PLOS ONE

Genetic associations with resistance to *Meloidogyne enterolobii* in guava (*Psidium* sp.) using cross-genera SNPs and comparative genomics to *Eucalyptus* highlight evolutionary conservation across the Myrtaceae

Carlos Antonio Fernandes Santos[1☯], Soniane Rodrigues da Costa[2☯], Leonardo Silva Boiteux[3], Dario Grattapaglia[4]*, Orzenil Bonfim Silva-Junior[4☯]

1 Embrapa Semi-Arid, Petrolina, Pernambuco, Brazil, 2 Graduate program in Genetic Resources, Universidade Estadual de Feira de Santana, Feira de Santana, Bahia, Brazil, 3 Embrapa Vegetable Crops (CNPH), Brasilia, Distrito Federal, Brazil, 4 Embrapa Genetic Resources and Biotechnology (CENARGEN), Brasília, Distrito Federal, Brazil

☯ These authors contributed equally to this work.
* dario.grattapaglia@embrapa.br

Abstract

Tropical fruit tree species constitute a yet untapped supply of outstanding diversity of taste and nutritional value, barely developed from the genetics standpoint, with scarce or no genomic resources to tackle the challenges arising in modern breeding practice. We generated a *de novo* genome assembly of the *Psidium guajava*, the super fruit "apple of the tropics", and successfully transferred 14,268 SNP probesets from *Eucalyptus* to *Psidium* at the nucleotide level, to detect genomic loci linked to resistance to the root knot nematode (RKN) *Meloidogyne enterolobii* derived from the wild relative *P. guineense*. Significantly associated loci with resistance across alternative analytical frameworks, were detected at two SNPs on chromosome 3 in a pseudo-assembly of *Psidium guajava* genome built using a syntenic path approach with the *Eucalyptus grandis* genome to determine the order and orientation of the contigs. The *P. guineense*-derived resistance response to RKN and disease onset is conceivably triggered by mineral nutrients and phytohormone homeostasis or signaling with the involvement of the miRNA pathway. Hotspots of mapped resistance quantitative trait loci and functional annotation in the same genomic region of *Eucalyptus* provide further indirect support to our results, highlighting the evolutionary conservation of genomes across genera of Myrtaceae in the adaptation to pathogens. Marker assisted introgression of the resistance loci mapped should accelerate the development of improved guava cultivars and hybrid rootstocks.

Introduction

Among the economically important species of family Myrtaceae, and within the large group encompassing ~3,500 fleshy fruit species in the family [1] the guava (*Psidium guajava* L.) stands out as the most commercially relevant crop [2]. Native to tropical America and widely distributed in subtropical and tropical countries, guava is frequently referred to as a super fruit due to its high polyphenolic and multivitamin content with important ethnopharmacological properties [3, 4]. Notwithstanding its recognized nutraceutical value, the estimated worldwide annual production in 2018 was only 6.75 million tons [5], just a fraction of the 129.6 million tons of apple, for example [6]. Guava still experiences the status of a very large number of mostly unknown tropical fruits of outstanding taste and nutritional value that are either undomesticated or mostly unexploited from the genetics, genomics and breeding standpoint.

Despite the availability of extensive genetic variation in the genus *Psidium*, the genetic basis of current guava cultivars is narrow, resulting in significant susceptibility to some diseases [7]. Following the first report of its occurrence in guava in Brazil [8], the root-know nematode (RKN) *Meloidogyne enterolobii* is currently the most economically important pathogen of guava in the Neotropics [9, 10]. *Meloidogyne enterolobii* (= *M. mayaguensis*) is a highly polyphagous species with a host range similar to that of the two major RKN species, *M. incognita* and *M. javanica* [11]. However, *M. enterolobii* is a greater agricultural threat with wider virulence profile, having the ability of 'breaking down' a wide range of resistance factors effective against other major *Meloidogyne* species in many crops [11, 12].

While controlling polyphagous RKN species by crop rotation in woody perennials is not an alternative, chemical control is increasingly banned due to environmental and human health concerns. The deployment of natural plant resistance is thus the most sustainable alternative toward intensive crop production. Major advances have been made in species of *Prunus* with a suite of resistance genes to RKN species finely mapped or cloned and used in breeding for durable resistance [13, 14]. In contrast, despite the current use of resistant rootstocks, very little is known about genes and defense mechanisms underlying the RKN resistance in other mainstream woody crops such as coffee and grapevine [15]. Sources of resistance to *M. enterolobii* have been detected in wild *Psidium* relatives but not in *P. guajava* accessions [9]. The deployment of interspecific (*P. guajava* × *P. guineense*) rootstock hybrids with resistance derived from *P. guineense* Swartz is currently the only viable management strategy (Fig 1). Nevertheless, cultivar development could be more efficient by the still unexploited prospects of introgressing disease resistance genes from *P. guineense* into *P. guajava*.

Progress has been made in the Myrtaceae family by generating genomic resources for forest trees and to a much lesser extent for fruit species [16]. Major emphasis has been in *Eucalyptus* species, for which a reference genome for *E. grandis* [17] and a multi-species SNP platform EuCHIP60K [18] are available. Alike *Eucalyptus*, *Psidium* displays a chromosome number $x = 11$, a basic complement largely conserved across the family [19, 20]. Recently, a chromosome-level assembly of the *P. guajava* genome corroborated its high collinearity to the *E. grandis* genome [21]. To date several microsatellites have been published for *Psidium* [22, 23], while SNP marker data have only been generated for *Psidium guajava* using genotyping by sequencing methods based on complexity reduction with restriction enzyme digestion [24, 25]. This approach is well known to suffer from poor data reproducibility and portability across experiments, especially in heterozygous genomes [26, 27]. Conversely, cross-genera transferability of *Eucalyptus* SNPs genotyped with the gold-standard Illumina Infinium™ EuCHIP60K platform to *Psidium* had been demonstrated early on [18], opening solid possibilities of genome-wide diversity and association studies in *Psidium* using high quality and portable SNP data across genomes.

