WRKY genes in black raspberry (Rubus occidentalis L.): duplicate and conquer

Winder Felipez
Universidade Federal de Pelotas
Camila Pegoraro
Universidade Federal de Pelotas
Luciano Maia
Universidade Federal de Pelotas
Antonio Costa de Oliveira (acostol@terra.com.br)
Universidade Federal de Pelotas

Research Article

Keywords: phylogeny, gene structure, subcellular location, expression regulation, differential expression

Posted Date: October 31st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2193697/v1

Abstract

WRKY transcription factors regulate several biological processes in plants, including responses to biotic stresses and tolerance to abiotic stresses, and are part of a gene superfamily in higher plants. There are many studies on the functions of WRKY proteins in several model species, aiming at identification and functional characterization, but there has not yet been a comprehensive analysis of the RoWRKY protein family in black raspberry (*Rubus occidentalis* L.) as shown here. In this study, the investigation of the complete genome of the black raspberry identified 62 *RoWRKY* genes that were evaluated and are unevenly distributed in all seven chromosomes. The proteins encoded by these genes were classified into four groups (I, II, III and IV), with those of group II divided into five subgroups (IIa - IIe) based on their conserved domains and zinc finger domain types. Motif analysis showed that all *RoWRKY*s contained one or two WRKY domains and that proteins from the same group had similar motif compositions. Five pairs of *RoWRKY* genes in segmental duplication and two pairs in tandem duplication were detected. Analysis of the structure of *RoWRKY* genes showed that they have 1–11 introns, with most *RoWRKY* genes consisting of two introns and three exons. A cis element analysis showed that all promoters of the *RoWRKYs* genes contain at least one cis stress-response element. Differential expression analysis of 10 samples of RNA-seq data, reviewed *RoWRKY* genes from black raspberry, show preferential or specific expression in tissue samples. These findings provide a complete overview of the evolution and modification of the RoWRKYs protein family, which will help the functional characterization of these proteins in the response to biotic and abiotic stresses of black raspberry.

Key Message

The 59 RoWRKY genes were classified into three groups, I, II, II and VI, and the group II was further divided into five subgroups. Differential expression analysis of 10 samples of RNA-seq data shows preferential or specific expression in tissue samples.

Introduction

The black raspberry (*Rubus occidentalis* L.) is native to eastern North America. The species is diploid (2n = 2x = 14) of the subgenus Idaeobatus (VanBuren et al. 2018a). Breeding of this species began in the early 21st century at the USDA ARS New York State Agricultural Experiment Station in Corvallis, OR and in Beltsville, MD, and with New Zealand HortResearch Inc (Jennings 2018). Molecular improvement for fruit traits was hampered by the lack of a high quality reference genome (Jibran et al. 2018). However, an almost complete chromosome-scale assembly of the black raspberry genome is now available (VanBuren et al. 2018b). The updated high-quality reference genome in black raspberry allows for comparative genomic studies and gene analyses, including those encoding transcription factors (TF) (Jiang et al. 2021). It also enables the application of marker-assisted breeding and genomic selection.

The high content of anthocyanins and ellagitannins present in black raspberry are important in human health due to anticancer activity, which has led to an increase in the production of the species in the last 15 years (Kula and Krauze-Baranowska 2016; Bushakra et al. 2018). In commercial production of black raspberry, there are disease-causing organisms, such as the aphid *Amphorophora agathonica* Hottes, which is a vector of the raspberry mosaic virus complex (Bushakra et al. 2015). In addition to this biotic factor, abiotic factors impact the species, such as the high temperatures that occur in the natural temperate habitats of black raspberry, requiring it to tolerate or adapt to local environmental conditions of production (Bradish et al. 2016). The interactions of the multiple biotic and abiotic stresses affecting plants need to be understood, from the molecular mechanisms of stress response and tolerance, to identifying and characterizing the genetic components such as the transcription factors involved in these mechanisms, including WRKYs, which mediate the stress response in different species (Yoon et al. 2020).

The FT WRKYs superfamily plays a significant role in various biotic and abiotic stress responses (Meng et al. 2016). In apple (*Malus x domestica*) culture *MdWRKY40a* and *MdWRKY54h* play negative roles in the defense against infection by the *B. dothidea* pathogen. On the other hand, *MdWRKY40*, *MdWRKY60* and *MdWKRY33s* may play important roles in the response to pathogens and are conserved in some angiosperm plants (Zhang et al. 2021). In the pear (*Pyrus ussuriensis*) culture, *PuWRKY31* induces sucrose levels in the fruit (Li et al. 2020b). WRKYs are considered defenders in cultures *Oryza sativa japonica*, *Populus trichocarpa*, *Arabidopsis thaliana*, and *Physcomitrella patens* (He et al., 2010). It is evident that the WRKY superfamily plays an important role in the plant responses to and tolerate stress through molecular signaling mechanisms.

The primary structure of FT WRKYs is composed of the WRKY domain and the zinc finger motif, a 60 amino acid region that is highly conserved by family members. These proteins are transcriptional regulatory factors with preferential binding to the cis W-box element and have the potential to regulate the expression of a variety of target genes (Eulgem et al. 2000). The classification of the superfamily depends on the number of domains and their characteristic zinc finger motif, leading to a subclassification of three subfamilies or groups. Group I is composed of TFs that have two WRKY domains, while group II (II-a, b, c, d, e) and III contain TFs with only one WRKY domain. However, group I and II TFs have the same zinc finger patterns C2 - H2 (C - X4–5 - C - X22–23 - H - X1 – H) and group III TFs have a zinc finger motif C2-HC binder (C - X7 - C - X23 - H - X1 – C) pattern (Eulgem et al. 2000; Li et al. 2020a). The evolution of FT WRKYs in plants started from a structure similar to the zinc finger motif, and lateral gene transfer between plants and fungi may have taken them outside the plant kingdom. Also, it is possible to transfer WRKY genes from algae to a primitive fungus (Rinerson et al. 2015). By portions of the zinc finger motif, the proteins WRKY, GCm1, and flYWCH likely share a common ancestor in evolution (Rinerson et al. 2016).

The multiple functions of transcription factors, including WRKYs, can simultaneously control several pathways in plants subjected to stress conditions, becoming a potential tool for stress tolerance manipulation (Hrmova and Hussain 2021). Therefore, it is important to identify the WRKY proteins, as well as their subcellular and chromosomal gene location, in addition to the phylogenetic analysis, classification and identification of conserved motifs in the complete genome of the black raspberry (*Rubus occidentalis*). This characterization will provide a complete overview of the evolution and modification of the WRKY protein family and will help to determine the functional features of black raspberry *RoWRKY* genes in response to biotic and abiotic stresses.

Materials And Methods

Identification of WRKY proteins in the black raspberry genome

The entire genome sequence of the black raspberry (*Rubus occidentalis* Whole Genome v3.0 Assembly & Annotation) was downloaded from the Rosaceae Genomes Database (GDR)) (VanBuren et al. 2018b). To assemble the possible candidate amino acid sequences of black raspberry WRKY, the HMM model of the WRKY domain (PF03106) was downloaded from Pfam (http://pfam.xfam.org/family/PF03106/hmm) (Finn et al. 2014). Hmmer software (Finn et al. 2011) was also used for similarity research in the annotated proteins in black raspberry using $1e^{-3}$ as the upper limit of the e-value. Also BLASTp were performed with WRKY sequences from *Arabidopsis thaliana* (Rushton et al. 2010) with a value of e^{-10} on the black raspberry proteome. All protein sequences obtained were examined for the presence of the WRKY domain using the Web CD-Search Tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Lu et al. 2020). The ExPASY ProtParam (https://web.expasy.org/protparam/) (Gasteiger et al. 2003) was used to predict the isoelectric point (pI), molecular weight (MW) and grand mean hydrophobicity (GRAVY) of each *RoWRKY*.

Subcellular Localization Of Wrky Family Transcription Factors

Subcellular localization was predicted using WoLF PSORT (https://wolfpsort.hgc.jp/) (Horton et al. 2007) and TargetP-2.0 (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al. 2007).

Analysis of cis elements in RoWRKY gene promoters

For each *RoWRKY* gene, a 2000 bp sequence upstream of the start codon was retrieved from the black raspberry genome (*Robus occidentalis* Genome v3.0 Assembly) by applying integrative genomics visualization -IGV (Robinson et al. 2011). This sequence was submitted to the PlantCARE website to investigate cis-acting regulatory elements (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al. 2002).

Multiple Alignment And Phylogenetic Analyses

A phylogenetic tree was constructed to compare black raspberry WRKY proteins. Multiple alignment of WRKY protein sequences was performed with ClustalW software using standard parameters (Larkin et al. 2007). The phylogenetic tree was constructed using BEAST v.2.5 software (Bouckaert et al. 2019), with the UPGMA clustering method (Sneath and Sokal 1973). A bootstrap analysis was conducted using 10,000,000 replicates and evolutionary distances were calculated using the JTT matrix-based method (Jones et al. 1992).

Analysis Of Gene Structure And Identification Of Conservation Of Motifs

To investigate the diversity and structure of *RoWRKY* family members, genomic sequences for their exon/intron were used and plotted on TBtools (Chen et al. 2020a), based on black raspberry genome annotation information (*Rubus occidentalis* whole genome assembly v3.0) RoWRKY protein sequences were used to identify conserved motifs, Expectation Maximization Tool for Motive Elicitation MEME (https://meme-suite.org/meme/tools/meme) (Bailey et al. 2006). The parameters were as follows: number of repetitions: any; maximum number of reasons: 10; and optimal motif widths: 8 to 50 amino acid residues. The functions and locations of the reasons were queried in the databases MOTIF Search (https://www.genome.jp/tools/motif/) (Kanehisa and Goto 2000) and ELM (http://elm.eu.org/) (Kumar et al. 2022), a resource of the Eukaryotic Linear Motif for functional protein sites.

