plant) per hectare. Once raspberries were collected, they were transported in plastic bags on ice and were subsequently stored at -80 °C until further analysis. In addition, *in vitro* raspberry plants were used as virus-free plants (i.e., raspberries not infected by ToRSV).

2.2. Identification of tomato ringspot virus (ToRSV)

The identification of ToRSV was performed using two approaches. First, a double-antibody sandwich ELISA (DAS-ELISA) was used to detect tomato ringspot virus (ToRSV), following the manufacturer's protocol (Agdia, USA). Second, ToRSV was identified by molecular approaches. To this end, total RNA was extracted from 500 mg of raspberry leaves. Total RNA extraction was performed using the RNA Thermo Scientific MagJET Plant® kit according to the manufacturer's recommendations. RNA visualization was performed over 1% agarose gels in TAE 1X buffer, with each well containing 2 μl of the RNA samples (500 ng/μL), and loading buffer supplemented with GelRed™ Nucleic Acid Gel Stain (Biotium, USA).

The molecular identification of ToRSV was performed using the U1 and D1 primers and RT-PCR amplification program described by Griesbach (1995). Then, PCR amplicons were cloned using the StrataClone PCR Cloning (Agilent) kit in accordance with the manufacturer's instructions, and these cloned PCR products were bi-directional sequenced by Macrogen Inc. (Seoul, Korea). Sequence analysis was conducted by BLAST searches against the non-redundant NCBI protein database, with the Geneious R6 program (Biomatters Ltd., New Zealand).

2.3. cDNA library preparation and transcriptome sequencing

Total RNA was extracted from leaf from five raspberry plants infected by ToRSV and from five raspberry plants not infected by ToRSV, as mentioned above. Total RNAs were pooled in equal amounts (5 μg) to generate a mixed cDNA library of ToRSV-infected and healthy *R. idaeus* raspberries. Two cDNA libraries (one infected by ToRSV and other do not infected by ToRSV) were sequenced in paired-end mode with an Illumina HiSeq™ 2000 sequencer. cDNA library preparation and sequencing were conducted by Macrogen Inc. (Seoul, Korea), and raw sequencing data were deposited into the NCBI database (BioProject accession number: PRJNA354231).

2.4. Transcriptome de novo assembly

Quality check of raw Illumina sequences was performed using FastQC (v0.11.4) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed 21 October 2019. High-quality Illumina reads from the two libraries were obtained by removing low-quality reads and adapter sequences using Trimmomatic (v0.33) (Bolder et al., 2014). High-quality reads from those libraries were concatenated and *de novo* transcriptome assembly was performed with Trinity software (v2014-04-13) (Grabherr et al., 2011) and default parameters. To remove redundant transcripts, we used the iAssembler (v1.3.2) pipeline that performs iterative assemblies with MIRA (4 cycles) and CAP3 (1 cycle), followed by automated error detection and correction with MEGABLAST (Zheng et al., 2011). Finally, we retained unigenes (assembled transcripts) over 200 bp for further analysis.

2.5. Transcriptome annotation

Open reading frames (ORFs) were predicted from unigenes using TransDecoder (v2.0.1) with default parameters (i.e., protein encoding least 100 amino acids).

For transcriptome annotation, we used the Trinotate pipeline (v2.0.2) (http://trinotate.github.io/, last accessed 21 October 2019). Briefly, unigenes and predicted proteins were annotated by sequence similarity using BLASTx and BLASTp (v2.2.28+) (Camacho et al., 2009) against Swiss-Prot and KOBAS databases (e-value cut-off of 1×10^{-5}). Predicted proteins were annotated using profile hidden Markov models with

HMMER (v3.1b2) (Eddy, 1998) against Pfam-A database. Based on these annotations, Gene Ontology (GO), Pfam and Kyoto Encyclopedia of Genes and Genomes (KEGG) terms were assigned to each unigene. In addition, prediction for signal peptides, transmembrane domains and rRNA transcripts was conducted by SignalP (v4.1) (Petersen et al., 2011), TMHMM (v2.0) (Krogh et al., 2001), and RNAMMER (v1.2) (Lagesen et al., 2007), respectively. Finally, all annotations were loaded into the Trinotate SQLite database and a final annotation report was generated. The maximum e-value for reporting the best hit and associated annotations for each unigene were no more than $E = 1e^{-5}$.

GO functional classification was performed using WEGO software (Ye et al., 2006). KEGG terms were mapped against the KEGG database using KEGG Mapper – Reconstruct Pathway tool (http://www.genome.jp/kegg/tool/map_pathway.html, last accessed 21 October 2019). Finally, we evaluated the quality and completeness of the reference transcriptome using BUSCO (v1.1b1) (Simão et al., 2015). We ran BUSCO on the assembled transcripts against a database of highly conserved single-copy genes in Eukaryota, which includes 429 genes.

2.6. Transcriptome profiling and differential gene expression

High-quality reads from each library were mapped to the annotated transcriptome using Bowtie (v0.12.8) (Langmead et al., 2009). For transcript quantification, we used RSEM software (v1.2.19) (Li and Dewey, 2011), and transcript expression values, of each library, were normalized by sequencing depth and converted into TPM (Transcripts Per Million) (Li and Dewey, 2011).

Given our RNA-Seq experiment have no biological replicates, we decided to use two independent approaches to identify differentially expressed genes (DEGs) between raspberries infected by ToRSV and not infected by ToRSV. First, we used the IsoDE software (v1.0), which is a bootstrap-based differential gene expression approach that employs a traditional bootstrapping approach to resample RNA-Seq reads and estimates gene expression levels from both samples (Al Seesi et al., 2014). In IsoDE, the Expectation-Maximization IsoEM algorithm is used during the inference process to ensure accurate length normalization (Al Seesi et al., 2014). This is a non-parametric method that does not assume an underlying statistical distribution of the data and has been demonstrated that performs equally well or even better with transcriptomic datasets with few or no biological replicates (Al Seesi et al., 2014). We used the IsoDE bootstrap calculator (http://dna.engr.uconn.edu/~software/cgi-bin/calc/calc.cgi, last accessed 21 October 2017) for computing the bootstrap support needed to achieve a significance level of $p < 0.05$, given a number of bootstrap samples of 20 for each condition. To detect DEGs, IsoDE takes as input the FPKM estimates from bootstrap samples generated for both samples and calculates a fold change in the gene expression level for each gene between the two conditions, providing a p-value for each fold-change estimation.

DEGs were also identified using EBSeq (Leng et al., 2013) and read count data obtained from RSEM (Li and Dewey, 2011). EBSeq is an empirical Bayesian approach that evaluates the posterior probabilities of differentially and non-differentially expressed genes, assuming negative binomial distribution of the data (Leng et al., 2013). When biological replicates are not available, EBSeq estimates the variance by polling similar genes together, and a median normalization procedure similar to DESeq is applied for accounting for different sequencing depths providing a Bayesian false discovery rate (FDR) associated with each gene (Leng et al., 2013). To use comparable thresholds for determining the differentially expressed genes between the two conditions and reducing false-positives, we used a 2-fold change difference controlled with an FDR at 0.05. Finally, only those genes identified as differentially expressed using both IsoDe and EBSeq software were considered for further analysis.

Gene enrichment analysis was performed using GOseq R package (v1.24.0) (Young et al., 2010). Briefly, GOseq determines enrichment of GO terms, Pfam protein domains and KEGG pathway categories accounting for over-detection of differential expression for long or highly expressed transcripts (Young et al., 2010). GOseq was run using gene annotation length as bias data and our GO, Pfam and KEGG category-mapping files obtained from the Trinotate annotation pipeline. P-values for overrepresented GO, Pfam and KEGG categories were adjusted using the Benjamini and Hochberg method for FDR control (Benjamin and Hochberg, 1995). GO enrichments were visualized using REVIGO (Supek et al., 2011).