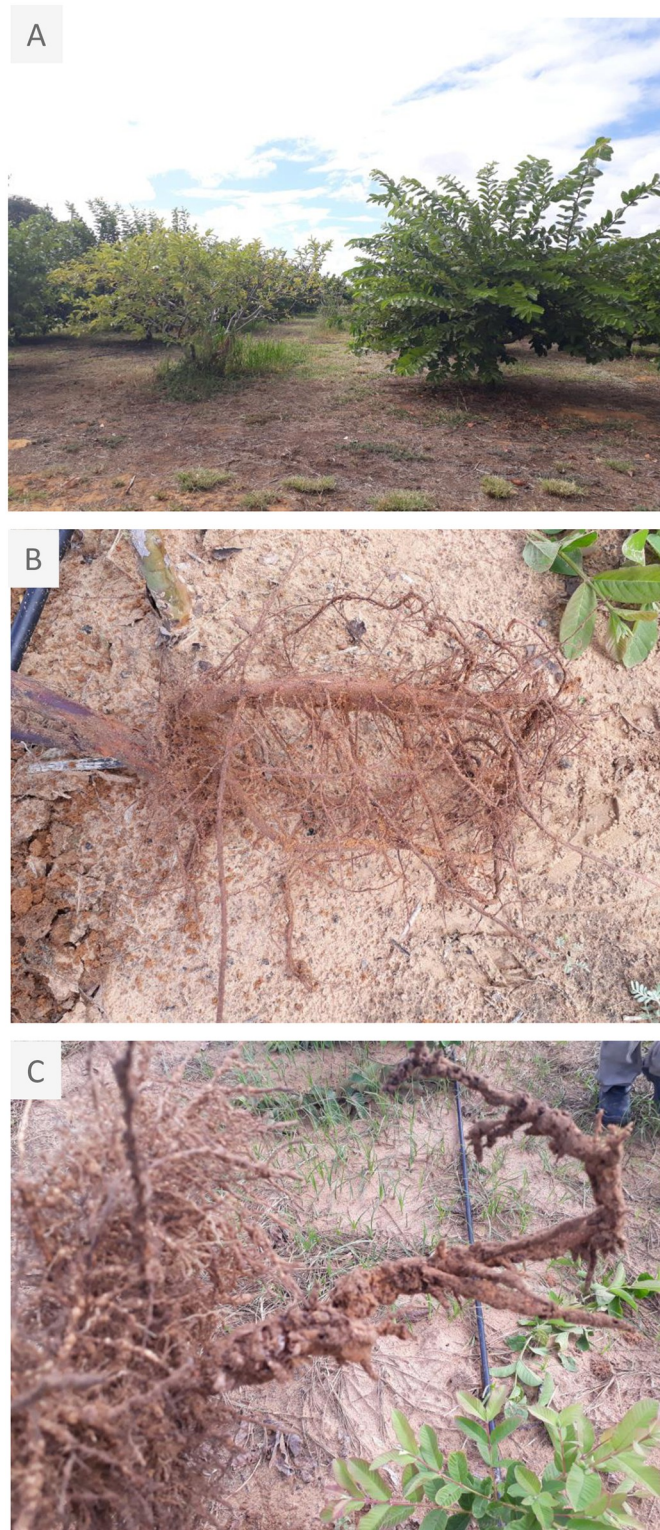**Fig 1.** **(A)** *Psidium guajava* plant attacked by *M. enterolobii* root knot nematode (left) and a *P. guajava* grafted plant onto a resistant *P. guajava x P. guineense* hybrid (right); **(B)** healthy roots of resistant *P. guajava x P. guineense* hybrid and **(C)** *M. enterolobii* RKN infected roots of a susceptible *P. guajava* plant.

https://doi.org/10.1371/journal.pone.0273959.g001

Association mapping has allowed important advances in mapping resistance loci to a wide array of *Meloidogyne* species across distinct hosts providing tools to assist breeding [28]. In this work, we generated a *de novo* assembly of the *P. guajava* genome and successfully transferred a large number of SNP probesets from *Eucalyptus* to *Psidium* at the nucleotide level to carry out an association study for *M. enterolobii* resistance response. Using these resources, we describe the detection of genomic regions harboring resistance loci to RKN derived from the wild relative *P. guineense*. Additionally, our results highlight the evolutionary conservation of genomes across genera of Myrtaceae in the adaptation to pathogens.

Material and methods

Psidium guajava × *P. guineense* association mapping population

To provide the best opportunity for recombination, an association mapping population was created by allowing open pollination among F_1 interspecific root–stock hybrids ('BRS Guaraçá' = *P. guajava* × *P. guineense*). A total of 189 outbred F_2 plants from open pollinated fruits harvested from 22 BRS Guaraçá F_1 hybrid trees were evaluated for *M. enterolobii* reaction in a controlled inoculation experiment at Embrapa Semi-Arid, Petrolina-PE, Brazil. Sample collection from native *Psidium* populations was granted by the Brazilian Institute of the Environment (IBAMA), authorization number CGEN Nº 001-B/2013.

Meloidogyne enterolobii inoculation and plant reaction evaluation

Seedlings of individual F_2 plants (with $\approx$ 25 cm in height) were inoculated with 10,000 eggs + second-stage juveniles (J2). The *M. enterolobii* inoculum was extracted from guava roots collected in a commercial area in the city of Petrolina-PE, using classic methods [29]. Each plant was inoculated with 2 mL of the suspension in each one of two hollows in the soil at a distance of 1.5 cm from the stem and 2.5 cm deep. At 120 days after inoculation, the plants were collected and the roots were carefully washed in water within a plastic container to avoid loss of nematode eggs. A qualitative binary phenotype of presence/absence of root galls and a quantitative Reproduction Factor (RF) trait were measured. Individual roots (5 grams) were processed, using a blender for 30 sec. Eggs and J2 were counted, determining the total number of nematodes (final population) and the reproduction factor (RF = final nematode population/ the initial inoculum), was calculated [30]. Plants were classified as resistant when a reduction $\geq$ 90% in RF was observed in relation to the susceptible plants [29]. For the subsequent quantitative analysis, the RF values were transformed into $\sqrt{RF} + 0.5$ due to the frequent occurrence of individuals with zero or very low RF values.

Genetic material and SNP genotyping

DNA extraction was performed from young leaves using an optimized protocol for high quality DNA from woody plants [31]. Plants were genotyped with the Infinium EUChip60K containing 60,904 *Eucalyptus* SNPs [18]. SNPs on the EUChip60K are coded with the chromosome number followed by the physical address in base pairs on version 1.1 of the *E. grandis* genome [17] which were then updated to the version 2.0 genome deposited at https://jgi.doe.gov/data-and-tools/. SNP genotyping was performed at GeneSeek (Lincoln, NE, USA).

Quality control of informative SNPs

Due to the complexity of the cross-genera genotyping assay, a clustering procedure was applied to the *Psidium* derived EUChip60K intensity data. We used the standard GenCall algorithm in GenomeStudio 2.0 (Illumina, Inc. San Diego, USA) following its best practices and

criteria described earlier [18]. After these QC steps, and in order to extract the maximum amount of quality genotypic data, the optiCall algorithm was used to ascertain genotypes at common and low-frequency variants [32].

Allelic association tests and analyses of mixed linear models

Allelic and genotypic association tests were employed, respectively, to test for the association between single SNP alleles and genotypes and the binary presence/absence of root galls and the RF phenotype. Allelic and genotypic association tests were carried out with a Fisher's exact test and Cochran-Armitage trend test, respectively. Bonferroni correction and the Benjamini & Hochberg (B&H) and Benjamini & Yekutieli (B&Y) 'false discovery rate' (FDR) procedures were implemented with PLINK [33]. A mixed linear model (MLM) analysis was carried out with the transformed RF phenotype using TASSEL 5.2.65 [34]. An 'identical by state' (IBS) matrix estimated in PLINK 1.9 [33] was used to account for the confounding effects of both population and family structure.