Chromosomal Localization And Gene Duplication

The chromosomal location image of the *RoWRKY* genes was generated by the TBtools software (Chen et al. 2020a), according to chromosomal position information provided in the genomic database for Rosaceae-GDR (https://www.rosaceae.org). To identify tandem and segmental duplications, two genes of the same species, located in the same clade of the phylogenetic tree, were defined as co-paralogs. The GDR browser was consulted in order to detect segmental duplication of the target genes (VanBuren et al. 2018b). The local alignment of two protein sequences was calculated using the Smith-Waterman algorithm (http://www.ebi.ac.uk/Tools/psa/).

Transcriptomic analysis of RoWRKY genes in black raspberry

The expression patterns of *RoWRKYs* genes were analyzed based on published RNA-seq data on NCBI BioProject ID PRJNA430858 (VanBuren et al. 2018b). For the analysis of differential expression, 10 samples of RNA-seq data were collected that were sequenced by the Illumina HiSeq4000, which are: canes (SRR7274864), root (SRR7274865), young leaf (SRR7274866), mature leaf (SRR7274867), red fruit (SRR7274868), ripened fruit (SRR7274869), flower bud (SRR7274870), green fruit (SRR7274871), leaf of post 24hr jasmonate spraying (SRR7274872), leaf of post 48hr jasmonate spraying (SRR7274873).

Data processing was performed in three steps. a) quality control and adjustment of samples: SRA toolkit (Leinonen et al. 2011) was used to download the data samples, FastQC (Wingett and Andrews 2018) was employed to analyze and visualize the quality of readings, Trimmomatic ver. 0.39 (Bolger et al. 2014) was applied to remove the low quality and library adaptors; b)The reads were mapped against the Black raspberry (*R. occidentalis* v3.0) reference genome (VanBuren et al. 2018b) using the software STAR (Dobin et al. 2013). In the next step, (c) for the counting of total reads aligned by gene in the

different libraries, FeatureCounts (Liao et al. 2014) was used. The quantification (d) was performed using the packages limma, EdgerR and DESeq2 in the R software (Law et al. 2018). In this protocol, a CPM normalization method (counts per million) was used, which are counts scaled by the total number of reads and the expression of *RoWRKYs* genes per library. A heatmap was produced using TBTools (Chen et al. 2020a), then a multidimensional scale graph (MDS) was generated to verify the repeatability of the sample and the general difference between the samples, a graph of sample mean variance (MeanVar) and Biological Variation Coefficient (BCV) as a function of gene abundance (in log2 counts per million).

Results

Identification Of The Wrky Protein Family In Black Raspberry

The statistical alignment of the Hidden Markov Model (HMMsearch) in HMMER and the BLASTp local alignment of the black raspberry proteome, was used to identify a total of 62 WRKY transcription factors in this species, and 59 *RoWRKY* genes were characterized with complete heptapeptide (WRKYGQK) domains and based on their nomenclature of the WRKY protein family (PF03106) (Table 1).

Three sequences without a complete heptapeptide domain (WRKYGQK), including variations in the heptapeptides (WGKYGQM, WGKYCQM and WGKYGVM) were discarded and not analyzed. The length of the peptide ranged from 170 amino acid residues (aa) (RoWRKY49) to 1572 aa (RoWRKY24). The coding sequences (CDS) ranged from 513 to 4887 nucleotides. The molecular weight of the predicted proteins ranged from 18648.66 (RoWRKY49) to 179205.2 kDA (RoWRKY24). The isoelectric point ranged from 4.75 (RoWRKY58) to 9.96 (*RoWRKY27*). The large mean hydrophobicity ranged from −0.2695 (RoWRKY55) to -1.186 (RoWRKY33).

Table 1
Information on *RoWRKYs* genes in the black raspberry genome

Transcript ID	*Gene name*	*Group*	*Chr.*	*StartPos*	*EndPos*	*Strand*	*CDS*	*aa*	*MW*	*pI*	*GRAVY*
Ro01_G04143	RoWRKY1	IId	1	8004614	8006199	-	1361	320	34741.1584	9.650610924	-0.484375
Ro01_G05057	RoWRKY2	I	1	486462	491212	-	1954	521	57813.4867	5.52670536	-0.854318618
Ro01_G11245	RoWRKY3	IIb	1	21623505	21626419	-	1836	456	49593.0792	7.005485344	-0.570394737
Ro01_G11672	RoWRKY4	IIc	1	12667509	12669853	-	672	223	25066.263	5.71398983	-0.838565022
Ro01_G14731	RoWRKY5	IIc	1	13729330	13731052	-	1152	313	34833.4001	5.932592583	-0.82971246
Ro01_G29921	RoWRKY6	IIc	1	12864681	12868801	+	983	187	21736.4665	9.41781559	-0.805882353
Ro02_G00571	RoWRKY7	IId	2	3604411	3606236	-	1620	325	35382.8985	9.526380348	-0.529846154
Ro02_G00778	RoWRKY8	IId	2	2276033	2278868	-	2385	337	36622.19	9.510843468	-0.497329377
Ro02_G02477	RoWRKY9	IIa	2	201460	203052	-	1415	282	31663.0201	8.819355583	-0.778014184
Ro02_G03986	RoWRKY10	IId	2	9397448	9400257	+	1154	271	30042.7428	9.627789116	-0.72804428
Ro02_G34805	RoWRKY11	IIa	2	214719	216547	-	1437	355	38949.3513	6.94552021	-0.550422535
Ro02_G35659	RoWRKY12	III	2	30494983	30497277	-	1896	631	70604.8016	5.874503136	-0.393343899
Ro02_G35662	RoWRKY13	III	2	30492753	30493624	-	585	194	22223.8089	6.515703773	-0.809278351
Ro03_G05285	RoWRKY14	IIe	3	40023145	40025358	-	2024	276	30439.28	5.251832008	-0.772463768
Ro03_G10658	RoWRKY15	I	3	3721286	3724842	+	2124	693	76270.1518	8.919861794	-0.806349206
Ro03_G12091	RoWRKY16	I	3	2733850	2738731	-	2991	723	77939.6323	5.872741127	-0.793914246
Ro03_G13413	RoWRKY17	IIc	3	39334276	39336694	-	1842	350	38586.0963	6.975815392	-0.847142857
Ro03_G13495	RoWRKY18	IIb	3	42815055	42817810	-	2150	570	61799.6277	6.874641991	-0.724210526
Ro03_G15815	RoWRKY19	IIe	3	37586963	37590391	-	1817	433	46181.6653	8.063786125	-0.629792148
Ro03_G17200	RoWRKY20	IIb	3	27707338	27710148	+	1888	545	59648.1746	6.195529747	-0.555045872
Ro03_G17983	RoWRKY21	I	3	26713247	26721585	-	3642	503	56039.149	8.066300392	-0.927833002
Ro04_G00011	RoWRKY22	I	4	3328160	3331146	-	2180	592	65356.0227	8.263122368	-0.980236486
Ro04_G00111	RoWRKY23	IIb	4	2758434	2761817	+	1803	600	65399.3516	5.967889595	-0.7185
Ro04_G02811	RoWRKY24	I	4	5449926	5456804	-	4719	1572	179205.2	5.528524208	-0.398600509
Ro04_G07106	RoWRKY25	IIc	4	6800803	6802632	-	798	188	21409.737	9.517677116	-0.887765957
Ro04_G15164	RoWRKY26	IIc	4	3463308	3467140	+	2390	330	35681.5731	5.915370369	-0.777575758
Ro04_G17961	RoWRKY27	IId	4	27718283	27719905	-	1423	300	32684.5354	9.965538979	-0.676333333
Ro04_G19918	RoWRKY28	IIc	4	8611791	8614860	-	1426	420	46246.6665	6.710604668	-0.841190476
Ro05_G08626	RoWRKY29	IId	5	9635829	9637625	+	1379	311	34817.1475	9.299129295	-0.55659164
Ro05_G10257	RoWRKY30	IIb	5	30858965	30861918	+	1920	545	60508.7653	5.344138527	-0.903119266
Ro05_G16431	RoWRKY31	IIb	5	1618499	1620470	+	1311	436	47880.6458	8.442795753	-0.853440367
Ro05_G16799	RoWRKY32	IIe	5	7328837	7329918	-	918	305	34813.7713	5.158559227	-0.698360656
Ro05_G19154	RoWRKY33	IIe	5	40543638	40545702	-	972	281	32458.202	4.758867455	-1.186120996
Ro05_G22566	RoWRKY34	III	5	6701342	6705195	+	3492	359	40457.2833	5.121784401	-0.774930362
Ro06_G05663	RoWRKY35	I	6	29647066	29649427	+	1950	512	56672.2893	7.647709084	-0.926367187
Ro06_G06104	RoWRKY36	IV	6	31287949	31290314	+	1131	290	32426.7799	5.372558022	-0.765862069
Ro06_G06729	RoWRKY37	IIb	6	21395792	21398705	+	1964	547	59853.4546	5.125251579	-0.789396709
Ro06_G09223	RoWRKY38	I	6	24597005	24601526	+	3559	697	76828.2192	5.698643303	-0.717647059
Ro06_G14746	RoWRKY39	III	6	29132792	29133812	+	624	207	24062.7768	6.304603767	-0.607729469

aa = amino acid, pI = isoelectric point MW = molecular weight, GRAVY = large average hydrophobicity