2.7. Identification of SSR and SNP markers

To identify microsatellite markers in the annotated transcriptome, we used the MIcroSatellite MISA software (http://pgrc.ipk-gatersleben.de/misa/, last accessed 21 October 2019). The identification criteria used for mono-, di-, tri-, tetra-, penta- and hexanucleotides were 20, 10, 7, 5, 5, and 5 repeats, respectively. The maximum number of base interrupting two SSRs in a compound microsatellite was set at 100 bp. In addition, we searched for potential single nucleotide polymorphism (SNP) markers in infected and not infected by ToRSV raspberry samples, using an approach comprising the use of Bowtie (v0.12.8) (Langmead et al., 2009), samtools (v1.3.1) (Li et al., 2009), and bcftools (v1.3.1) (Li, 2011). High-quality reads of each raspberry sample were aligned to the reference transcriptome using Bowtie (Langmead et al., 2009), and BAM files were processed by the mplieup function of samtools to produce BCF files (Li et al., 2009). These BCF files were then used to call genotypes, with the bcftools call function (Li, 2011).

3. Results

3.1. De novo transcriptome assembly and functional annotation of R. idaeus (Var. Amira) reference transcriptome

Because no reference genome exists for *R. idaeus* (Var. Amira), raw reads from the two cDNA libraries (i.e., infected and not infected by ToRSV) were combined and assembled into a reference transcriptome. A combined total of 229.7 million short reads were sequenced using Illumina technology. After low-quality filtering, 225.6 million (98.2%) were used for *de novo* transcriptome assembly. This assembly yielded a total of 68,853 high-quality unigenes or contigs, with an N50 of 3,375 bp. An overview of the sequencing and assembly process is shown in Table 1. The BUSCO results indicated that the reference transcriptome was of high quality, with 90% of the BUSCO eukaryote orthologs found.

Several complementary methods were used to annotate the *R. idaeus* (Var. Amira) reference transcriptome. First, transcripts and proteins were compared against the Swiss-Prot protein database using BLASTx and BLASTp (Camacho et al., 2009), respectively. A total of 42,189 (61.3%) transcripts/proteins shared sequence homology with known proteins from the Swiss-Prot database, whereas 26,664 (38.7%) do not match to known proteins and seem to be taxon-restricted genes. Then, by mapping Entrez Gene IDs to GO annotations, 49,074 (71.3%) transcripts were assigned to one or more GO terms.

Figure 1 shows the number and percentage of *R. idaeus* (Var. Amira) transcripts categorized into 51 functional groups, belonging to the three main GO ontologies: biological process, molecular function and cellular component. In total, 275,953 GO assignments were obtained: 103,351 were assigned to biological processes, 99,779 were assigned to molecular functions, and 72,823 were assigned to cellular processes. Under the biological process ontology: cellular process (29,684; 60.5%) was the largest group, which was followed by metabolic process (27,819; 56.7%), biological regulation (11,064; 22.5%), pigmentation (10,106; 20.6%) and response to stimulus (9,822; 20.0%). Under the molecular function ontology: binding (30,958; 63.1%) was the largest category, which was followed by catalytic activity (25,508; 52.0%), transporter activity (3,414; 7.0%), and transcription regulatory activity (2,135; 4.4%). Under the cellular component ontology: cell and cell part (both with 34,306; 69.9%) were the largest categories, which was followed by organelle

Table 1. Summary statistics of *R. idaeus* (Var. Amira) sequencing and *de novo* reference transcriptome assembly.

Feature	Number
Total raw reads in infected library	120,082,600
Total raw reads in not infected library	109,725,146
Total high-quality reads in infected library	117,920,221
Total high-quality reads in not infected library	107,656,511
Total (Trinity) contigs	145,893
Total non-redundant (iAssembler) unigenes	116,967
Total predicted proteins/genes (ORFs ≥100 aa)*	68,853
Longest gene	16,146 bp
Mean gene	2,698 bp
Unigene N50	3,375 bp
Unigene N90	1,586 bp
Total assembled bases	185,755,243 bp
GC content	41.64%

* This feature corresponds to the total number of genes/proteins that comprise the *R. idaeus* (Var. Amira) reference transcriptome.

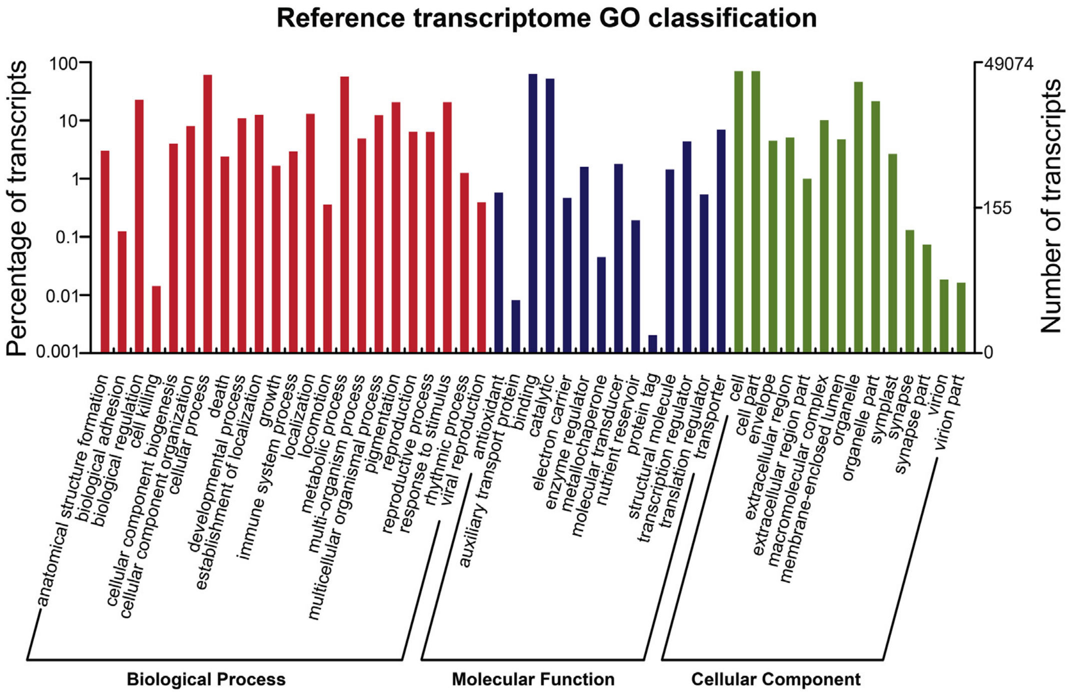


Figure 1. Gene Ontology (GO) classification of the *R. idaeus* (Var. Amira) reference transcriptome. Histogram showing summary for the three main GO categories: biological process (red), molecular function (blue) and cellular component (green). The left axis indicates the percentage of sequences of each category, whereas the right axis shows the total number of genes in each category.

(22,570; 46.0%), organelle part (10,477; 21.3%), and macromolecular complex (4,967; 10.1%).

In addition to the GO analysis, we conducted Pfam protein domains and KEGG pathway analyses to provide a deeper understanding of the transcript functions in the *R. idaeus* (Var. Amira) reference transcriptome. We found that 45,814 (66.5%) proteins were assigned to one or more Pfam domain/family terms (Figure 2A). Among the proteins with a detectable Pfam domain, 31,747 (69.3%) had sequence homology, based on BLAST hits, in the Swiss-Prot database. In contrast, 14,067 proteins (30.7%) had no BLAST hits but protein domains were detected using HMMER (Figure 2A). Then, Pfam domains/families were ranked according to the frequency of occurrence, and we found that Leucine-Rich Repeats (LRR) and Pentatricopeptide Repeats (PPR) were the most abundant ones. Figure 2B shows the top 15 most abundant Pfam domains/families in the *R. idaeus* (Var. Amira) reference transcriptome. In addition, KEGG pathway analysis indicated that 19,537 (28.4%) proteins were assigned to six main categories; these included: metabolism (9,438; 48.3%), human diseases (6,097; 31.2%), organismal systems (5,063; 25.9%), genetic information processing (4,607; 23.6%), environmental information processing (3,737; 19.1%), and cellular processes (3,197; 16.4%). Among the 42 pathways shown in Figure 2C, the most represented were signal transduction (3,593 proteins), carbohydrate metabolism (2,108 proteins), and translation (1,444 proteins).