Comparative genomic analyses of *Eucalyptus* associated SNPs in the *Psidium* genome

At the time of this study, a genome sequence assembly was not available for *P. guajava* or *P. guineense*. To allow bona fide comparative genomic analyses of the associated genomic regions using *Eucalyptus* SNPs onto the *Psidium* genome, we sequenced the genome of the *Psidium guajava* S_2 plant UENFGO8.1–10 derived from two generations of selfing using PACBIO Sequel I technology. The FALCON-Unzip v.1.1.5 pipeline [35] was applied to produce an initial assembly. Contigs were aligned to the *Eucalyptus grandis* v2.0 genome sequence in Phytozome v13 [36] using the Cactus whole-genome multiple alignment program [37]. To obtain the full value of the Cactus protocol and to mitigate the effect of the fragmentary assembly on alignment quality, we aligned the contigs to a set of three chromosome-level related genomes. We included the *E. grandis* (Myrtaceae) and the most likely sister to Myrtaceae, the *Punica granatum* (Lythraceae), as well as *Vitis vinifera* (Vitaceae) as the ultimate outgroup. The tree topology was generated starting from proteome data to reconstruct phylogenies that chart the relationships among these organisms. Protein sequences into the genomes were downloaded from the NCBI repository. For *P. guajava* proteins, the Augustus pipeline [38] was run on the contig assembly using the extrinsic evidence provided by aligning RNA-Seq data available for the species in NCBI SRA. Single-copy genes in the BUSCOv3 program [39, 40] were used to identify shared subsets from the different sets of protein data across the genomes. Proteins were aligned using MAFFT and filtered with trimAl, and the maximum likelihood tree was built using RAxML [39, 41] to estimate the species phylogeny. The final non-overlapping whole-genome multiple alignments with ProgressiveCactus described as a HAL file [42] were extracted for the eleven chromosomes of the *E. grandis* and linked further using the algorithm implemented in Ragout2 [43].

The reference genome assembly using synteny in Ragout2 was used to ensure consistency of the transference of the SNP probesets from the target genome of the *P. guajava*. For this purpose, starting from a BED file with probesets coordinate annotation on the *E. grandis* genome, we used the halLiftOver procedure in the HAL package to perform a base-by-base mapping between guava and *Eucalyptus*. The output which refers to the probeset sequences in the *E. grandis* genome to their corresponding locations in guava was written in the PSL format and then converted to the AXT format using the program utilities in the Kent's Utilities package from the UCSC Genome Browser tools [44]. The corresponding locations and sequence content of probesets in guava were inspected to accept only those that matched the same

original positions in the query genome and recovered the same assayed substitution at the nucleotide-level. Otherwise, the probeset location was called erroneous and discarded. Finally, we performed a transcript consistency analysis using the comprehensive transcript set provided for the *E. grandis* genome assembly. We used the CAT pipeline [45] for each of one of 46,280 transcripts in the *E. grandis* annotation to determine their contig location and orientation in the contigs into the *P. guajava* genome with respect to the alignments between the two genomes from the HAL file and extrinsic evidences based on public mRNA-Seq data. After the comparative annotation of protein-coding genes, an evaluation of the putative effects of genetic variants was carried out using a database prepared using the SnpEff v5.0c pipeline [46].

Results

EUChip60K performance in *Psidium*

A total of 6,879 SNPs were successfully assayed from the 60,904 EuCHIP60K SNPs using the GenCall clustering algorithm with a call rate CR$\geq$ 0.90 and no minimum allele frequency (MAF) threshold. The normalized X-Y intensity data for these 6,879 SNPs were exported from Genome Studio (GS) and genotypes ascertained using the optiCall algorithm (default options) resulting in 6,225 SNPs. When a MAF $\geq$0.01 was applied to the 6,879 SNPs only 521 SNPs were retained with the standard GenCall algorithm in GS, while optiCall delivered 4,143 polymorphic SNPs, an eight-fold improvement (S1 File).

Phenotypic evaluation

Out of the 189 originally inoculated plants, 175 ultimately survived to final evaluation. For the binary phenotype, 92 resistant plants were free of *M. enterolobii* induced galls while 83 plants displayed conspicuous root galls and were classified as susceptible. The untransformed Reproduction Factor (RF) phenotype varied from zero to 2.53 and when considered in a simple binary fashion only six of the 175 plants were scored with a reproduction factor RF> 1.0, thus classified as susceptible, while all others, with RF <1.0, were rated as resistant (S1 File). Although these binary counts fitted the same epistatic models proposed earlier in an inbred F_2 population [47] only one out of the 4,143 SNP data segregated in a 1:2:1 ratio indicating that the population did not behave as a regular inbred F_2 from a single F_1 plant but rather, as expected, as a mixture of outbred crossed offspring.

Allelic and genotypic association analyses

Following multiple test corrections three and four SNPs displayed significant allelic and genotypic association respectively for the binary trait of *M. enterolobii* galls according to the Cochran-Armitage trend test (with adjusted p-value < 5.0E-3) (Table 1). The combination of allelic and genotypic association tests indicated two SNPs (*viz.* EuBR03s29615246 and EuBR03s30383415) in significant association with the overall resistance reaction. The SNP EuBR03s30383415 found significant by both the allelic and genotypic association, was later found significantly associated by the MLM analysis as well. The additional SNPs also on chromosome 3, EuBR03s16993500, EuBR03s37875650, EuBR03s21599380 showed variable signal depending on the statistical association test employed and were not considered further.

Association mapping with a mixed linear model

The MLM analysis incorporating population and kinship covariates resulted in one SNP (EuBR03s30383415) associated with *M. enterolobii* resistance using TASSEL following the specified threshold p-value of <5.0E^{-04}, with coefficient of determination (R^2) of 0.107

Table 1. SNP markers displaying allelic and genotypic associations and corresponding adjusted p-values for the binary trait presence/absence of *Meloidogyne enterolobii* galls in 175 outbred F$_2$ plants derived from open pollination among F$_1$ hybrid *Psidium guajava* × *P. guineense* plants. SNPs are coded with the chromosome number followed by the physical address in base pairs on version 2.0 of the *E. grandis* genome.

SNP marker	Chromosome	Adjusted p-value		
		Bonferroni	**FDR/B&H**[*]	**FDR/B&Y**[**]
Allelic association—Fisher's exact test				
EuBR03s29615246	03	$6.26E^{-16}$	$6.26E^{-16}$	$5.21E^{-15}$
EuBR03s16993500	03	$1.48E^{-05}$	$7.40E^{-06}$	$6.16E^{-05}$
EuBR03s30383415	03	0.002022	0.000674	0.005616
EuBR03s37875650	03	0.05376	0.01344	0.112
EuBR03s11126682	03	0.5225	0.1045	0.8705
Genotypic association—Cochran-Armitage trend test				
EuBR03s29615246	03	$2.84E^{-07}$	$2.84E^{-07}$	$2.37E^{-06}$
EuBR03s37875650	03	0.000289	0.000145	0.001203
EuBR03s30383415	03	0.001052	0.000351	0.002921
EuBR03s21599380	03	0.001934	0.000484	0.004028
EuBR03s16993500	03	0.03814	0.007628	0.06355
EuBR03s7520846	03	0.05023	0.008372	0.06974

[*]False discovery rate (FDR) Benjamini & Hochberg (B&H).

[**]False discovery rate (FDR) Benjamini & Yekutieli (B&Y).