Ro06_G14747	RoWRKY40	III	6	29142106	29144027	+	1390	327	37108.7447	5.788051033	-0.785015291
Ro06_G16328	RoWRKY41	IIb	6	219293	222133	-	1920	542	59131.3352	5.996252251	-0.591881919
Ro06_G16671	RoWRKY42	IIa	6	9075470	9078186	-	1922	321	35597.4632	8.715174675	-0.799688474
Ro06_G17302	RoWRKY43	I	6	20699892	20703187	-	2033	471	51291.2415	6.417258644	-0.91104034
Ro06_G17579	RoWRKY44	I	6	2626048	2629135	-	1605	534	58259.8364	8.345384026	-0.873782772
Ro06_G24980	RoWRKY45	IIe	6	7384224	7389857	+	4887	261	28526.367	5.052952385	-0.695785441
Ro06_G28717	RoWRKY46	I	6	23986840	23989927	+	1605	534	58259.8364	8.345384026	-0.873782772
Ro06_G29026	RoWRKY47	I	6	32246811	32249876	-	1653	550	60624.6455	8.95589962	-0.802
Ro07_G04379	RoWRKY48	IIe	7	2159617	2161306	-	1442	338	36994.7226	8.400568962	-0.704142012
Ro07_G04383	RoWRKY49	IV	7	2144733	2146053	-	513	170	18648.7	7.986939812	-0.768823529
Ro07_G04678	RoWRKY50	III	7	37004358	37007280	-	1506	349	39329.0047	5.839490318	-0.855873926
Ro07_G04679	RoWRKY51	III	7	37001846	37004086	+	1050	349	39070.9602	6.004550743	-0.746704871
Ro07_G06810	RoWRKY52	III	7	38645734	38647993	+	1961	378	41617.712	6.172964668	-0.628835979
Ro07_G14025	RoWRKY53	IIe	7	38958476	38960678	-	2025	428	46046.8154	5.270759392	-0.73271028
Ro07_G14616	RoWRKY54	IIc	7	6272668	6276945	-	1247	245	28144.4658	8.533244896	-1.088979592
Ro07_G17080	RoWRKY55	III	7	16277554	16284077	+	4795	1467	165921.3356	5.741954613	-0.269529652
Ro07_G18694	RoWRKY56	III	7	1633346	1636563	+	2290	353	38941.6711	5.376309395	-0.71529745
Ro07_G33471	RoWRKY57	III	7	5076284	5084214	+	846	281	31792.4618	4.893234825	-0.407117438
Ro07_G33472	RoWRKY58	III	7	5053673	5055956	+	1185	394	44060.196	4.752046776	-0.689593909
Ro07_G33475	RoWRKY59	III	7	5100454	5110794	+	3180	1059	119778.1416	4.926940346	-0.386968839
Ro02_G04519			*2*	*4715117*	*4725115*	*-*	*7086*	*965*	*107885.98*	*5.567743*	*-0.636477*
Ro06_G09963			*6*	*19044339*	*19046664*	*+*	*1314*	*320*	*35747.761*	*5.551771*	*-0.835*
Ro06_G28795			*6*	*25843863*	*25847599*	*-*	*3004*	*306*	*33939.584*	*9.27605*	*-0.51634*

aa = amino acid, pI = isoelectric point MW = molecular weight, GRAVY = large average hydrophobicity

Subcellular localization of the RoWRKY protein family

Wolf SORT evaluation revealed that within the 59 RoWRKY proteins, all have transport sequences targeting the nucleus, 26 the chloroplast, 15 the cytosol, nine the plasma membrane, eight the cytoskeleton, seven the vacuolar membrane, five the mitochondria, five have extracellular targets, three the cytosol-nucleus, two the peroxisome, one the cytoskeleton-nucleus, one the endoplasmic reticulum and none the Golgi complex (Supplementary S1).

TargetP analyses revealed that four RoWRKY proteins present transport sequences targeted to mitochondria (RoWRKY12 = 0.01; RoWRKY21 = 0.29; RoWRKY24 = 0.08; RoWRKY55 = 0.029), six to chloroplast (RoWRKY6 = 0.02; RoWRKY12 = 0.04; RoWRKY21 = 0.01, RoWRKY2*4* = 0.12; RoWRKY55 = 0.0014; RoWRKY57 = 0.001), three to thylakoid luminal transit peptide (RoWRKY7 = 0.01; RoWRKY26 = 0.02; RoWRKY = 0,001), two signal peptide in the secretory pathways of the cell (RoWRKY6 = 0.03; RoWRKY57 = 0.007) and 59 in other compartments, where values indicate score (0.00–1.00) and reliability class (1–5; the best class is 1).

Chromosomal localization and RoWRKY gene duplication

The 59 *RoWRKY* genes were distributed in seven chromosomes of the black raspberry genome. Among them, chromosome 6 had the highest density of *RoWRKY* genes with 13 members; chromosome 7 had the second highest density with 12 genes; eight *RoWRKY* genes were located on chromosome 3; seven genes were located on each of chromosomes 2 and 4, and six genes were located on each of chromosomes 1 and 5 (Fig. 2).

The *RoWRKY* family showed two types of gene duplication (Supplementary S3). Segmental duplication shows 20 pairs of paralogous genes detected, of which, five pairs of *RoWRKY* genes contain more than 70% similarity (*RoWRKY40-RoWRKY39* = 100%; *RoWRKY12- RoWRKY13* = 82.0%; *RoWRKY24-RoWRKY22* = 75,5%; *RoWRKY46-RoWRKY44* = 100%; *RoWRKY54-RoWRKY6* = 73,2%). The tandem duplication presents 12 pairs of paralogous genes, of which, two pairs of *RoWRKY* genes were detected at a distance of less than 100kb on chromosome 2 (*RoWRKY57-RoWRKY59* = 75.2% similarity) and on chromosome 3 (*RoWRKY12-RoWRKY13* = 82%).

Number of transcripts and expression patterns of RoWRKY genes in various tissues

Of the total of 33,252 genes annotated in the black raspberry genome, the transcriptional expression of genes for each sample was: canes (10,161 = 31%), flower bud (9,390 = 28%), green leaf (10,166 = 31%), mature leaf (11,603 = 35%), leaf MJ24(10,548 = 32%), leaf MJ48 (10,872 = 33%), red fruit (11,603 = 35%), ripened fruit (11,916 = 36%), root (10,161 = 31%) and young leaf (10,391 = 31%). However, the expression of *RoWRKY* genes for the samples were: canes (44 = 0.43%), flower bud (42 = 0.45%), green leaf (41 = 0.40%), mature leaf (41 = 0.40%), leaf MJ24 (41 = 0.39%), leaf MJ48 (42 = 0.39%), red fruit (42 = 0.36%), ripened fruit (41 = 0.34%), root (44 = 0.43%) and young leaf (43 = 0.41%) (Supplementary S4).

Figure 3A, shows the relationship between samples and treatments using multidimensional scaling (MDS). It shows the robustness of the data regarding number and types of samples, i.e., mature leaf, red fruit and ripened fruit, regarding expression when treatments MJ48 leaf and MJ24 leaf are compared, in addition to young leaf sample results. In Fig. 3B, a visual representation of the mean-variance relationship using the plotMeanVar was generated as indicated from the ten samples. Differential expression analysis, in edgeR calculates the mean-square-variance relationship (dispersion) to moderate the degree of dispersion between traits (genes) was demonstrated in Fig. 3C. Finally, a heat map displaying the expression of *RoWRKY* genes of the ten samples is shown (Fig. 3D).

Discussion

Identification of the WRKY protein family in black raspberry

WRKY FTs are key regulators of many processes in plants, exhibiting specific responses to biotic stresses and involvement in tolerance to abiotic stresses (Pandey and Somssich 2009; Ramegowda et al. 2020). In model species such as *Arabidopsis thaliana* (Eulgem et al. 2000) as well as in non-model as those of the botanical family Rosaceae (Qiao et al. 2015), WRKY TFs and their involvement to promote resistance and tolerance to different stresses were identified and functionally characterized (Li et al. 2020a). The availability of assembly and annotation of the complete genome of the black raspberry in its version V3.0 (VanBuren et al. 2018b), made it possible to analyze the entire genome of the *RoWRKY* family. The identification of 59 *RoWRKY* genes in black raspberry corresponds to a slightly lower number than in other species, for example, there are 75 *AtWRKY* genes in Arabidopsis (Eulgem et al. 2000), 103 *PbWRKY* genes in pear (*Pyrus bretschneideri*) (Huang et al. 2015) and 127 *MdWRKY* genes in apple (*Malus domestica*) (Meng et al. 2016). In contrast, the rose (*Rosa chinensis*) contains 56 *RcWRKY* (Liu et al. 2019a), and 59 *FvWRKY* genes in wild strawberry (*Fragaria vesca*) (Zhou et al. 2016), and 61 *PpWRKY* genes in peach (*Prunus persica*) (Yanbing et al. 2016) were identified.

The present findings in black raspberry, an important polydrupe fruit and a plant resistant to the biotic stress caused by the fungus *Verticillium dahliae* (Bushakra et al. 2016), could become a model plant of the species of the genus *Rubus* and have an agricultural and economic potential (Fuks 1984; Graham and Brennan 2018; Moreno–medina et al. 2020). *RoWRKY* transcription factors contribute to the recent identification of *WRKY* genes in Rosaceae plant species such as apple, pear, strawberry and peach.