3.2. Gene expression profiles and identification of differentially expressed genes

To characterize the transcriptome of *R. idaeus* (Var. Amira) in response to tomato ringspot virus (ToRSV), we collected samples from *R. idaeus* infected and not infected by ToRSV. Two libraries, one from

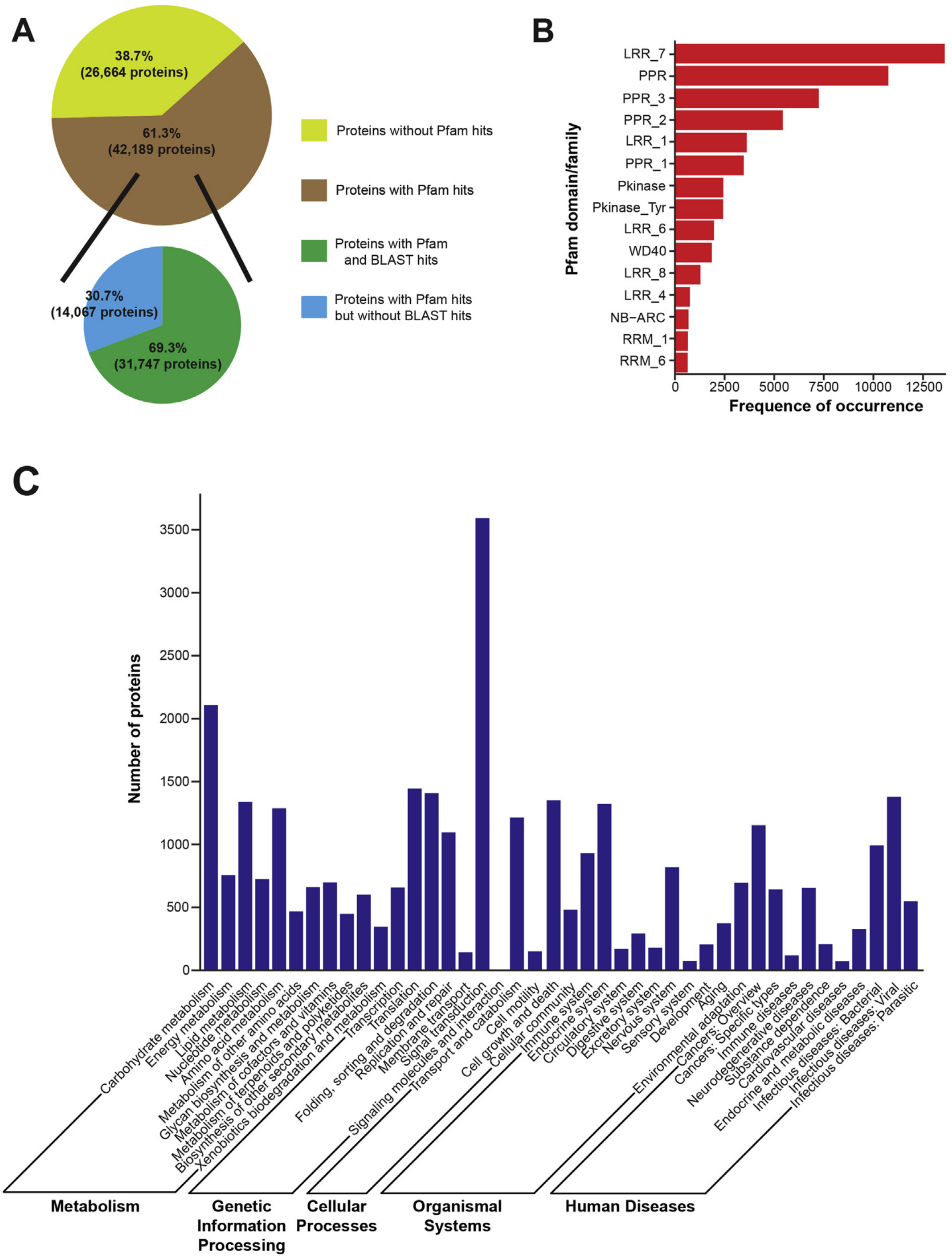
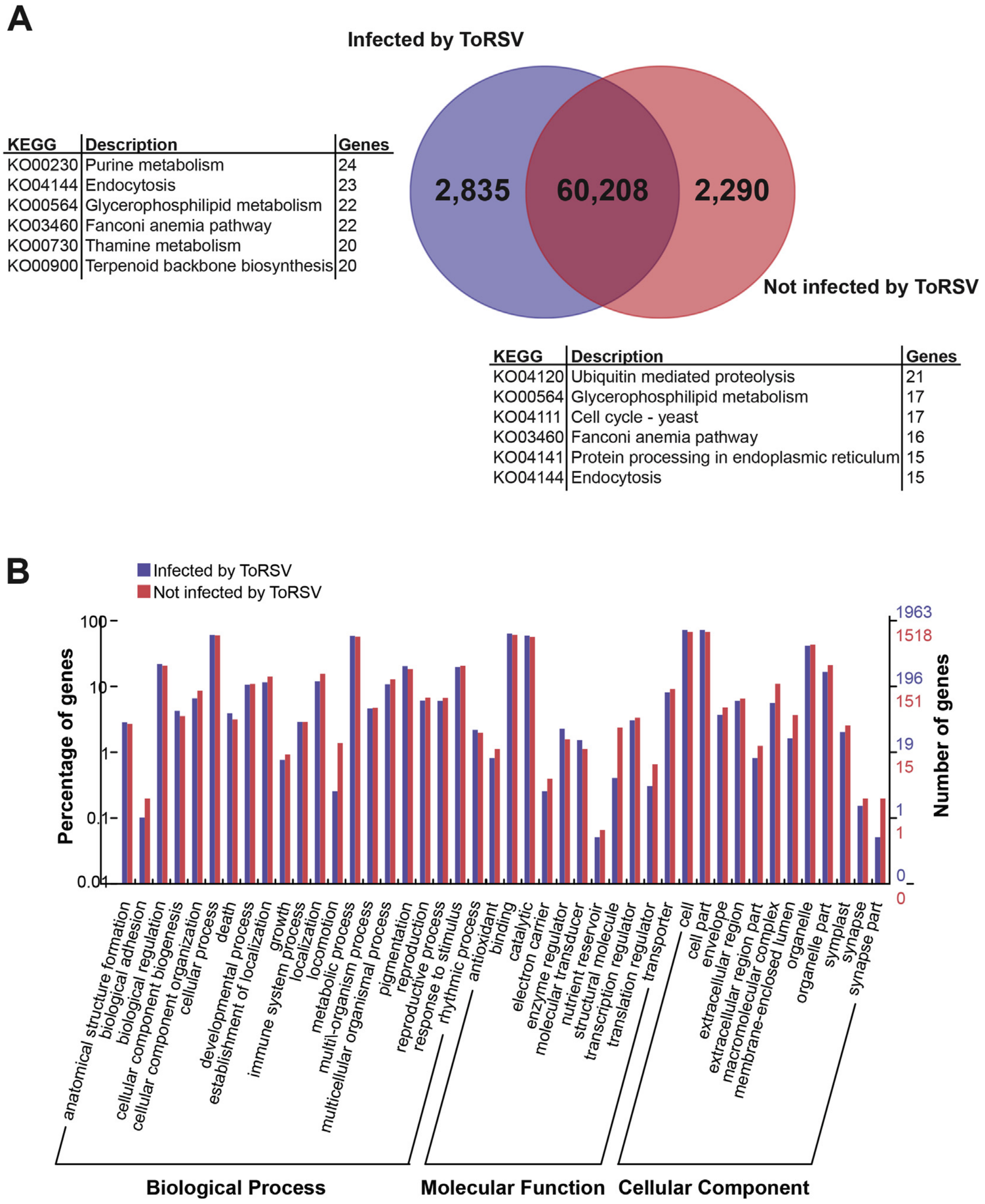