(Fig 2; Table 2), suggesting that factors residing in this genomic region may explain a considerable fraction of the *M. enterolobii* resistance phenotype. Three additional SNPs were suggestive of association, one also on chromosome 3 and the other two on different chromosomes. Two of the four SNPs detected in the MLM analysis, namely EuBR03s30383415 and EuBR03s16993500, both on chromosome 3, had also been detected by both the allelic and the genotypic association analyses.

Genome assembly of *P. guajava* and comparative genomic analyses to the *E. grandis* genome

The assembled contigs for our whole genome shotgun of *P. guajava* totaled 359.4 Mb into 510 sequences and covered 78% of the 463 Mb genome size estimated for *P. guajava* by flow cytometry [19]. This assembly was deposited at DDBJ/ENA/GenBank under the accession JAGHRR000000000. The version described in this article is JAGHRR010000000. These resources were registered under the BioProject ID PRJNA713343. These contigs were further linked to provide a reference genome assembly from the whole-genome alignments against the *E. grandis* assembly using a syntenic path approach. Some of the original contigs were broken to fit into the hierarchical block construction, suggesting the presence of rearrangements of the homologous sequences between the two genomes across large segments (>10 kbp). At smaller distances in the range of the probeset sequence length (100 bp) the resulting syntenic path covered 32.1% of our *de novo* guava's assembly, which allowed us to successfully reallocate 14,268 probesets of the EucHIP60k. These probesets are distributed across 377 contigs (328 Mb) into the assembly. Comparative transcript identification resulted in the annotation of 21,240 protein-coding loci. Importantly, out of the 4,143 SNP probesets that detected polymorphism, 1,241 were fully converted between the genomes at the nucleotide-level resolution and matched the original allelic variation of the SNP into the genotyping assay. A detailed analysis was therefore possible for these variant positions in the *P. guajava* genome. Prediction of effects at these genomic variants revealed diverse types of impact on 6,201 transcripts in the
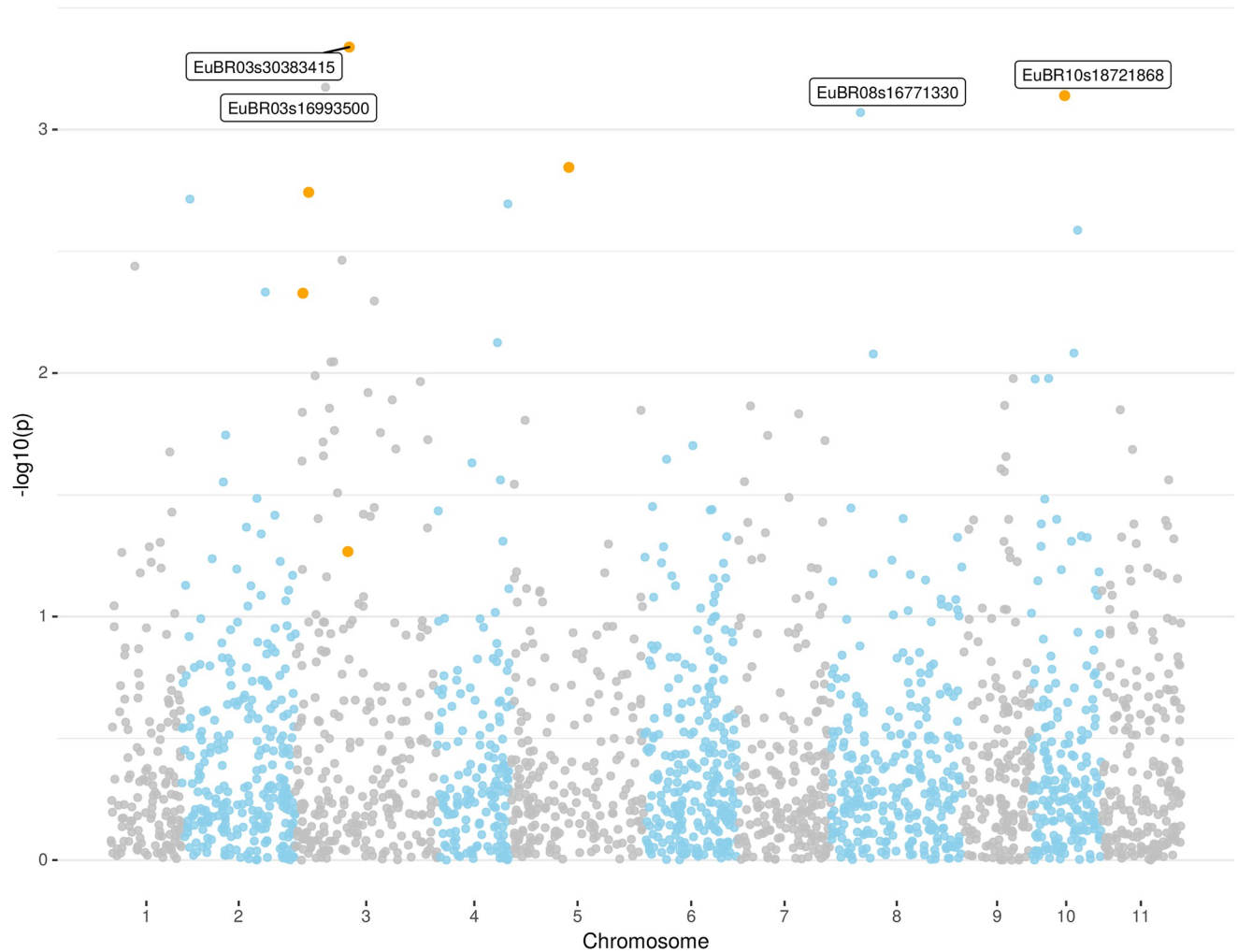
Fig 2. Manhattan plot of the association analysis for *M. enterolobii* root knot nematode (RKN) resistance in guava from a MLM analysis. Significantly associated SNPs (see Table 2) are labeled and those for which SNP probesets were reallocated on the *Psidium guajava* genome are highlighted by orange dots.

https://doi.org/10.1371/journal.pone.0273959.g002

P. guajava genome from which 2,379 have putative homologues in the *E. grandis* genome assembly. A SNP annotation file (VCF) (S2 File) and an Illumina GenomeStudio™ 2.0 cluster file (EGT) for the successfully reallocated probesets are provided (S3 File). The EGT file will be particularly valuable for processing and quality control of SNP data for future *Psidium* genotyping experiments using the EucHIP60K.

Table 2. SNPs displaying significant associations and their coefficients of determination (R^2) according to a mixed linear model (MLM) analyses in TASSEL for reproduction factor of *Meloidogyne enterolobii* root know nematode in 175 outbred F_2 plants derived from open pollination among F_1 hybrid *Psidium guajava* × *P. guineense* plants. SNPs are coded with the chromosome number followed by the physical address in base pairs on version 2.0 of the *E. grandis* genome.