Subcellular localization of the RoWRKY protein Family

WRKY proteins are part of a multigene family and are involved in many responses to adverse environmental, effects (Eulgem and Somssich 2007; Rushton et al. 2012). Several members of the family are present in different cellular compartments as observed in Arabidopsis (*A. thaliana*), millet (*P. glaucum*), sorghum (*S. bicolor*) and pear (*P. betulaefolia*) species (Van Aken et al. 2013; Liu et al. 2019b; Baillo et al. 2020; Chanwala et al. 2020). In this study, analyzing the subcellular localization of RoWRKY proteins, it was found that several members present signaling for transport to many cellular compartments of the cell black raspberry. For example, all 59 RoWRKY proteins have nuclear localization sequences, similarly to peanuts (*Arachis hypogaea*), which out of 158 AhWRKY, 157 have nuclear localization sequence (Zhao et al. 2020). Specific WRKY proteins were also located in the nucleus such as PbWRKY53 from pear (*P. betulaefolia*) (Liu et al. 2019b), MbWRKY1 of the apple (*M. domestica*) (Han et al. 2018) and AtWRKY13 and AtWRKY57 of Arabipdosis (*A. thaliana*) (Van Aken et al. 2013). In addition, there are studies of functional categorization of cellular components of WRKY in the complete genomes of millet (*P. glaucum*) (Chanwala et al. 2020) and sorghum (*S. bicolor*) (Baillo et al. 2020), that could be used to complement the analysis of RoWRKY.

The subcellular localization of WRKY proteins in the nucleus was expected, as translation in the cytoplasm must return to the nucleus to bind to the promoter region of genes (W-Box) to regulate transcription. This action is aided by the signal peptides which are short proteins that act as a signal fragment or transport from the cytoplasm to the nucleus(Lu et al. 2021).

We also detected signaling to mitochondria and chloroplast. Some genes are located in other cellular compartments such as chloroplasts and mitochondria that require transcription factors encoded by the nucleus (Lee et al. 2019). On the other hand, organelles send signals to the nucleus to coordinate nuclear and organelle activities(Woodson and Chory 2008) .

Cis elements in the promoters of RoWRKY genes

Gene promoters have cis-acting elements, which act as the control center for gene transcription. The promoter regions of 59 *RoWRKY* genes exhibited several conserved cis-acting regulatory elements involved in various functions, such as abiotic and biotic stress responses (MBS, LTR, ARE, TC-rich repeat and GC motif), light-responsive elements (G -box, GT1-motif, Box4 and TCT-motif) and phytohormone regulation (ABRE, TCA element, TGA element, TGACG motif, CGTCA element, AuxRR nucleus and GARE motif). The presence of many cis-acting elements that are involved in environmental stress responses, light-responsive elements and phytohormones indicates the involvement of WRKYs in different biological processes. The expression of *WRKYs* genes can occur through the binding of a WRKY FT to the W-box or by the binding of another FT to a different cis element along the *WRKY* promoters, making an up- or down-regulation, by activating or repressing transcription (Phukan et al. 2016). Consistent with the results obtained in the promoters of the *RoWRKY*

genes, which presented one or more W-box sequences. Induced WRKYs rapidly bind to promoters of *WRKY* genes, inducing them by self-regulation and cross-regulation, building a regulatory network of WRKY FTs expression (Birkenbihl et al. 2018). In apple (*M. domestica*) the cis-acting elements W-box, GARE and ERE were found in the region of the *MdWRKY100* promoter, which may play a response to fungal and environmental stresses and in the developmental pathways (Zhang et al. 2019).

Similar to this study, the promoters of 17 *MdWRKYs* contain a cis methyl jasmonate (MeJA) response element, 12 *MdWRKY* promoters contain a salicylic acid response element, and 15 *MdWRKY* promoters contain cis elements G-Box, ABRE, CAAT-box and TATA-box, which are elements related to defense and stress (Zhang et al. 2021). *MdWRKY31* has been shown to bind to the promoters of the *MdRAV1* and *MdABIs* genes via the cis element TTGACC to mediate abscisic acid (ABA) sensitivity (Zhao et al. 2019). Therefore, cis elements play an important role in binding to WRKY-promoting regions to respond to plant growth, development, defense and stress tolerance.

Phylogeny, gene structure and motif analysis of RoWRKY protein in black raspberry

The structure of WRKY transcription factors is composed of the conservation of the WRKY domain number and the features of its zinc finger motif, which binds to the cis W-box element in the promoters of its target genes, the most essential feature of this superfamily (Eulgem et al. 2000; Zhang and Wang 2005). The classification of *RoWRKY* genes was performed according to the approach used in other species, based on the generated phylogenetic tree topology. The divisions of subfamilies or groups described in Arabidopsis were also adapted: groups I, II and III according to the number of WRKY domains and the type of zinc finger motif, together with the subdivision of group II into subgroups IIa, IIb, IIc, IId and IIe (Eulgem et al. 2000). In this study, 12, 32 and 13 *RoWRKY* genes were classified in groups I, II and III respectively, however, two *RoWRKY* genes were classified in group IV because they did not belong to any group. Group II was larger with 32 members and represented 54% of all *RoWRKY* genes.

These results are consistent with the approximations of the highest percentage size of WRKY representation in group II in Arabidopsis (63%), apple (62%), vine (65%) and rice (47%) (Meng et al. 2016). Among the subgroups, subgroup IIb and IIc have the highest number of genes, each with eight *RoWRKY* genes or 27% of the genes attributing to group II. In apple subgroup IIb (25%), Arabidopsis IIc (40%), vine IIc (41%) and rice IIc (33%) have the highest number of WRKY genes of group II (Meng et al. 2016).

Considering that most RoWRKY proteins contained the conserved WRKYGQK domain, four genes (RoWRKY4, RoWRKY59, RoWRKY57 and RoWRKY55) were identified with the WRKYGKK and WGKYGQM variant and three domains were removed for having incomplete variants of the WRKY domain. Similar variations were identified in the heptapeptide of 13 SbWRKY proteins from sorghum (*S. bicolor*) (Baillo et al. 2020). There are also similar variations identified in the MdWRKY heptapeptide from apple (*M. domestica*), in the proteins MdWRKY78, MdWRKY53, MdWRKY92 and MdWRKY96 (WRKYGKK) and other variations of MdWRKY51 and MdWRKY38 (WRKYG*K), MdWRKY114 (WRKYARS), MdWRKY20 (WRKYGKI), MdWRKY42 (WRKYGKS) (Meng et al. 2016). Furthermore, identical variations of the heptapeptide were found in peach (*P. persica*), PpWRKY33 (WRKYGKK) and PpWRKY35 (WRKYGMK) proteins (Yanbing et al. 2016). However, WRKYGKK variations in soybean (*G. max*), in proteins GmWRKY6 and GmWRKY21 do not allow binding to the cis W-box element (Zhou et al. 2008). The NtWRKY12 protein in tobacco with the WRKYGKK variant recognizes other binding sequences (TTTCCAC) instead of the normal W-box (van Verk et al. 2008). Therefore, further investigation is needed to identify the DNA-binding sequences of the variants that are part of the heptapeptide WRKYGQK.

Within the primary structure of the *RoWRKY* genes, a similar exon-intron pattern was detected. The number of introns in *RoWRKY* genes ranged from one to 11, which is almost similar to that reported in cucumber (*C. sativus*), which ranged from zero to 13 (Chen et al. 2020b). Also, these values range from zero to 17 introns in willow (*S. suchowensis*) (Bi et al. 2016) and one to 20 introns in strawberry (Zhou et al. 2016). The *RoWRKY* genes did not show absence of introns, indicating that a process of evolution of the gene family may be taking place. If there were no introns, it could be explained by the mechanisms of retroposition of genes without introns (retroduplicated genes), duplication of genes without existing introns and horizontal gene transfer (He et al. 2011). Variations in intron sizes within and between *RoWRKY* groups may have resulted from inversion, fusion or duplication events (Li et al. 2016). The similarity of exon-intron pattern provides important clues about the evolution of *RoWRKY* genes.

Of the 10 functional motifs identified in RoWRKY proteins, 1, 2, 3, 4 and 5 corresponded to the WRKY domain (Fig. 1C). Zinc finger domains are present in most RoWRKY members. Motif 6 (KTVREPRVVVQTRSE) is likely to represent the nuclear localization signal and histone methyltransferase complex (TVREPRV/EPRV) distributed mainly in groups I and II (IIe and IIc). In Arabidopsis, epigenetic regulation by histone methyltransferase is attributed to *AtWRKY53*, a senescence regulator (Banerjee and Roychoudhury 2015). In motif 7, 8 submotifs were identified, where the initiating motif of vegetative septum formation (FtsL/DivIC) (Villanelo et al. 2011) and the leucine zipper motif (bZIP_2) (Jakoby et al. 2002), are motifs that are distributed in subgroups IIa and IIe of RoWRKY proteins, having specific functions of process regulation, pathogen defense and, among others, of stress signaling. However, motif 9 (WRKY-GCM1 zinc fingers) is a plant zinc finger domain (Babu et al. 2006), that has functional diversity through expression divergence rather than protein sequence divergence. This motif is uniquely distributed in proteins RoWRKY30, RoWRKY24, RoWRKY31, RoWRKY18, RoWRKY29, RoWRKY20 and RoWRKY33. Therefore, not all identified motifs of RoWRKY proteins are functional in their predicted structure, but some motifs conforming to submotifs have specific functions.