Figure 2. Pfam protein/family domain and Kyoto Encyclopedia of Genes and Genomes (KEGG) classification of the *R. idaeus* (Var. Amira) reference transcriptome. A) Pie charts showing the categorization of the *R. idaeus* (Var. Amira) reference transcriptome based on Pfam and BLAST annotations. A total of 42,189 proteins were assigned to one or more Pfam terms. Proteins annotated with Pfam hits are further subdivided into proteins with Pfam and BLAST hits and proteins with Pfam hits but without BLAST hits. **B)** The 15 most represented Pfam domain/family terms present in the *R. idaeus* (Var. Amira) reference transcriptome. **C)** The histogram shows the KEGG classification of the *R. idaeus* (Var. Amira) reference transcriptome. Proteins were classified into metabolism, genetic information processing, cellular processes, organismal systems, and human diseases.

each condition, were sequenced using Illumina technology and more than 100 million high-quality reads were obtained from each library (Table 1). We compared gene expression profiles of these two samples by mapping high-quality reads to the reference transcriptome using RSEM (Li and Dewey, 2011). A total of 62,498 and 63,043 genes were detected in infected and not infected by ToRSV raspberry samples, respectively; and 60,208 genes were expressed in both samples (Figure 3A). KEGG pathway analysis showed that genes expressed in only one condition were enriched in different biological pathways. For example, raspberry infected by ToRSV was enriched in purine metabolism (KO00230, 24 genes), endocytosis (KO04144, 23 genes), and glycerophospholipid metabolism (KO00564, 22 genes), while raspberry not infected by ToRSV



KEGG	Description	Genes
KO00230	Purine metabolism	24
KO04144	Endocytosis	23
KO00564	Glycerophosphilipid metabolism	22
KO03460	Fanconi anemia pathway	22
KO00730	Thamine metabolism	20
KO00900	Terpenoid backbone biosynthesis	20

KEGG	Description	Genes
KO04120	Ubiquitin mediated proteolysis	21
KO00564	Glycerophosphilipid metabolism	17
KO04111	Cell cycle - yeast	17
KO03460	Fanconi anemia pathway	16
KO04141	Protein processing in endoplasmic reticulum	15
KO04144	Endocytosis	15

Figure 3. Comparison of expressed transcripts between *R. idaeus* (Var. Amira) infected and not infected by ToRSV samples. A) Venn diagram showing the transcripts expressed in each samples. A total of 60,208 genes are co-expressed in infected by ToRSV and not infected by ToRSV samples. The table at the right of the Venn diagram shows the top six most represented KEGG terms in the infected by ToRSV sample, while the table at the left of the Venn diagram illustrates the top six most enriched KEGG terms in the not infected by ToRSV sample. **B).** Histogram depicts functional classification, based on GO terms, of genes expressed exclusively in both samples.

was enriched in ubiquitin-mediated proteolysis (KO04120, 21 genes), glycerophospholipid metabolism (KO00564, 17 genes), and cell cycle - yeast (KO04111, 17 genes) (Figure 3A). GO functional classification of genes expressed exclusively in each raspberry sample is shown in Figure 3B.

Given we have no biological replicates for each condition, DEGs were determined by applying two approaches: a bootstrap-based (i.e., IsoDE) and an empirical Bayesian (i.e., EBSeq) (Al Seesi et al., 2014; Leng et al., 2013). We determined a transcript as differentially expressed by filtering their fold changes (FC > 2) and p-values or FDR at 0.05. Under these criteria and using IsoDE, we found 8,705 genes with differential expression in infected by ToRSV raspberry sample when compared to the not infected by ToRSV sample. While using EBSeq, the number of DEGs between both conditions reached 8,307. Although both IsoDE and EBSeq software detected a similar number of DEGs between conditions, we have focused our further analysis only on those genes that were consistently detected as differentially expressed using both software. To know this, we constructed a Venn diagram and found that 2,340 genes were overlapped (Figure 4A). Of these DEGs, 1,512 (65%) had BLAST hit in the Swiss-Prot database, while 828 (35%) genes had no sequence similarity and seem to be taxon-restricted genes. Based on this analysis, the number of up-regulated DEGs in *R. idaeus* (Var. Amira) infected by ToRSV

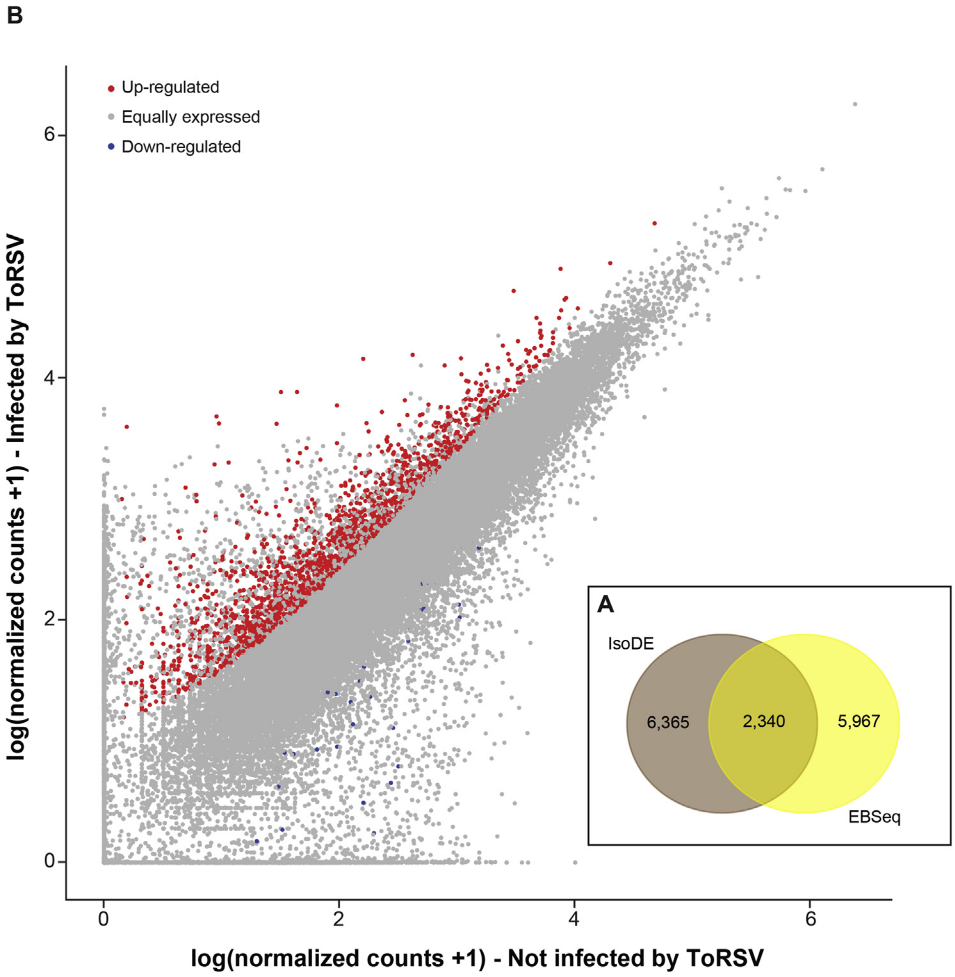


Figure 4. Differential gene expression between *R. idaeus* (Var. Amira) samples infected and not infected by ToRSV. A). Venn diagram showing the number of differentially expressed genes detected by IsoDE and EBSeq. Only those genes found as differentially expressed using both software were used in further analysis. **B)** Scatterplot comparing the gene expression levels (measured as normalized read counts) between infected and not infected by ToRSV raspberry samples. DEGs classified as up-regulated in the infected by ToRSV sample are shown in red, while down-regulated genes are depicted in blue. Equally expressed genes between both conditions are shown in grey.

raspberry sample was 2,291 (98%) and the remaining 49 DEGs (2%) were up-regulated in the not infected by ToRSV sample (Figure 4B). The normalized count data of both samples are displayed in Figure 4B, showing all DEGs that were detected by both IsoDE and EBSeq software.