SNP	*Eucalyptus/Psidium* chromosome	p-value	R^2
EuBR03s30383415	03	4.59E^{-04}	0.107
EuBR03s16993500	03	6.70E^{-04}	0.097
EuBR10s18721868	10	7.26E^{-04}	0.094
EuBR08s16771330	08	8.51E^{-04}	0.115

https://doi.org/10.1371/journal.pone.0273959.t002

The probe sequences for the *M. enterolobii* resistance associated SNPs EuBR03s29615246 and EuBR03s30383415 spaced by 768 kbp on the *Eucalyptus* genome were found spaced nearly 470 kbp apart onto a single sequence of the *P. guajava* assembly that spans 3.14 Mbp in size (JAGHRR010000008.1). The physical interval covered by this contig in the pseudo-assembly of the *Psidium guajava* using the syntenic path alignment approach to the *Eucalyptus grandis* assembly corresponds to a genomic region of about 7.45 Mb along Chromosome 3 (coordinates 27,580,466 to 35,980,336 of the Egrandisv2.0).

Our pipeline for protein-coding gene annotation predicted 102 gene models with coordinates within the contig of the *Psidium guajava* assembly containing the probes for SNPs EuBR03s29615246 and EuBR03s30383415. The gene density of 32.5 genes/Mbp is similar to its corresponding putative orthologous region in the *Eucalyptus grandis* assembly, which has a density of 36.1 genes/Mbp (269 gene models distributed along 7.5 Mbp). Interestingly, the comparison of the repertoire of putative NBS-LRR genes between the genomes along this orthologous region shows a different picture. While the *Eucalyptus* genome contains 31 models of physical arrangement in this gene family within this locus, we could identify only 2 models within the corresponding region in the *Psidium guajava* assembly.

In terms of the predicted impact based exclusively on a bioinformatics analysis, these two SNPs are variants with moderate effect causing non-synonymous changes in respect to the conceptual translation of the transcripts J3R85_001472 /Eucgr.C01744 (Uniprot id: A0A059CPK6) and J3R85_001443/Eucgr.C01791 (Uniprot id: A0A059CPT9), respectively. SNP EuBR03s30383415 is also a modifier variant occurring downstream to the gene locus for J3R85_001444/ Eucgr.C01790 (Uniprot id: A0A059CQ27). The physical location, corresponding probesets and features of the predicted variant genes for these two associated SNPs on chromosome 3 are summarized (Fig 3; Table 3) and further discussed below.

Discussion

The EuCHIP60K provides genome-wide SNP genotyping in *Psidium*

Our genetic association experiment was driven by successful genotyping and precise reallocation of SNP probesets from the *Eucalyptus* genome to a *de novo* assembly of the *Psidium* genome. Typically, for poorly funded orphan crops, SNP genotyping has been carried out by one of the several methods of restriction enzyme-based reduced representation sequencing [48]. This approach has recently been used to investigate the genetic diversity of *Psidium guajava* [24], as well as related species of the genus [25]. These methods offer the advantage of simultaneous SNP discovery and genotyping with no upfront costs, but suffer from well-known pitfalls and limitations in data quality, reproducibility and especially portability across experiments, particularly for highly heterozygous genomes [26, 48, 49]. Although SNPs datasets are generated, the ascertained SNPs vary across experiments and as such they do not constitute a true legacy genomic resource for future widespread and long-term use by the community, nor they allow consolidation of data across studies in the same manner as the fixed content chip-based SNPs data and resource we have described and provided in this work.

Although fixed content chips are currently the gold standard for SNP genotyping, successful attempts to transfer platforms across species and genera have been limited to domestic animals [50], showing a linear decrease of 1.5% in SNP call rate per million years divergence and exponential decay of polymorphisms retention [51], consistent with theoretical expectations [52]. In plants, to the best of our knowledge no large-scale transferability evaluations of SNP arrays across genera have been reported. The 6,879 out of 60,904 SNPs successfully transferred from *Eucalyptus* to *Psidium* in this work corroborate our earlier estimates of ~10%
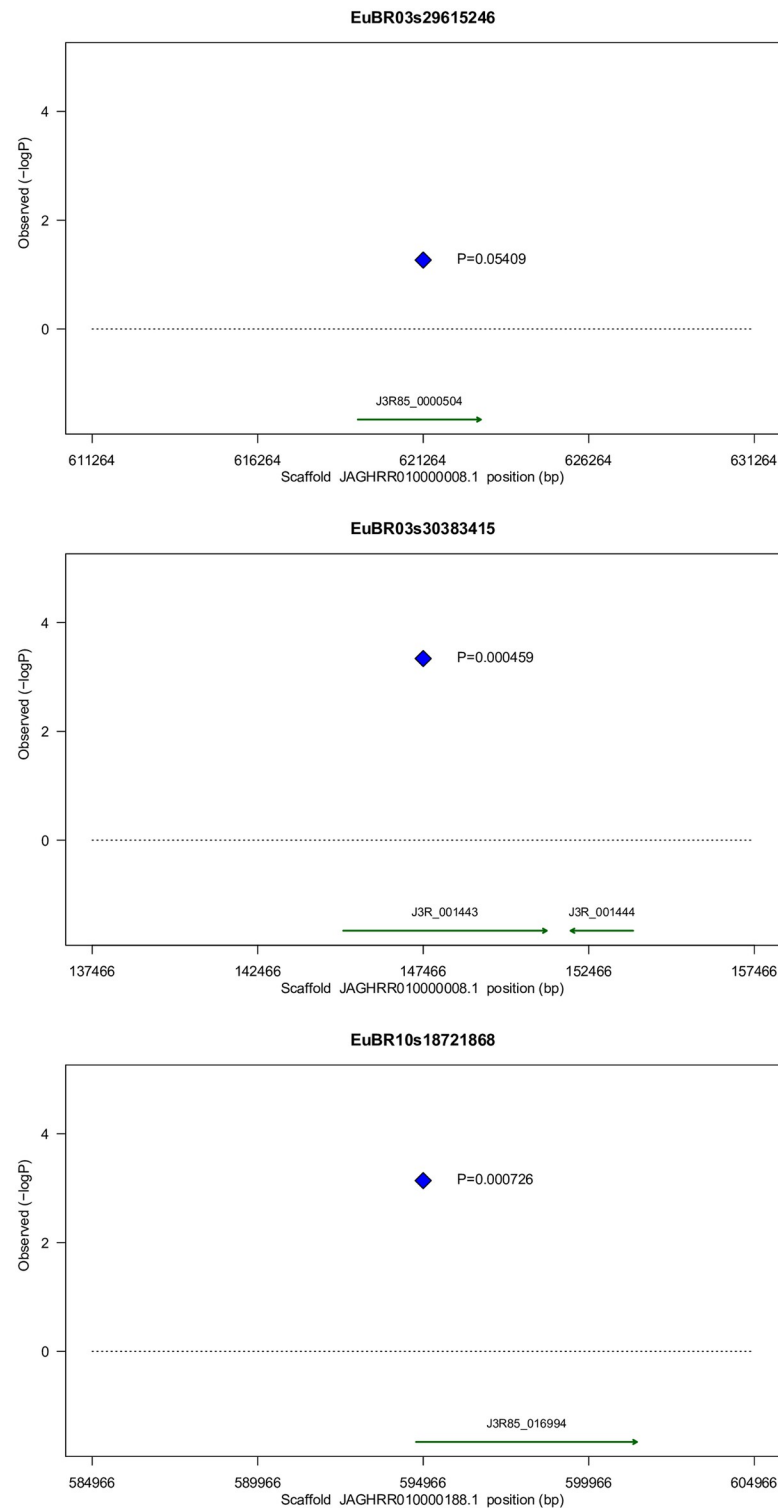
Fig 3. Schematic diagrams of the physical scaffold location of the three main SNPs on chromosomes 3 and 10 of the *Psidium guajava* genome associated with RKN resistance with their corresponding p-values and associated functionally relevant defense genes annotated (see text and Table 3 for details). Coordinates are centered at the SNP (blue diamond) and expand 10 Kbp to both sides.

https://doi.org/10.1371/journal.pone.0273959.g003

Table 3. Summary of the main features of two SNP variants associated with RKN resistance on chromosome 3 and chromosome 10 of *Psidium guajava* likely to cause functional or regulation changes on predicted genes (see text for details).