Chromosomal localization and RoWRKY gene duplication

The chromosomal location of the 59 *RoWRKY* genes on the seven chromosomes of the black raspberry genome is almost similar to the strawberry (*Fragaria vesca*), which contains seven chromosomes and 59 *FvWRKY* genes (Zhou et al. 2016), changing their distribution intensity in both genomes. The presence of segmental and tandem duplication in the black raspberry genome differs by the number of copies and their distribution, suggesting that duplication events occurred after the divergence between black raspberry and strawberry, ca. 75 million years ago (Xiang et al. 2017). Gene duplication

events affect genome expansion, family size, and the distribution of genes in chromosomes, which are important factors for functional prediction (Zhou et al. 2016). However, whole genome, segmental, and tandem duplication were major contributors to the expansion of the *WRKY* gene family (Chen et al. 2019).

The tandem duplication of chromosomal regions leads to the expansion and evolution of the gene family, through structural and functional divergence over time (De Grassi et al. 2008; Chen et al. 2019). In this analysis, two pairs of genes were identified in tandem duplication (*RoWRKY57-RoWRKY59 = 75.2%* and *RoWRKY12-RoWRKY13* = 82.0%, with > 70% similarity), which is a lower number than the reported for strawberry (Zhou et al. 2016), Arabidopsis (Cannon et al. 2004) and similar to sorghum (Baillo et al. 2020). Despite these differences, tandem duplications may have shaped the evolution of the *RoWRKY* gene family in the black raspberry. A segmental duplication is the result of large-scale genomic events such as polyploidy or duplication of large chromosomal regions (Cannon et al. 2004). In this analysis, 20 pairs of paralogous genes were identified, of which five show similar segmental duplication (*RoWRKY40-RoWRKY39* = 100%; *RoWRKY12-RoWRKY13* = 82%; *RoWRKY24-RoWRKY22* = 75,7%; *RoWRKY46-RoWRKY44* = 100%; *RoWRKY54-RoWRKY6* = 73.2%) and the comparison of promoter regions of segmental duplicate (*RoWRKY40-RoWRKY39* = 45.9%; *RoWRKY12-RoWRKY13* = 53.9%; *RoWRKY24-RoWRKY22* = 45.0%; *RoWRKY46*-RoWRKY44 = 45.9% ; *RoWRKY54-RoWRKY6* = 46.2%) and tandem duplicate (*RoWRKY57-RoWRKY59* = 47.3% and *RoWRKY12-RoWRKY13* = 53.9%). These results showed a higher divergence in the regulatory regions than in the coding region, suggesting events of neofunctionalization and/or subfunctionalization. The strawberry has the same segmental and tandem duplication of *RoWRKY* genes, suggesting that the duplication events occurred before the formation of these species from a common ancestor.

Tandem duplications and segmental duplications have played essential roles in the evolution and diversification of the *WRKY* gene family in plant species (Zhu et al. 2014). Duplication events are significant for *WRKY* diversification, as duplicated genes can acquire new functions (Baillo et al. 2020). Gene duplication can result in sub-functionalization. For example, expansion of functions among wheat *WRKY* gene family members has occurred through tandem and whole genome duplication (Hassan et al. 2019). Therefore, the location, distribution, and tandem and segmental duplications provide evidence for the evolution and prediction of functions of black raspberry *RoWRKY* genes.

The presence of transcription factors WRKYs in this species was reported in the mapping of quantitative traits at the -QTL locus (S0141:9310–97411: ID = G14616) of the whole genome and v1.0 annotation of the black raspberry (*R. occidentalis*) (VanBuren et al. 2016). The functional evidence of TF WRKY in black raspberry remains to be explored, although the presence of chromosomal locations was identified, their functions could be computationally inferred, but could also be experimentally corroborated.

Differential expression of RoWRKY **genes**

Our heat map data showed that most of the *RoWRKY* genes were expressed in the black raspberry tissue and organ samples. These results indicate that they can participate in growth and development. For example, the genes *RoWRKY7*, *RoWRKY37* and *RoWRKY2* showed strong expression in all tissue types. However, nine *RoWRKY* genes (*RoWRKY6, RoWRKY17, RoWRKY29, RoWRKY31, RoWRKY39, RoWRKY43, RoWRKY50, RoWRKY55, RoWRKY56*) were not expressed in the 10 samples studied. In addition, some genes (*RoWRKY19* and *RoWRKY34*) were only expressed in the flower bud sample.

According to the results and information obtained in this study, the importance of WRKY in the growth, development and response to biotic and abiotic stresses in plants is evident. There are report of differential RNA-seq expression, for example, 13 *RiWRKY* genes out of 75 genes expressed in red raspberry (*Rubus ideaus* L.) provide information on the genetics of thorn development (Khadgi and Weber 2020). In addition, *ReWRKY35* and *ReWRKY31* expression annotation participates in the genetic variability of *Rubus ellipticus* (Sharma et al. 2021). Also, in the root tissue sample, the expression of *WRKY* genes in *Phytophthora rubi* rot resistance in red raspberry 'Latham (*Rubus idaeus*) was evidenced (Ward and Weber 2012).

Differential expression analyzes of strawberry (*Fragaria × ananassa*) RNA-seq data processing, show for receptacle tissues during specific stages of ripening (74 *FaWRKY* genes for White_R, Turning_R and Red_R) and achene (102 *FaWRKY* genes for White_A, Turning_A and Red_A). Also, it shows differential expression in response to infection by the fungus *Colletotrichum fructicola* (54 *FaWRKY* genes for Inf_24h, Inf_72h, Inf_96h)(Garrido-Gala et al. 2022). Therefore, gene expression studies provide functional information on two WRKY transcription factors in Rosaceae species.

Conclusion

The complete genome of the black raspberry (*Rubus occidentalis*) contains 62 identified and 59 characterized members of the *RoWRKY* gene family, and these are located on seven chromosomes. The subcellular location of RoWRKY members is in many cellular compartments. Segmental and tandem duplications may be associated with paused fusion events of the black raspberry genome. The genome distribution, organization and gene structure suggest a complex evolutionary history of this family in the black raspberry. The cis-acting elements in the promoters of *RoWRKY* genes are involved in various biotic and abiotic stress response functions and biological processes. Also, the motifs and submotifs of RoWRKY proteins have specific functions of process regulation, pathogen defense and other stress signaling functions. Duplications show higher divergence on the promoter regions, indicating events of neo/subfunctionalization in the family. Differential expression of *RoWRKY* genes in tissue samples analyzed shows preferential or specific expression in tissue samples from plant organs. In addition, *RoWRKY* genes have a stress response, when sprayed with phytohormone and also without treatment.

Declarations

Supplementary Information

Acknowledgements

Authors are thankful to CNPq, CAPES and FAPERGS for grants and fellowships.

Authors' contributions

ACO conceived the idea, WF performed the analyses and wrote the draft. ACO, CP and LCM edited the manuscript.

Availability of data and materials

Material is available as supplementary files

The authors declare no conflict of interest

Ethics approval and consent to participate

NA

Consent for publication

All the authors agree with the publication of the manuscript.

Competing interests

The authors declare no competing interests.

Author details

WF – winder.felipezz@gmail.com

CP – pegorarocamilanp@gmail.com

LCM- lucianoc.maia@gmail.com

References

1. Babu MM, Iyer L, Balaji S, Aravind L (2006) The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. Nucleic Acids Res 34:6505–6520. https://doi.org/10.1093/nar/gkl888
2. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34:369–373. https://doi.org/10.1093/nar/gkl198
3. Baillo EH, Hanif MS, Guo Y, et al (2020) Genome-wide identification of WRKY transcription factor family members in sorghum (Sorghum bicolor (L.) moench). PLoS One 15:1–24. https://doi.org/10.1371/journal.pone.0236651
4. Banerjee A, Roychoudhury A (2015) WRKY Proteins: Signaling and Regulation of Expression during Abiotic Stress Responses. Sci World J 2015:1–17. https://doi.org/10.1155/2015/807560
5. Bi C, Xu Y, Ye Q, et al (2016) Genome-wide identification and characterization of WRKY gene family in Salix suchowensis. PeerJ 2016:1–29. https://doi.org/10.7717/peerj.2437
6. Birkenbihl RP, Kracher B, Ross A, et al (2018) Principles and characteristics of the Arabidopsis WRKY regulatory network during early MAMP-triggered immunity. Plant J 96:487–502. https://doi.org/10.1111/tpj.14043
7. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
8. Bouckaert R, Vaughan TG, Barido-Sottani J, et al (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol 15:1–28. https://doi.org/10.1371/journal.pcbi.1006650
9. Bradish CM, Bushakra JM, Dossett M, et al (2016) Genotyping and phenotyping heat tolerance in black raspberry (Rubus occidentalis L.). Acta Hortic 1127:321–324. https://doi.org/10.17660/ActaHortic.2016.1127.50
10. Bushakra JM, Bassil N V., Weiland JE, et al (2016) Comparative RNA-seq for the investigation of tolerance to Verticillium wilt in black raspberry. Acta Hortic 1133:103–114. https://doi.org/10.17660/ActaHortic.2016.1133.16
11. Bushakra JM, Bryant DW, Dossett M, et al (2015) A genetic linkage map of black raspberry (Rubus occidentalis) and the mapping of Ag 4 conferring resistance to the aphid Amphorophora agathonica. Theor Appl Genet 128:1631–1646. https://doi.org/10.1007/s00122-015-2541-x
12. Bushakra JM, Dossett M, Carter KA, et al (2018) Characterization of aphid resistance loci in black raspberry (Rubus occidentalis L.). Mol Breed 38:1–19. https://doi.org/10.1007/s11032-018-0839-5
13. Cannon SB, Mitra A, Baumgarten A, et al (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. BMC Plant Biol 4:1–21. https://doi.org/10.1186/1471-2229-4-10