Although we found a relatively high number of genes consistently detected as differentially expressed with IsoDE and EBSeq, we found a low percentage of down-regulated genes between infected and non-infected ToRSV raspberry samples (Figure 4B). This striking result might be attributed to the strict criteria used to define a gene with differential expression, but also might be explained by the different normalization methods used by each software during the process of differential expression analysis (length normalization based on gene isoforms and IsoEM algorithm in IsoDE and median normalization similar to DESeq in EBSeq).

3.3. Gene enrichment analysis of differentially expressed genes

To understand the transcriptome response of *R. idaeus* (Var. Amira) against ToRSV infection and putative metabolic pathways involved in resistance/susceptibility to ToRSV in raspberries, we focused our enrichment analysis on those genes detected as differentially expressed by both IsoDE and EBSeq software (i.e., 2,340 DEGs) (Figure 4). GO enrichment analysis identified 119 significantly overrepresented GO categories (adjusted p-value < 0.05, BH method), including 59 biological processes, 53 molecular functions, and 7 cellular components. Enriched GO terms were visualized using the Uniprot database as background and the default semantic similarity measures (Simrel), implemented in REVIGO (Supek et al., 2011). This analysis showed that biological processes associated with metabolic process, carbohydrate metabolism, transport of virus in host – tissue to tissue, acidic amino acid transport, jasmonic acid metabolic process, flavonoid biosynthetic process, wax biosynthetic process, positive regulation of translational initiation, response to wounding, lipid metabolic process, sphingolipid biosynthetic process, cellular glycan metabolic process, plant-type secondary cell wall biogenesis and maintenance of seed dormancy were significantly overrepresented among the DEGs in *R. idaeus* (Var. Amira) (Figure 5A). This analysis also showed that molecular functions related to symporter activity, inositol phosphoceramide synthase activity, jasmonate O-methyltransferase activity, transferase activity, transferring acyl groups other than amino-acyl groups, calcium-dependent phospholipid binding, beta-glucosidase activity, oxidoreductase activity, copper ion binding, receptor binding, nutrient reservoir activity, terpene synthase activity, lyase activity, heme binding, sequence-specific DNA binding transcription factor activity and hydroquinone:oxygen oxidoreductase activity were enriched among the 2,340 DEGs (Figure 5B).

We also performed Pfam and KEGG enrichment analyses to identify specific protein domains and metabolic pathways/enzymes that were significantly enriched among the DEGs. As a result, 63 Pfam domains and 22 KEGG pathways/enzymes were significantly enriched among the DEGs between *R. idaeus* (Var. Amira) samples infected and not infected by ToRSV. We found Pfam domains associated with terpene synthase, multicopper oxidase, glycosil hydrolase, transferase, ABC transporter function, as the most overrepresented proteins

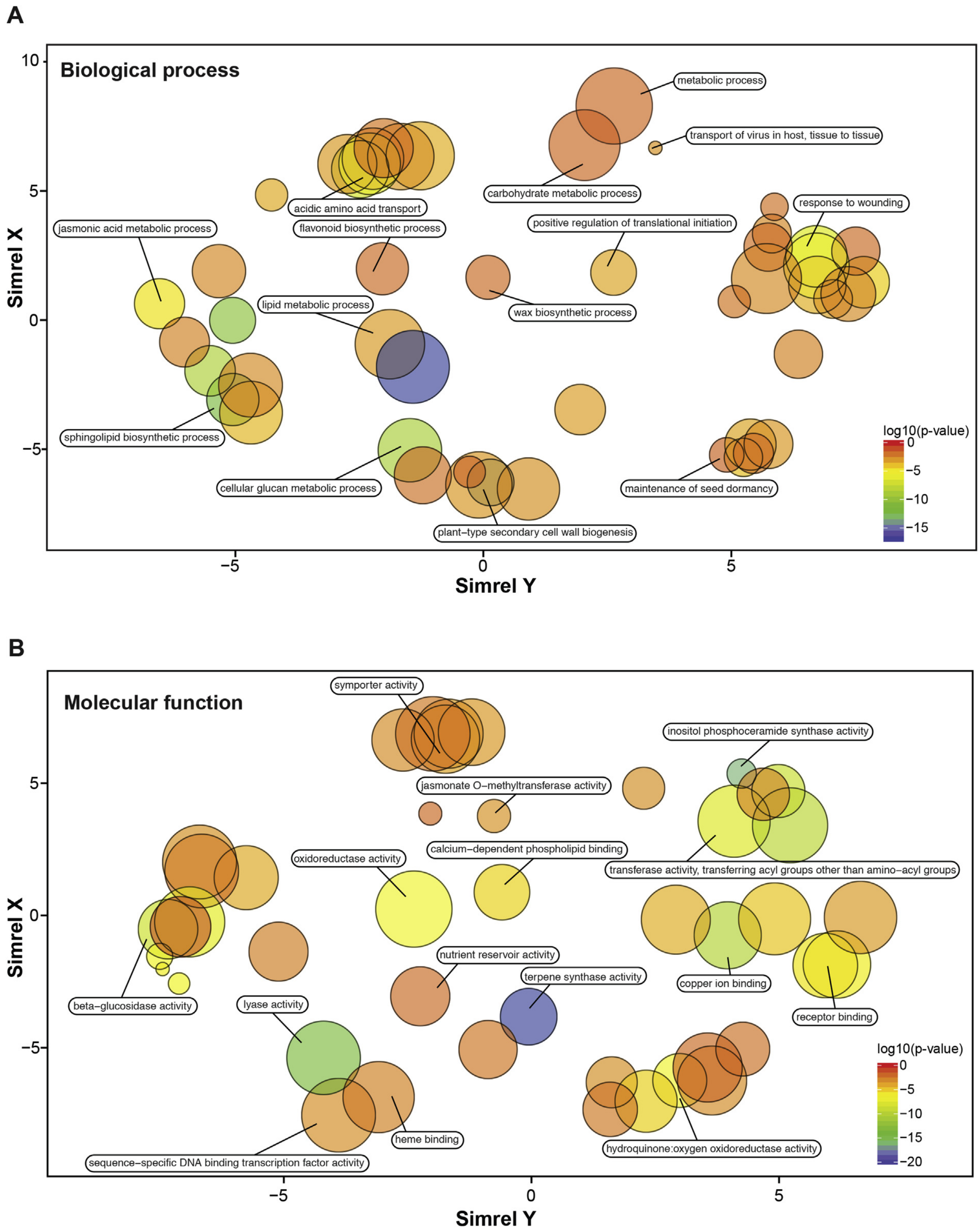


Figure 5. Gene Ontology (GO) enrichment analysis of DEGs between *R. idaeus* (Var. Amira) infected and not infected by ToRSV using REVIGO. The scatterplots show the cluster representatives (terms remaining after redundancy) in a two-dimensional space derived by applying multi-dimensional scaling to a matrix of GO terms semantic similarities. Bubble colour indicates the p-value for FDR derived from the GOseq analysis. The circle size represents the frequency of the GO term in the Uniprot database (more general terms are represented by larger size bubbles). **A)** GO enrichment under the biological process category. **B)** GO enrichment under the molecular function category.

among the DEGs (Figure 6A). In addition, we found that laccases were the most enriched enzymes among the DEGs, followed by beta-glucosidases and xyloglucan:xyloglucosyl transferases (Figure 6B). Pfam and KEGG enrichment results are consistent with GO enrichment analysis and might suggest that in the immune response of *R. idaeus* (Var. Amira) against the ToRSV participates genes/enzymes from different pathways.