Feature	EuBR03s3083415	EuBR03s29615246	EuBR10s18721868
P. guajava assembly scaffold	JAGHRR010000008.1	JAGHRR010000008.1	JAGHRR010000188.1
P. guajava assembly locus	147,466	621,264	594,966
Reference allele	G	G	C
Major allele (A) associated with the RKN resistant phenotype)	A	A	T
Minor allele (B) associated with the RKN susceptible phenotype	G	G	C
Minor allele frequency	0.253	0.467	0.271
Counts of genotype AA	102	86	97
Counts of genotype AB	57	9	56
Counts of genotype BB	6	71	8
Downstream gene variant	ncbi: J3R85_001444; uniprot:EUGRSUZ_C01790	-	-
Intron variant	-	-	-
Missense variant	ncbi: J3R85_001443; uniprot ortholog:EUGRSUZ_C01791	ncbi: J3R85_001472; uniprot ortholog:EUGRSUZ_C01744	ncbi:J3R85_0015914; uniprot ortholog:EUGRSUZ_J01461
synonymous_variant			
Protein name	EUGRSUZ_C01790: ENT domain-containing protein; EUGRSUZ_C01791:DUF642	EUGRSUZ_C01744: F-box domain-containing protein	EUGRSUZ_J01461: Histone acetyltransférase
SNP probe sequence [variant SNP site]	GTGCCCTATGAGTCGRAGGGCAAAGGCGGGTTCAAGCGCGCTGTCCTGCCGGTTCCAGGCC [A/G] TGTCCATGAGGACCAGGATCATGTTCTACAGCACGTTCTACACCATGAGGAGTGACGATT	ATTGGGCACTATGCCACGGCTGTTGCTGATTTAACCCTTACTAGCCTCCATAATGTCACT [A/G] AGAGAGGGCTTTGGGTCATGGCAATGGTCATGGTTTGCAAAGGTTGAGGTCTTTGATAG	CGTACAATGATACATTCTGCCCATCGTGCCACAATTGTTGCAGAGGTGGGACCATCAAGTG [C/T] GGGGGGATTTCTGTTTCCAAGGCAAGCTCCACAAGGAAAGAATAATGAGGAAATTGAA

transferability [18] consistent with the estimated divergence time of 60–65 Million years (90% decrease in SNP calling rate) between tribes Eucalypteae and Myrteae [53], with a slightly higher than theoretically expected polymorphism retention of 6.8% (4,143 SNPs in 60,904) following data clustering with the optiCall algorithm. Although almost 7,000 SNPs, out of which 4,143 were informative in this particular population, might seem a small number, it represents a substantial increase in high quality marker availability for *P. guajava*, opening great prospects for genetic and breeding applications. We speculate that this higher-than-expected SNP retention might be due to the deliberate design of the Eucalyptus multispecies EUChip60K toward conserved genomic regions [18], likely subject to stronger purifying selection allowing better opportunity for polymorphism retention of targeted sites.

M. enterolobii resistance in *P. guajava* reveals evolutionary conservation of chromosome 3 loci across genera of Myrtaceae

Despite the limited power of our genetic association experiment, all analytical approaches consistently pointed to a clear-cut signal involving two SNPs (EuBR03s29615246 and EuBR03s30383415) associated with *M. enterolobii* resistance on a specific sequence of the assembly of *Psidium guajava* (JAGHRR010000008.1) syntenic to chromosome 3 of *Eucalyptus grandis* (Chr03:27,580,466–35,980,336). This result is supported by the accurate sequence-level reallocation of SNP probesets from *Eucalyptus* chromosome 3 on our *de novo P. guajava* genome assembly, and corroborated by the highly conserved syntenic relationship between these two genomes [21]. In fact, a number of studies in *Eucalyptus* have reported disease resistance loci for different fungal pathogens also pathogenic in *Psidium* on chromosome 3 [54–59]. Genome annotations have also highlighted the highest densities of clusters and superclusters of NBS-LRR (nucleotide binding site-leucine-rich repeat) resistance genes on *Eucalyptus* chromosomes 3, 5, 6, 8 and 10 [60]. Observations at the gene level have shown extensive syntenic blocks of 778 genes in 522 gene families shared between *Eucalyptus* chromosome 3 and *Populus* chromosome XVIII, with the most common family represented by 33 disease resistance genes [17]. All these evidences not only provide further indirect support to our results, but also contribute to highlight a seemingly strong evolutionary conservation of the role of chromosome 3 across genera of Myrtaceae and possibly beyond, in the adaptation to different pathogens. Furthermore, the detection of major effect loci encompassing regions with NBS-LRR genes has been a common feature for resistance to RKNs across many pathosystems of both annual [28] and woody perennial [15] crops. To the best of our knowledge, our report is the first for a species in the large family Myrtaceae.

Associated SNPs are located within or in close proximity to functionally relevant defense genes in the *Psidium* genome

The precise nucleotide-level reallocation of *Eucalyptus* SNP probesets onto the *Psidium* genome and functional effect prediction, allowed an in-depth examination of the two SNPs located in gene loci likely linked to the defense response on chromosome 3. This, in turn, allowed proposing plausible molecular mechanisms underlying the resistance response at the gene level with an emphasis for the SNPs on chromosome 3 (Fig 3; Table 3). Although we are fully aware that conclusive proof will require additional experimental evidence beyond the scope of this initial SNP discovery, we contend that the following discussion provides a potential roadmap to guide follow-up experiments. Furthermore, we tested and excluded the hypothesis that the gene loci underlying the significant SNPs could be orthologs to the *Ma* gene, a large-spectrum *Meloidogyne* species resistance locus described in *Prunus* linkage group

7 (Prupe.7G065400) [14]. Thus, the gene loci discussed below more likely represent novel sources of RKN resistance, possibly specific to Myrtaceae.

SNP EuBR03s29615246 is located in the gene J3R85_001472 and its *Eucalyptus* ortholog gene Eucgr.C01744 (Uniprot id: A0A059CPK6). This gene codes for a F-box protein belonging to the leucine-rich repeat (LRR) family involved in cell cycle control and glucose signaling, with orthology relationship to Arabidopsis EBF1/EBF2 [61]. In Arabidopsis, EBF1 and EBF2 play a specific role in the recognition of the EIN3 (ethylene-insensitive3) transcription factor (s) and facilitate their subsequent SCF-dependent ubiquitylation and degradation via the ubiquitin/26S proteasome pathway [62]. This pathway is established as an important posttranscriptional mechanism that allows eukaryotic cells to respond rapidly to signal molecules and changes in environmental conditions [63].