14. Chanwala J, Satpati S, Dixit A, et al (2020) Genome-wide identification and expression analysis of WRKY transcription factors in pearl millet (Pennisetum glaucum) under dehydration and salinity stress. BMC Genomics 21:1–16. https://doi.org/10.21203/rs.2.24672/v1
15. Chen C, Chen H, Zhang Y, et al (2020a) TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant 13:1194–1202. https://doi.org/10.1016/j.molp.2020.06.009
16. Chen C, Chen X, Han J, et al (2020b) Genome-wide analysis of the WRKY gene family in the cucumber genome and transcriptome-wide identification of WRKY transcription factors that respond to biotic and abiotic stresses. BMC Plant Biol 20:1–19. https://doi.org/10.1186/s12870-020-02625-8
17. Chen X, Li C, Wang H, Guo Z (2019) WRKY transcription factors : evolution , binding, and action. Phytopathol Res 1:1–15. https://doi.org/https://doi.org/10.1186/s42483-019-0022-x
18. De Grassi A, Lanave C, Saccone C (2008) Genome duplication and gene-family evolution: The case of three OXPHOS gene families. Gene 421:1–6. https://doi.org/10.1016/j.gene.2008.05.011
19. Dobin A, Davis CA, Schlesinger F, et al (2013) STAR: ultrafast universal RNA-seq aligne. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635
20. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2:953–971. https://doi.org/10.1038/nprot.2007.131
21. Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends Plant Sci 5:199–206. https://doi.org/10.1016/S1360-1385(00)01600-9
22. Eulgem T, Somssich IE (2007) Networks of WRKY transcription factors in defense signaling. Curr Opin Plant Biol 10:366–371. https://doi.org/10.1016/j.pbi.2007.04.020
23. Finn RD, Bateman A, Clements J, et al (2014) Pfam: The protein families database. Nucleic Acids Res 42:222–230. https://doi.org/10.1093/nar/gkt1223
24. Finn RD, Clements J, Eddy SR (2011) HMMER web server: Interactive sequence similarity searching. Nucleic Acids Res 39:29–37. https://doi.org/10.1093/nar/gkr367
25. Fuks R (1984) Rubus L (rosaceae) do estado do Rio de Janeiro. Rodriguesia 36:3–31
26. Garrido-Gala J, Higuera J-J, Rodríguez-Franco A, et al (2022) A Comprehensive Study of the WRKY Transcription Factor Family in Strawberry. Plants 11:1585. https://doi.org/10.3390/plants11121585
27. Gasteiger E, Gattiker A, Hoogland C, et al (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788. https://doi.org/10.1093/nar/gkg563
28. Graham J, Brennan R (2018) Introduction to the Rubus Genus. In: Raspberry. Springer, Cham, pp 1–16
29. Han D, Ding H, Chai L, et al (2018) Isolation and characterization of MbWRKY1, a WRKY transcription factor gene from Malus baccata (L.) borkh involved in drought tolerance. J Plant Sci 98:1–23. https://doi.org/10.1139/cjps-2017-0355
30. Hassan S, Lethin J, Blomberg R, et al (2019) In silico based screening of WRKY genes for identifying functional genes regulated by WRKY under salt stress. Comput Biol Chem 83:1–11. https://doi.org/10.1016/j.compbiolchem.2019.107131
31. He S, Zou M, Guo B (2011) The roles and evolutionary patterns of intronless genes in deuterostomes. Comp Funct Genomics 2011:. https://doi.org/10.1155/2011/680673
32. Horton P, Park KJ, Obayashi T, et al (2007) WoLF PSORT: Protein localization predictor. Nucleic Acids Res 35:585–587. https://doi.org/10.1093/nar/gkm259
33. Hrmova M, Hussain SS (2021) Plant transcription factors involved in drought and associated stresses. Int J Mol Sci 22:. https://doi.org/10.3390/ijms22115662
34. Huang X, Li K, Xu X, et al (2015) Genome-wide analysis of WRKY transcription factors in white pear (Pyrus bretschneideri) reveals evolution and patterns under drought stress. BMC Genomics 16:1–14. https://doi.org/10.1186/s12864-015-2233-6
35. Jakoby M, Weisshaar B, Dröge-Laser W, et al (2002) bZIP transcription factors in Arabidopsis. Trends Plant Sci 7:106–111. https://doi.org/10.1016/S1360-1385(01)02223-3
36. Jennings SN (2018) Advances in Rubus Breeding. In: Raspberry. Springer, Cham, pp 17–28
37. Jiang X, Chen P, Zhang X, et al (2021) Comparative analysis of the SPL gene family in five Rosaceae species: Fragaria vesca, Malus domestica, Prunus persica, Rubus occidentalis, and Pyrus pyrifolia. Open Life Sci 16:160–171. https://doi.org/10.1515/biol-2021-0020
38. Jibran R, Dzierzon H, Bassil N, et al (2018) Chromosome-scale scaffolding of the black raspberry (Rubus occidentalis L.) genome based on chromatin interaction data. Hortic Res 5:11. https://doi.org/10.1038/s41438-017-0013-y
39. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8:275–282. https://doi.org/10.1093/bioinformatics/8.3.275
40. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27
41. Khadgi A, Weber CA (2020) RNA-Seq Analysis of Prickled and Prickle-Free Epidermis Provides Insight into the Genetics of Prickle Development in Red Raspberry (Rubus ideaus L.). Agronomy 10:. https://doi.org/10.3390/agronomy10121904
42. Kula M, Krauze-Baranowska M (2016) Rubus occidentalis: The black raspberry - Its potential in the prevention of cancer. Nutr Cancer 68:18–28. https://doi.org/10.1080/01635581.2016.1115095

43. Kumar M, Michael S, Alvarado-Valverde J, et al (2022) The Eukaryotic Linear Motif resource: 2022 release. Nucleic Acids Res 50:D497–D508. https://doi.org/10.1093/nar/gkab975
44. Kun He, An-Yuan Guo, Ge Gao, Qi-Hui Zhu, Xiao-Chuan Liu, He Zhang, Xin Chen, Xiaocheng Gu and JL (2010) Computational Identification of Plant Transcription Factors and the Construction of the PlantTFDB Database. In: Ladunga I (ed) Computational Biology of Transcription Factor Binding. Springer Science+Business Media, LLC
45. Larkin MA, Blackshields G, Brown NP, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948. https://doi.org/10.1093/bioinformatics/btm404
46. Law CW, Alhamdoosh M, Su S, et al (2018) RNA-seq analysis is easy as 1-2-3 with limma , Glimma and edgeR [version 3 ; referees : 3 approved]. F1000Research 5:1–29. https://doi.org/https://doi.org/10.12688/f1000research.9005.1
47. Lee DW, Lee S, Lee J, et al (2019) Molecular Mechanism of the Specificity of Protein Import into Chloroplasts and Mitochondria in Plant Cells. Mol Plant 12:951–966. https://doi.org/10.1016/j.molp.2019.03.003
48. Leinonen R, Sugawara H, Shumway M (2011) The Sequence Read Archive. Nucleic Acids Res 39:2010–2012. https://doi.org/10.1093/nar/gkq1019
49. Lescot M, Déhais P, Thijs G, et al (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30:325–327. https://doi.org/10.1093/nar/30.1.325
50. Li M-Y, Xu Z-S, Tian C, et al (2016) Genomic identification of WRKY transcription factors in carrot (Daucus carota) and analysis of evolution and homologous groups for plants. Sci Rep 6:1–17. https://doi.org/10.1038/srep23101
51. Li W, Pang S, Lu Z, Jin B (2020a) Function and Mechanism of WRKY Transcription Factors in Abiotic Stress Responses of Plants. Plants 9:1–15. https://doi.org/10.3390/plants9111515
52. Li X, Guo W, Li J, et al (2020b) Histone acetylation at the promoter for the transcription factor PUWRKY31 affects sucrose accumulation in pear fruit. Plant Physiol 182:2035–2046. https://doi.org/10.1104/PP.20.00002
53. Liao Y, Smyth GK, Shi W (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930. https://doi.org/10.1093/bioinformatics/btt656
54. Liu X, Li D, Zhang S, et al (2019a) Genome-wide characterization of the rose (Rosa chinensis) WRKY family and role of RcWRKY41 in gray mold resistance. BMC Plant Biol 19:1–12. https://doi.org/10.1186/s12870-019-2139-6
55. Liu Y, Yang T, Lin Z, et al (2019b) A WRKY transcription factor PbrWRKY53 from Pyrus betulaefolia is involved in drought tolerance and AsA accumulation. Plant Biotechnol J 17:1770–1787. https://doi.org/10.1111/pbi.13099
56. Lu J, Wu T, Zhang B, et al (2021) Types of nuclear localization signals and mechanisms of protein import into the nucleus. Cell Commun Signal 19:1–10. https://doi.org/10.1186/s12964-021-00741-y
57. Lu S, Wang J, Chitsaz F, et al (2020) CDD/SPARCLE: The conserved domain database in 2020. Nucleic Acids Res 48:D265–D268. https://doi.org/10.1093/nar/gkz991
58. Meng D, Li Y, Bai Y, et al (2016) Genome-wide identification and characterization of WRKY transcriptional factor family in apple and analysis of their responses to waterlogging and drought stress. Plant Physiol Biochem 103:71–83. https://doi.org/10.1016/j.plaphy.2016.02.006
59. Moreno–medina BL, Casierra–posada F, Albesiano S (2020) Rubus alutaceus (Rosaceae), A new species for colombia with agronomic potential. Rev Bras Frutic 42:1–12. https://doi.org/10.1590/0100-29452020542
60. Pandey SP, Somssich IE (2009) The Role of WRKY Transcription Factors in Plant Immunity. Plant Physiol 150:1648–1655. https://doi.org/10.1104/pp.109.138990
61. Phukan UJ, Jeena GS, Shukla RK (2016) WRKY Transcription Factors: Molecular Regulation and Stress Responses in Plants. Front Plant Sci 7:1–14. https://doi.org/10.3389/fpls.2016.00760
62. Qiao X, Li M, Li L, et al (2015) Genome-wide identification and comparative analysis of the heat shock transcription factor family in Chinese white pear (Pyrus bretschneideri) and five other Rosaceae species. BMC Plant Biol 15:1–16. https://doi.org/10.1186/s12870-014-0401-5
63. Ramegowda V, Da Costa MVJ, Harihar S, et al (2020) Abiotic and biotic stress interactions in plants: A cross-tolerance perspective. In: Priming-mediated stress and cross-stress tolerance in crop plants. Academic Press, pp 267–302
64. Rinerson CI, Rabara RC, Tripathi P, et al (2016) Structure and Evolution of WRKY Transcription Factors. In: Plant transcription factors. Academic Press, pp 657–675
65. Rinerson CI, Rabara RC, Tripathi P, et al (2015) The evolution of WRKY transcription factors. BMC Plant Biol 15:1–18. https://doi.org/10.1186/s12870-015-0456-y
66. Robinson JT, Thorvaldsdóttir H, Winckler W, et al (2011) Integrative Genome Viewer. Nat Biotechnol 29:24–6. https://doi.org/10.1038/nbt.1754.Integrative
67. Rushton DL, Tripathi P, Rabara RC, et al (2012) WRKY transcription factors: key components in abscisic acid signalling. Plant Biotechnol J 10:2–11. https://doi.org/10.1111/j.1467-7652.2011.00634.x
68. Rushton PJ, Somssich IE, Ringler P, Shen QJ (2010) WRKY transcription factors. Trends Plant Sci 15:247–258. https://doi.org/10.1016/j.tplants.2010.02.006
69. Sharma S, Kaur R, Kumar K, et al (2021) Genetic variability in Rubus ellipticus collections assessed by morphological traits and EST-SSR markers. J Plant Biochem Biotechnol 30:37–55. https://doi.org/10.1007/s13562-020-00567-8