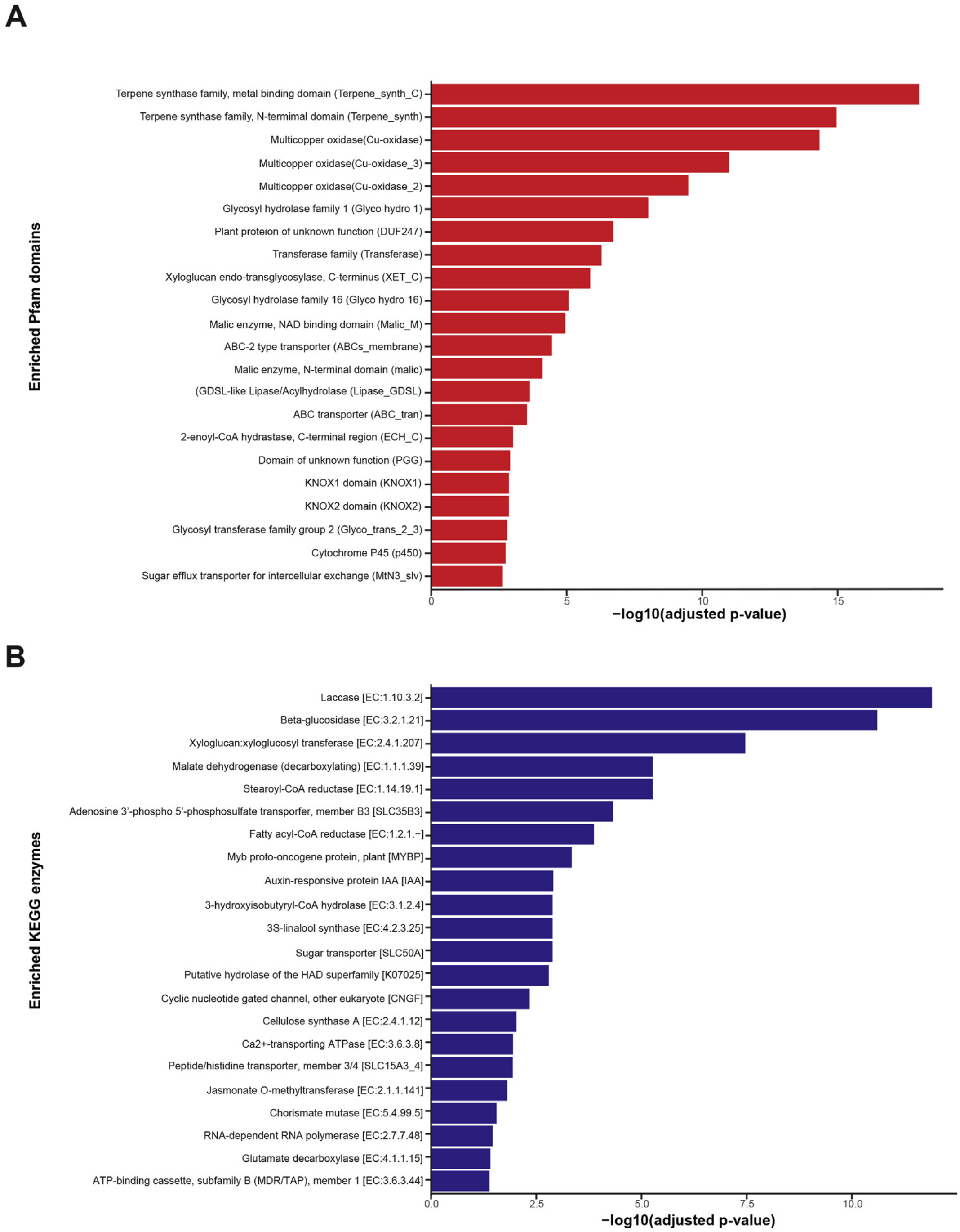


Figure 6. Pfam domain and KEGG enzyme enrichment analyses of DEGs between *R. idaeus* (Var. Amira) infected and not infected by ToRSV. A) The histogram shows the top 22 most Pfam enriched terms among the DEGs. **B)** The histogram depicts the 22 KEGG enzyme enriched among the DEGs.

3.4. Detection of SSRs and SNPs

To identify molecular markers for *R. idaeus* (Var. Amira), we performed a general screening for microsatellites (SSRs) in the raspberry reference transcriptome. This screening identified a total of 8,304 SSRs with 240 motifs, which are distributed in 7,569 transcripts. Among SSRs, di-nucleotide motifs (4,450; 53.6%) were the most abundant, followed by tri- (1,725; 20.8%), mono- (1,426; 17.2%), tetra- (414; 4.9%), hexa- (176; 2.1%), and penta-nucleotides (113; 1.4%). The top five motif repeats were (AG/CT)n [3,295; 39.7%], (A/T)n [1,405; 16.9%], (AT/AT)n [859; 10.3%], (AAG/CTT)n [765; 9.2%], and (AC/GT)n [296; 3.6%].

In addition, we searched for single-nucleotide polymorphisms (SNPs) that were present in the infected and not infected by ToRSV raspberry samples. In the infected by ToRSV raspberry sample, a total of 125,719 high-quality SNPs was found, among which 70,646 were transitions and 55,073 were transversions. In the not infected by ToRSV raspberry sample, there were 26,612 high-quality SNPs, among which 14,909 were transitions and 11,703 were transversions. The SSRs and SNPs identified in this study constitute an important and valuable genomic resource for further studies on genetic diversity and breeding programs in *R. idaeus* (Var. Amira).

4. Discussion

4.1. Characterization of the Rubus idaeus (Var. Amira) transcriptome and comparison with other Rubus species

With the development of sequencing technology, many plants such as *Arabidopsis*, rice, maize, and sorghum have had their genomes decoded (The Arabidopsis Genomie Initiative, 2000; Goff et al., 2002; Schable et al., 2009; Paterson et al., 2009). These astonishing achievements have speeded up the research on the understanding of molecular mechanisms that underlie plant growth, development, and physiology. However, for some non-model plans and minor crops, it is infeasible to carry out whole genome sequencing projects because of the expensive cost. To overcome this issue, next-generation sequencing technologies provide an opportunity to mine the genes and deepen our understanding of growth, development, and physiology in non-model plants.

In this study, we generated and annotated a comprehensive transcriptome of *Rubus idaeus* (Var. Amira) by combining deep transcriptome data from raspberries exposed to ToRSV infection and healthy raspberries. More than 200 million short reads were generated and assembled into 68,853 predicted protein-coding sequences, resulting in a high-quality transcriptome compared to that of other *Rubus* species (Garcia-Seco et al., 2015; Hyung Hyun et al., 2014). Our reference transcriptome showed better assembly statistics with longer average length of transcripts, higher ratio of annotated genes into putative functions, and higher proportion of genes assigned to the corresponding KEGG pathways, when it is compared to other *Rubus* sp. transcriptomes (Garcia-Seco et al., 2015; Hyung Hyun et al., 2014).