On the same contig where EuBR03s29615246 resides, the second SNP, EuBR03s30383415, was detected as associated. This SNP was declared significant by all analytical approaches, ranked as the strongest association with the lowest p-value in the MLM analysis (Tables 1 and 2; Fig 3). This SNP results in a non-synonymous change on the product of the gene J3R85_001443, which is an ortholog of Eucgr.C01791 (Uniprot id: A0A059CPT9) in *Eucalyptus*. Its product is a protein that contains a DUF642 domain and a galactose-binding-like domain fold, and it is member of a group of seven proteins in *Eucalyptus* that have orthology relationships to a group of ten proteins in *Arabidopsis*. Remarkably, two members of this group are AT4G32460 (BIIDXI locus) and AT2G41800 (TEEBE locus), recently reported as highly induced genes by auxin during early interaction between the susceptible *A. thaliana* ecotype Columbia and the RKN *M. incognita* [64]. SNP EuBR03s30383415 is also predicted to cause an allelic change in the region downstream of another gene of *P. guajava*, J3R85_001444, which has orthology relationship to Eucgr.C01790 (Uniprot id: A0A059CQ27) in *Eucalyptus*. This gene codes for a protein containing an EMSY N-terminal domain (PF03735), a central Agenet domain (PTHR31917), and a probable coiled-coil motif at the C-terminus. The Agenet domain is member of the superfamily Agenet/Tudor [65] and in plants it was suggested to act as a link between DNA replication, transcription and chromatin remodeling during flower development [66], a process involving hormonal regulation in which auxin plays a major role [67].

Based on these findings, we speculate about the possibility of SNPs EuBR03s29615246 and EuBR03s30383415 being important leads to identify underlying genes in the *Psidium guajava* assembly, highly syntenic to chromosome 3 in *Eucalyptus*, whose expression is likely to undergo changes during pathogen invasion due to effects of chemical modulation of plant hormone levels or signaling contributing to the resistant response in P. *guajava* × *P. guineense* plants. Although this hypothesis awaits gene expression experiments to be tested, we have advanced our discussion based on available evidences regarding the disease status of RKN infected trees.

M. enterolobii resistance might be triggered by mineral nutrients and phytohormone homeostasis or signaling subject to miRNAs modulation

Studies in guava [68], coffee [69], peach and almond [70] have shown that disease status of RKN infected trees is related to symptoms triggered by nutritional imbalances in the concentration of nitrogen, calcium, manganese and magnesium in several tissues in adult plants. Furthermore, because our binary phenotypic data fitted the same epistatic models proposed earlier [47], we further considered the hypothetical involvement of a second genetic locus interacting with the underlying genes on chromosome 3. Particularly, the associated SNP EuBR10s18721868 (Table 2; Fig 3) residing on the 700 kb long sequence JAGHRR010000188.1

of the *P. guajava* assembly looked promising enough in demanding further investigation. EuBR10s18721868 is a synonymous variant inside of gene J3R85_016994 whose ortholog in *Eucalyptus* is Eucgr.J01461 on chromosome 10. This gene encodes a CREB-binding protein (CBP)/p300 protein of the subfamily of highly conserved histone acetyltransferase (HAT) and histone deacetylase (HDAC). These proteins are involved in various physiological events and their homologs in plants, the HAC genes, were recently suggested to be involved in ethylene signaling [71].

Studies in animals, plants, and viruses have suggested that microRNA function may affect synonymous codon choices in the vicinity of its target sites [72]. Following that suggestion, we performed a preliminary search for individual miRNA in the scaffold JAGHRR010000188.1, that resulted in two different families of miRNA namely miR172 and miR393. MicroRNA miR393 has been shown to take part in plant metal homeostasis, uptake and accumulation of various nutrient ions under low-nutrient conditions including nitrogen and divalent cations [73, 74]. microRNA identification on this sequence showed the highest number of loci for miR172 and it also contains one of the only two loci that were found to code for miR393 in the genome assembly of *P. guajava*.

Interestingly, nitrogen supply has shown to promote the upregulation of miR393 for targeting auxin receptor genes that encode F-box proteins such as that encoded by the gene J3R85_001472 impacted by the SNP EuBR03s29615246 on chromosome 3. These proteins are involved in ubiquitin-mediated degradation of specific substrates during auxin signaling cations [74]. Moreover, miR393 targets transcripts that code for basic helix-loop-helix (bHLH) transcription factors and for the auxin receptors TIR1, AFB1, AFB2, and AFB3. Studies have shown that miR393/AFB3 is a unique N-responsive module that controls root system architecture in response to external and internal N availability in Arabidopsis [75]. The miR172 family has a recognized role in damping the expression of genes encoding the APETALA2 (AP2) transcription factor members of a large family of the AP2/EREBP family from *Arabidopsis thaliana* and other plants, which are responsible in part for mediating the response to ethylene [74].

Although we are aware that the preliminary analysis described above lacks direct supportive experimental data to ascertain the impact of miRNAs on *Psidium* response to RKN infection, these conjectures fit both the epistatic interaction model described previously [47] and the emerging view that specific developmental or stress events can be frequently subject to modulation by diverse miRNA families [74]. Therefore, it could be postulated that this region on chromosome 10 might constitute a distinct locus epistatic to the genetic locus on chromosome 3 harboring SNPs EuBR03s29615246 and EuBR03s30383415 potentially contributing to the *M. enterolobii* resistance response.

Concluding remarks

We mapped two genomic regions associated with resistance to *M. enterolobii* in a member of the Myrtaceae. Comparative genomics driven by significantly associated SNPs and a *de novo* assembly of the *P. guajava* genome support our conclusion that the continuous physical stretch of nearly 3.14 Mbp, encompassing a single sequence of *Psidium guajava* largely syntenic to *Eucalyptus grandis* chromosome 3, harboring SNPs EuBR03s29615246 and EuBR03s30383415, may constitute a resistance locus for RKN response derived from the wild relative *P. guineensis*. Detailed functional annotations and positioning of the gene loci targeted by these SNPs allowed us to provide a roadmap to putative underlying molecular mechanisms of resistance response at the gene level, to guide follow-up experimental work. Further indirect support to our data comes from a number of previous reports mapping resistance loci in the

syntenic genomic region in *Eucalyptus*, a region housing the highest density of NBS-LRR resistance genes and showing elevated synteny and gene retention in the genome evolution of woody plants. Additional SNPs were detected on other chromosomes suggesting that other loci besides those on chromosome 3 likely participate in the response, notably on chromosome 10, also supported by indirect evidences from *Eucalyptus* mapping studies.