70. Sneath PHA, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. San Francicso: Freeman, 1973. This Week ´s Citatation Class 18
71. Van Aken O, Zhang B, Law S, et al (2013) AtWRKY40 and AtWRKY63 Modulate the Expression of Stress-Responsive Nuclear Genes Encoding Mitochondrial and Chloroplast Proteins. Plant Physiol 162:254–271. https://doi.org/10.1104/pp.113.215996
72. van Verk MC, Pappaioannou D, Neeleman L, et al (2008) A Novel WRKY Transcription Factor Is Required for Induction of PR-1a Gene Expression by Salicylic Acid and Bacterial Elicitors. Plant Physiol 146:1983–1995. https://doi.org/10.1104/pp.107.112789
73. VanBuren R, Bryant D, Bushakra JM, et al (2018a) Sequence and Analysis of the Black Raspberry (Rubus occidentalis) Genome. In: The Genomes of Rosaceous Berries and Their Wild Relatives, Compendium of Plant Genomes. Springer, Cham, pp 185–197
74. VanBuren R, Bryant D, Bushakra JM, et al (2016) The genome of black raspberry (Rubus occidentalis). Plant J 87:535–547. https://doi.org/10.1111/tpj.13215
75. VanBuren R, Wai CM, Colle M, et al (2018b) A near complete, chromosome-scale assembly of the black raspberry (Rubus occidentalis) genome. Gigascience 7:1–9. https://doi.org/10.1093/gigascience/giy094
76. Villanelo F, Ordenes A, Brunet J, et al (2011) A model for the Escherichia coli FtsB/FtsL/FtsQ cell division complex. BMC Struct Biol 11:. https://doi.org/10.1186/1472-6807-11-28
77. Ward JA, Weber CA (2012) Comparative RNA-seq for the investigation of resistance to phytophthora root rot in the red raspberry “Latham.” Acta Hortic 946:67–72. https://doi.org/10.17660/ActaHortic.2012.946.7
78. Wingett SW, Andrews S (2018) FastQ Screen: A tool for multi-genome mapping and quality control[version 2; referees: 4 approved]. F1000Research 7:1–13. https://doi.org/10.12688/f1000research.15931.1
79. Woodson JD, Chory J (2008) Coordination of gene expression between organellar and nuclear genomes. Nat Rev Genet 9:383–395. https://doi.org/10.1038/nrg2348
80. Xiang Y, Huang CH, Hu Y, et al (2017) Evolution of rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. Mol Biol Evol 34:262–281. https://doi.org/10.1093/molbev/msw242
81. Yanbing G, Zhirui J, Fumei C, et al (2016) Genome-wide identification and expression analysis of the WRKY gene family in peach. Yi chuan = Hered 38:254–270. https://doi.org/10.16288/j.yczz.15-235
82. Yoon Y, Seo DH, Shin H, et al (2020) The role of stress-responsive transcription factors in modulating abiotic stress tolerance in plants. Agronomy 10:1–23. https://doi.org/10.3390/agronomy10060788
83. Zhang F, Wang F, Yang S, et al (2019) MdWRKY100 encodes a group I WRKY transcription factor in Malus domestica that positively regulates resistance to Colletotrichum gloeosporioides infection. Plant Sci 286:68–77. https://doi.org/10.1016/j.plantsci.2019.06.001
84. Zhang H, Yang L, Zhou R, et al (2021) Genome Wide Transcriptomic Analysis of WRKY Gene Family Response to Biotic Stresses in Malus × domestica. Am J Plant Sci 12:858–885. https://doi.org/10.4236/ajps.2021.125058
85. Zhang Y, Wang L (2005) The WRKY transcription factor superfamily: Its origin in eukaryotes and expansion in plants. BMC Evol Biol 5:1–12. https://doi.org/10.1186/1471-2148-5-1
86. Zhao N, He M, Li L, et al (2020) Identification and expression analysis of WRKY gene family under drought stress in peanut (Arachis hypogaea L.). PLoS One 15:1–25. https://doi.org/10.1371/journal.pone.0231396
87. Zhao XY, Qi CH, Jiang H, et al (2019) The MdWRKY31 transcription factor binds to the MdRAV1 promoter to mediate ABA sensitivity. Hortic Res 6:1–16. https://doi.org/10.1038/s41438-019-0147-1
88. Zhou H, Li Y, Zhang Q, et al (2016) Genome-Wide Analysis of the Expression of WRKY Family Genes in Different Developmental Stages of Wild Strawberry (Fragaria vesca) Fruit. PLoS One 11:1–18. https://doi.org/10.1371/journal.pone.0154312
89. Zhou QY, Tian AG, Zou HF, et al (2008) Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. Plant Biotechnol J 6:486–503. https://doi.org/10.1111/j.1467-7652.2008.00336.x
90. Zhu Y, Wu N, Song W, et al (2014) Soybean (Glycine max) expansin gene superfamily origins: Segmental and tandem duplication events followed by divergent selection among subfamilies. BMC Plant Biol 14:1–19. https://doi.org/10.1186/1471-2229-14-93