The functional characterization of the *R. idaeus* (Var. Amira) reference transcriptome shows that most of the transcripts were successfully assigned to different biological processes, molecular functions, and cellular components. Although no genomic information for *Rubus idaeus* and little information of Rosaceae family are available, this functional annotation and gene classification suggest that raspberry (Var. Amira) contains an extensive and diverse gene set, and also suggests that this study will provide a solid foundation for further investigations and identification of functional genes in *R. idaeus* (Var. Amira).

Protein domain analysis shows that *R. idaeus* transcriptome is characterized by the high presence of proteins with leucine-rich repeats (e.g., LRR_7, LRR_1, LRR_6, LRR_8, and LRR_4), followed by proteins with pentatricopeptide repeats (e.g., PPR_1, PPR_2, and PPR_3), and Pkinase domain-containing proteins. We also found the presence of genes encoding proteins with NB-ARC and WD40 domains. Proteins containing those domains are implicated in a variety of functions ranging from signal transduction and transcription regulation to cell cycle control to plant resistance and regulators of cell death. In plants, NB-ARC-containing proteins are thought to be associated with pathogenic recognition and subsequent activation of innate immune response (van Ooijen et al., 2008), whereas WD40-containing proteins have been related to resistance against different types of abiotic stress, as well as to the protein proteasomal degradation and damaged DNA repair pathway (Lee et al., 2010). Additionally, the presence of proteins containing RRM domains (i.e., RNA recognition motif) warns about the importance of metabolic processes that are proper of plants and their association with the post-transcriptional gene regulation (Lorković and Barta, 2002). Particularly in raspberries (e.g., *R. idaeus* Var. Amira), more studies are needed to determine the putative functions of genes containing those protein domains and their roles in immune response against pathogens and viruses.

Despite there is no comprehensive database in order to compare plant-specific metabolic pathways, we used the KEGG database and found many proteins with sequence similarity to different categories, including metabolism, genetic information processing, cellular processes, and human disease. The last category is tempting when we want to evaluate the transcriptome response in non-model plant species. In the transcriptome of *R. idaeus* (Var. Amira), we found many genes with sequence similarity to human genes, which are probably involved in the response to diseases or related to immune processes. Similarly, we also found transcripts with sequence similarity to several genes involved in different types of cancers ranging from lung, liver, colorectal, melanoma to the prostate. Taken together, this study shows that these types of similarities could be very useful in studies using RNA-Seq and plant-virus interactions in order to identify potential candidate genes involved in transcriptome immune response in cultivated fruit crops.

4.2. Differential gene expression of Rubus idaeus (Var. Amira) under ToRSV infection provides clues into the molecular mechanisms underlying immune response in raspberries

Viral infectious processes, such as the one provoked by ToRSV, can produce alterations in the metabolism of *Rubus* idaeus (Var. Amira). These changes can either be related to immune responses or to maintenance functions. In this study, we found a low number of expressed genes related to biological adhesion, nutrient reservoir and synapse part in the infected by ToRSV raspberry sample. This result might be explained by the immediate transcriptome response of *R. idaeus* (Var. Amira) against the virus by detecting the infection and waking up the immune response (Nümberger et al., 2004). On the other hand, uniquely expressed genes in the infected by ToRSV raspberry sample are associated mainly with enzyme regulator and molecular transducer functions, suggesting that the virus can influence protein regulation in *Rubus idaeus* (Var. Amira).

The identification of differentially expressed genes between both conditions, using both IsoDE and EBSeq software, suggests that these genes are differentially switched on/off due to the alterations produced by the ToRSV infection. The analysis of those DEGs suggests that complex interactions occur in *Rubus idaeus* (Var. Amira), with components of the jasmonic acid and sphingolipid biosynthesis pathways that seem to be active players in the transcriptome immune response of raspberries. This result is in conjunction with previous reports showing that the methyl salicylate and methyl jasmonate (MeJA), other components of the jasmonic acid pathway, are essential for the immune resistance against *Tobacco mosaic virus* (TMV) in *Nicotiana benthamiana* (Zhu et al., 2014). On the other hand, DEGs associated with inositol phosphoceramidas synthase activity is responsible for the formation of signal transduction complexes and secondary signal molecules that modulate the programmed cell death as part of the defensive system against fungal pathogens in plants (Mortimer et al., 2013; Mina et al., 2010). However, we cannot formally exclude the possibility that these results might be biased due to the lack of biological replicates in our RNA-Seq experiment.

Despite the limitation mentioned above, this study represents the first transcriptomic approach to provide clues about the immune response that takes place in the *R. idaeus* (Var. Amira) under ToRSV infection, giving a greater informative diagram than previous research based on candidate-gene approaches (Hyung Hyun et al., 2014; Dardick, 2007). GO enrichment of DEGs shows the prominent presence of genes associated with copper-iron binding and receptor binding. The over-representation of these genes might probably be due to their association with viral infectious processes through increasing the activity of these copper-related genes to give support to the possible ongoing metabolic deficiencies during virus infection into the plant (Marschner, 1995; Abdel-Ghany and Pilon, 2008). On the other hand, KEGG enzymes analysis on DEGs shows enrichment for laccase enzymes, which are thought to be part of the immune response to fungal necrotrophs infections (Mayer et al., 2001); however, laccase function in viral infections is still unknown. In addition, the enrichment of DEGs associated with terpene synthase activity might suggests that the plant-virus interaction generates a lack of control of the flow of ions and metabolites through the cell membrane, which ultimately produces the intervention of modulating-membrane proteins and receptors to capture the virus and restore the cell homeostasis (Wittstock and Gershenzon, 2002; Wink, 2010). However, we should highlight that further studies are needed to validate the putative functions of those DEGs.

Taken together, our study reveals that *R. idaeus* (Var. Amira) might use a complex and dynamic repertory of genes, proteins, and enzymes to tackle the immune alteration produced by ToRSV. In addition, this study provides clues into different metabolic pathways that might be involved in the immune response and susceptibility of *R. idaeus* (Var. Amira) against ToRSV.

5. Conclusions

Next-generation sequencing technology allowed for the identification of thousands of genes that are differentially expressed in *R. idaeus* (Var. Amira) infected by ToRSV. This study shows the usefulness of deep transcriptome sequencing as a powerful tool to understand differences in gene expression and their relation to biological and physiological processes in non-model plant species. The information from this study is a first attempt to provide clues in the metabolic and defensive response of raspberries against viral infections, and report valuable genomic resources, including sequence data, annotated transcripts, candidate genes and molecular markers (i.e., SSRs and SNPs), for further molecular studies in *R. idaeus* (Var. Amira).

Declarations

Author contribution statement

G. Gonzalez: Conceived and designed the experiments; Performed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

F. Aguilera: Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

V. D'Afonseca: Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was supported by FONDECYT (11121394).

Competing interest statement

The authors declare no conflict of interest.

Additional information

Raw sequencing data associated with this study has been deposited at NCBI SRA database under the BioProject accession number PRJNA354231. Additional datasets are available upon request from the corresponding author.

References

Abdel-Ghany, S., Pilon, M., 2008. MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in *Arabidopsis*. J. Biol. Chem. 283, 15932–15945.

Alice, L.A., Campbell, C.S., 1999. Phylogeny of *Rubus* (Rosaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. Am. J. Bot. 86, 81–97.

Al Seesi, S.A., Tiagueu, Y.T., Zelikovsky, A., Măndoiu, I.I., 2014. Bootstrap-based differential gene expression analysis for RNA-Seq data with and without replicates. BMC Genom. 15 (Suppl 8), S2.

Benjamin, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. 57, 289–300.

Bolder, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Cabeza, R., Koester, B., Liese, R., Lingner, A., Baumgarten, V., Dirks, J., et al., 2014. An RNA sequencing transcriptome analysis reveals novel insights into molecular aspects of the nitrate impact on the nodule activity of *Medicago truncatula*. Plant Physiol. 164, 400–411.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 421.