Figures

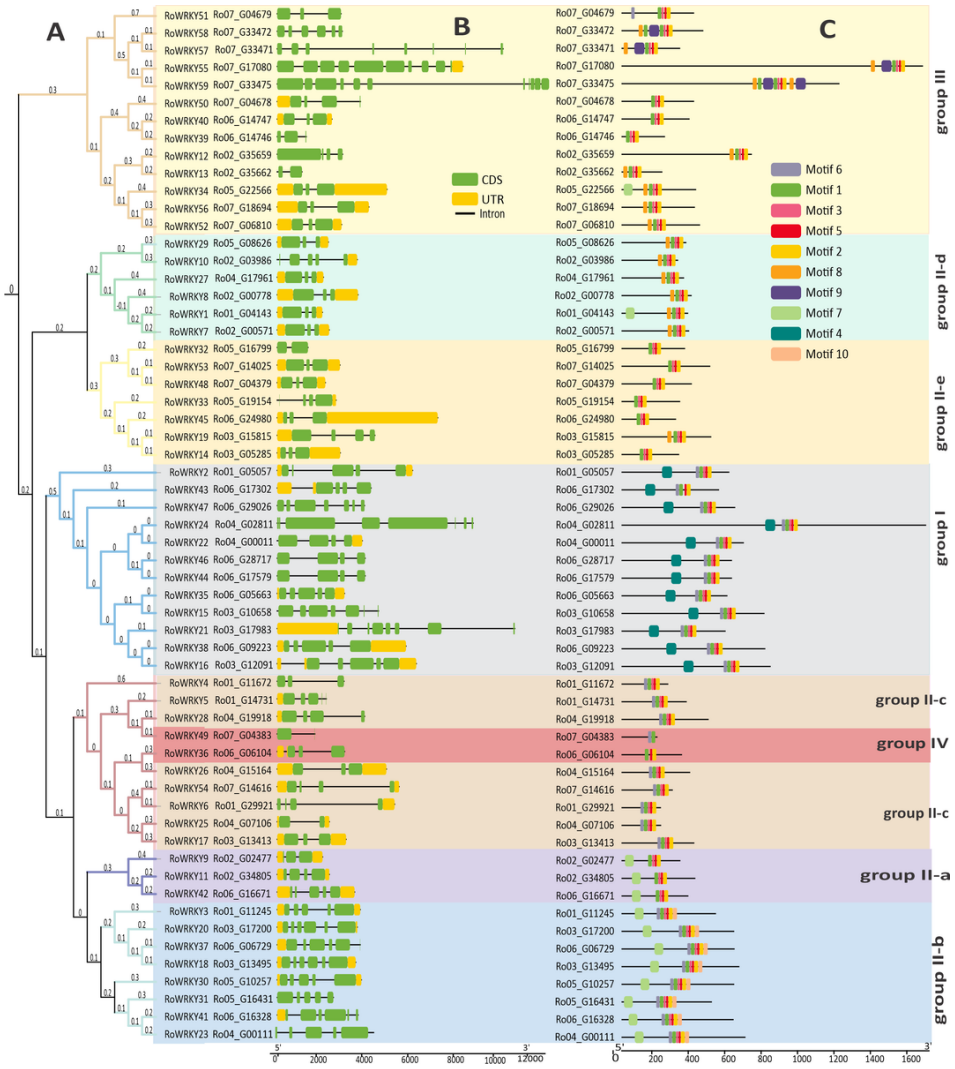

Figure 1

Phylogenetic relationships and structure of genes encoding the black raspberry RoWRKY proteins: (**A**) The unrooted tree was generated with the BEAST program using the full-length amino acid sequences of the 59 black raspberry RoWRKY proteins by the UPGMA method, with 10,000,000 bootstrap replications. RoWRKY protein subfamilies (I, IIa, IIb, IIc, IId, IIe, III and IV); (**B**) Exon/intron organization of raspberry *RoWRKY* genes. Yellow boxes represent exons and black lines represent introns. Untranslated regions (UTRs) are indicated by green boxes. The sizes of exons and introns can be estimated using the scale at the bottom; (**C**) Schematic representation of conserved motifs in black raspberry RoWRKY proteins, elucidated from publicly available data (NCBI CDD Domain – Pfam –18271 PSSMs). Each colored rectangular box represents a motif with the given name and motif consensus.

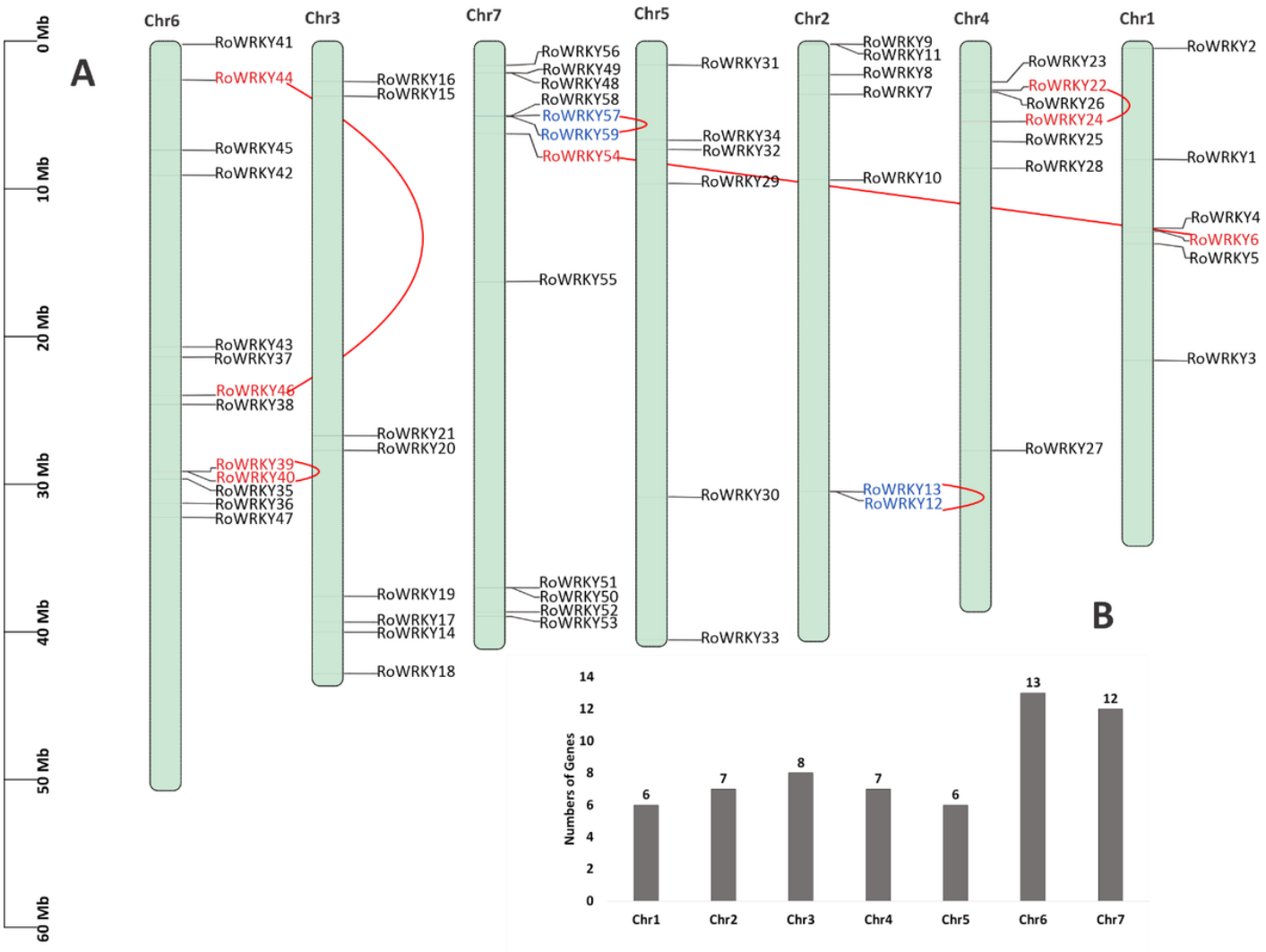

Figure 2

Chromosomal map and coordinates of *RoWRKY* gene duplication events: (**A**) The identity of each linkage group is indicated in the central part of each bar. The possible Segmental duplicated genes are connected by red color lines and the duplicated gene pair in tandem in blue color on chromosome. (**B**) The number of *RoWRKY* genes in each chromosome.

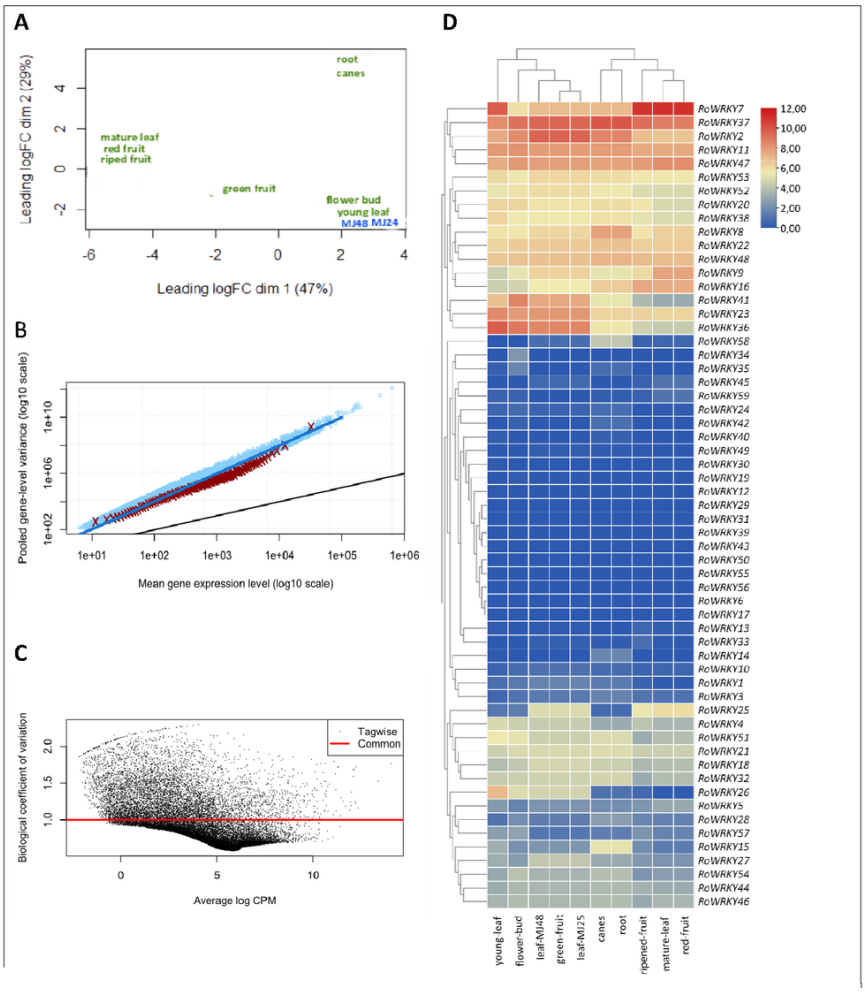

Figure 3

Analysis of gene expression in 10 samples of black raspberry: (**A**) Multidimensional scaling (MDS), indicating the dispersion of treatments and replications, as a function of the general pattern of gene expression of each sample, (**B**) mean-variance relationship and dispersion were plotted using edgeR's plotMeanVar function to explore the mean-variance relationship. Each point represents the estimated mean and variance for each gene, with pooled variances, as well as the common scatter of overlapping trend (untreated and treated samples), (**C**) Scatter plots, plotBCV illustrates the relationship of the biological coefficient of variation with the average log counts per million, (**D**) RoWRKY heat map (the values are presented in log2 from 0.00 to 12.00), the red color indicates the greater the number of reads or genes expressed according to the sequencing depth and blue indicates smaller and read or genes expressed according to sequencing depth.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- suplementaryS1S4.xlsx