Dardick, C., 2007. Compararive expression profilling of *Nicotiana benthamiana* leaves systemically infected with three fruit tree viruses. Mol. Plant Microbe Interact. 20, 1004–1017.

Deighton, N., Brennan, R., Finn, C., Davies, H.V., 2000. Antioxidant properties of domesticated and wild *Rubus* species. J. Sci. Food Agric. 80, 1307–1313 (200007)80:9<1307::AID-JSFA638>3.0.CO;2-P.

Eddy, S.R., 1998. Profile hidden Markov models. Bioinformatics 14, 755–763.

Garcia-Seco, D., Zhang, Y., Gutierrez-Mañero, F.J., Martin, C., Ramos-Solano, B., 2015. RNA-Seq analysis and transcriptome assembly for blackberry (*Rubus* sp. Var. Lochness) fruit. BMC Genom. 16, 5.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., et al., 2002. A draft sequence of the rice genome (*Oryza sativa* L. spp. japonica). Science 296, 92–100.

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 (7), 644–652.

Griesbach, J., 1995. Tomato ringspot virus by polymerase chain reaction. Plant Dis. 79, 1054–1056.

Hummer, K., Janick, J., 2009. Rosaceae: Taxonomy, Economic Importance, Genomics. Genetics and Genomics of Rosaceae. Springer, New York, pp. 1–17.

Hyung Hyun, T., Lee, S., Kumar, D., Rim, Y., Kumar, R., Yeol Lee, S., et al., 2014. RNA-seq analysis of *Rubus idaeus* cv. Nova: transcriptome sequencing and de novo assembly for subsequent functional genomics approaches. Plant Cell Rep. 33, 1617–1628.

Jean-Gilles, D., Li, L., Ma, H., Yuan, T., Chichester, C., Seeram, N., 2012. Anti-inflammatory effects of polyphenolic-enriched red raspberry in an antigen induced arthritis rat model. J. Agric. Food Chem. 60, 5755–5762.

Jennings, D., 1988. Raspberries and Blackberries: Their Breeding, Diseases and Growth. Academic Press, London.

Kähkönen, M.P., Hopia, A.I., Heinonen, M., 2001. Berry phenolics and their antioxidant activity. J. Agric. Food Chem. 49, 4076–4082.

Kakumanu, A., Ambavaram, M.M., Klumas, C., Krishnan, A., Batlang, U., Myers, E., et al., 2012. Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. Plant Physiol. 160, 846–867.

Kalkman, C., 2004. Rosaceae. Flowering Plants - Dicotyledons. Springer, Berlin Heidelberg, pp. 343–386.

Kamenetsky, R., Faigenboim, A., Shemesh Mayer, E., Michael, T.B., Gershberg, C., Kimhi, S., et al., 2015. Integrated transcriptome catalogue and organ-specific profiling of gene expression in fertile garlic (*Allium sativum* L.). BMC Genom. 16, 12.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.-H., Rognes, T., Ussery, D.W., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100–3108.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25.

Lee, S., Lee, J., Paek, K.-H., Kwon, S.-Y., Cho, H.S., Kim, S.J., et al., 2010. A novel WD40 protein, BnSWD1, is involved in salt stress in *Brassica napus*. Plant Biotechnol. Rep. 4, 165–172.

Leng, N., Dawson, J.A., Thomson, J.A., Ruotti, V., Rissman, A.I., Smits, B.M.G., et al., 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-Seq experiments. Bioinformatics 29, 1035–1043.

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 12, 323.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993.

Lorković, Z.J., Barta, A., 2002. Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plan *Arabidopsis thaliana*. Nucleic Acids Res. 30, 623–635.

Marschner, H., 1995. Mineral Nutrition of Higher Plants. Academic Press, London.

Martin, R.R., MacFarlane, S., Sabanadzovic, S., Quito, D., Poudel, B., Tzanetakis, I.E., 2013. Viruses and virus diseases of *Rubus*. Plant Dis. 97, 168–182.

Mayer, A.M., Staples, R.C., Gil-ad, N.L., 2001. Mechanisms of survival of necrotrophic fungal plant pathogens in hosts expressing the hypersensitive response. Phytochemistry 58, 33–41.

Medina, C., Matus, J., Zúñiga, M., San-Martín, C., Arce-Johnson, P., 2006. Occurrence and distribution of viruses in commercial plantings of *Rubus, Ribers* and *Vaccinium* species in Chile. Cienc. Investig. Agrar. 33, 23–28.

Mina, J.G., Okada, Y., Wansadhipathi-Kannangara, N.K., Pratt, S., Shams-Eldin, H., Schwarz, R.T., et al., 2010. Functional analysis of differentially expressed isoforms of the Arabidopsis inositol phosphorylceramide synthase. Plant Mol. Biol. 73, 399–407.

Morales, C., González, M., Hirzel, J., Riquelme, J., Herrera, G., Madariaga, M., et al., 2009. Aspectos relevantes en la producción de frambuesa (*Rubus idaeus* L.). Boletín INIA. 192, 118.

Mortimer, J.C., Yu, X., Albrecht, S., Sicilia, F., Huichalaf, M., Ampuero, D., et al., 2013. Abnormal glycosphingolipid mannosylation triggers salicylic acid-mediated responses in *Arabidopsis*. Plant Cell 25, 1881–1894.

Nümberger, T., Brunner, F., Kemmerling, B., Piater, L., 2004. Innate immunity in plants and animals: striking similarities and obvious differences. Immunol. Rev. 198, 249–266.

O'Rourke, J.A., Bolon, Y.-T., Bucciarelli, B., Vance, C.P., 2014. Legume genomics: understanding biology through DNA and RNA sequencing. Ann. Bot. 113, 1107–1120.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al., 2009. The *Sorghum bicolor* genome and the diversification of grasses. Nature 457, 551–556.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8 (10), 785–786.

Requena, A., Requena, M., Ezziyyani, M., Candela, M., 2007. Virosis en los principales cultivos hortícolas de la región Murcia. Horticul. 25, 12–20.

Schable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., et al., 2009. The B73 maize genome: complexity, diversity, and dynamics. Science 326, 1112–1115.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assenbly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.

Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO summarizes and visualized long list of Gene Ontology terms. PloS One 6, e21800.

The Arabidopsis Genomie Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thalliana*. Nature 408, 796–815.

van Ooijen, G., Mayr, G., Kasiem, M.M.A., Albrecht, M., Cornelissen, B.J.C., Takken, F.L.W., 2008. Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. J. Exp. Bot. 59, 1383–1397.

Wink, M., 2010. Annual Plant Reviews: Functions and Biotechnology of Plant Secondary Metabolites. Blackwell Publishing Ltd.

Wittstock, U., Gershenzon, J., 2002. Constitutive plant toxins and their role in defense against herbivores and pathogens. Curr. Opin. Plant Biol. 5, 300–307.

Ye, Y., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al., 2006. WEGO: a wel tool for plotting GO annotations. Nucleic Acids Res. 34, W293–W297.

Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A., 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 11, R14.

Yousefi, G., Yousefi, S., Emam-Djomeh, Z., 2013. A comparative study on different concentration methods of extracts obtained from two raspberries (*Rubus idaeus* L.) cultivars: evaluation of anthocyanins and phenolics contents and antioxidant activity. Int. J. Food Sci. Technol. 48, 1179–1186.

Zheng, Y., Zhao, L., Gao, J., Fei, Z., 2011. iAssembler: a package for *de novo* assembly of Roche-454/Sanger transcriptome sequence. BMC Bioinf. 12, 453.

Zhu, F., Xi, D.-H., Yuan, S., Xu, F., Zhang, D.-W., Lin, H.-H., 2014. Salicyclic acid and jasmonic acid are essential for systemic resistance against *Tobacco mosaic virus* in *Nicotiana benthamiana*. Mol. Plant Microbe Interact. 27, 567–577.