Controlling RKN species by the introgression of resistance loci represents a promising alternative to environmentally undesirable nematicides. This has been the preferred approach in a growing number of mainstream woody crops, such as *Prunus*, grapevine and coffee. Our work represents the first association study in a fruit crop of the large group of fleshy members of the Myrtaceae family, engaging the guava fruit in this select group of fruit crop that have been the subject of genomic investigation. Our contribution should advance the development of new guava cultivars through the development of inexpensive SNP based assays to monitor the introgression of the *P. guineense*-derived resistance into guava cultivars and rootstocks, and foster comparative genomic studies in other species of the genus that endure the damaging effects of this pathogenic nematode.

Supporting information

S1 File. Phenotypic data for the binary and quantitative (RF) RKN resistance traits and SNP genotype data for the 4,143 probesets in the EucHIP60K that have passed quality control using Genome Studio and displayed polymorphism following the optiCall analysis (SNP Call Rate ≥90% and MAF ≥0.05). Normalized X-Y data for the 6,879 successfully assayed SNPs (Call Rate ≥90%) were exported from GS and genotypes were ascertained using the optiCall algorithm (default options).
(XLSX)

S2 File. SNP annotation file (VCF format) for the 14,268 *Eucalyptus* probesets in the EucHIP60K successful reallocated onto the *Psidium* guajava genome assembly using whole-genome alignment and extraction of synteny blocks. This file highlights the 1,241 SNPs that were polymorphic in our study (GT ="0/1") and also displays the functional effect predictions resulting from the allelic changes at these genomic variants on the protein-coding loci in the *P. guajava* assembly using SnpEff terms (ANN field). Homology relationship with the protein-coding loci in the *E. grandis* assembly was inferred with the Comparative Annotation Toolkit program.
(VCF)

S3 File. Diploid Cluster file (EGT format of Illumina GenomeStudio™ 2.0 software) for the EucHIP60K Infinium assay for *Psidium sp*. genotyping. Cluster positions were derived from a set of 70 diverse samples of *Psidium guajava*, *Psidium guineensis* and their F_1 hybrids using the GenTrain algorithm. It corresponds to the SNP data from the 14,268 successful reallocated probesets onto the *P. guajava* assembly while the remaining SNP data were "zeroed". This file is provided to help processing and quality control of SNP data for future *Psidium sp*. genotyping using the EucHIP60K.
(EGT)

Acknowledgments

Priscila Grynberg and Dr. Marcos Costa, from Embrapa Genetic Resources and Biotechnology, for carrying the preliminary bioinformatics analysis on the miRNA identification.

Author Contributions

Conceptualization: Carlos Antonio Fernandes Santos, Leonardo Silva Boiteux.

Data curation: Soniane Rodrigues da Costa, Orzenil Bonfim Silva-Junior.

Formal analysis: Carlos Antonio Fernandes Santos, Orzenil Bonfim Silva-Junior.

Funding acquisition: Carlos Antonio Fernandes Santos, Dario Grattapaglia.

Investigation: Carlos Antonio Fernandes Santos, Soniane Rodrigues da Costa, Leonardo Silva Boiteux, Dario Grattapaglia.

Methodology: Carlos Antonio Fernandes Santos, Soniane Rodrigues da Costa, Orzenil Bonfim Silva-Junior.

Project administration: Dario Grattapaglia.

Resources: Carlos Antonio Fernandes Santos, Leonardo Silva Boiteux, Dario Grattapaglia.

Software: Orzenil Bonfim Silva-Junior.

Validation: Leonardo Silva Boiteux, Orzenil Bonfim Silva-Junior.

Writing – original draft: Carlos Antonio Fernandes Santos, Leonardo Silva Boiteux, Dario Grattapaglia, Orzenil Bonfim Silva-Junior.

Writing – review & editing: Dario Grattapaglia, Orzenil Bonfim Silva-Junior.

References

1. Biffin E, Lucas EJ, Craven LA, Ribeiro da Costa I, Harrington MG, Crisp MD. Evolution of exceptional species richness among lineages of fleshy-fruited Myrtaceae. Annals of Botany. 2010; 106(1):79–93. https://doi.org/10.1093/aob/mcq088 PMID: 20462850

2. Mitra SK. Important Myrtaceae fruit crops. Acta Horticulturae. 2010; 849:33–8. https://doi.org/10.17660/ActaHortic.2010.849.2

3. Gutiérrez RMP, Mitchell S, Solis RV. Psidium guajava: A review of its traditional uses, phytochemistry and pharmacology. Journal of Ethnopharmacology. 2008; 117(1):1–27. https://doi.org/10.1016/j.jep.2008.01.025 PMID: 18353572

4. Morais-Braga MFB, Carneiro JNP, Machado AJT, dos Santos ATL, Sales DL, Lima LF, et al. Psidium guajava L., from ethnobiology to scientific evaluation: Elucidating bioactivity against pathogenic microorganisms. Journal of Ethnopharmacology. 2016; 194:1140–52. https://doi.org/10.1016/j.jep.2016.11.017 PMID: 27845266

5. Altendorf S. Minor tropical fruits—Mainstreaming a niche market. In: FAO, editor. Food outlook—Biannual Report on Global Food Markets. 2018. Rome—Italy: FAO; 2018. p. 67–75.

6. FAOSTAT statistical database [Internet]. Food and Agricultural Organization, Rome (Date accessed: March 10, 2022). 2021 [cited March 10, 2022]. https://search.library.wisc.edu/catalog/9999890171702121.

7. Sitther V, Dhekney S, Pramanik S, Zee FTP, Yadav AK. Assessment of Genetic Diversity in Psidium guajava L. using Simple Sequence Repeat Markers. Hortscience. 2012; 47(9):S200–S1.

8. Carneiro RMDG, Moreira WA, Almeida MRA, Gomes ACMM. Primeiro registro de Meloidogyne mayaguensis em goiabeira no Brasil. Nematologia Brasileira. 2001; 25(2):223–8.

9. Carneiro RMDG, de Freitas VM, Mattos JK, Castro JMC, Gomes CB, Carneiro RG. Major guava nematodes and control prospects using resistance on *Psidium* spp. and non-host crops. ActaHortic. 2012; 959:41–9. Epub 959. https://doi.org/10.17660/ActaHortic.2012.959.4

10. Singh N. Emerging problem of guava decline caused by Meloidogyne enterolobii and Fusarium oxysporum f.sp. psidii. Indian Phytopathology. 2020; 73(2):373–4. https://doi.org/10.1007/s42360-020-00198-y

11. Castagnone-Sereno P. Meloidogyne enterolobii (= M. mayaguensis): profile of an emerging, highly pathogenic, root-knot nematode species. Nematology. 2012; 14(2):133–8. https://doi.org/10.1163/156854111X601650

12. Koutsovoulos GD, Poullet M, Elashry A, Kozlowski DKL, Sallet E, Da Rocha M, et al. Genome assembly and annotation of Meloidogyne enterolobii, an emerging parthenogenetic root-knot nematode. Scientific Data. 2020; 7(1):324. https://doi.org/10.1038/s41597-020-00666-0 PMID: 33020495

13. Claverie M, Dirlewanger E, Bosselut N, Van Ghelder C, Voisin R, Kleinhentz M, et al. The Ma gene for complete-spectrum resistance to *Meloidogyne* species in *Prunus* is a TNL with a huge repeated c-terminal post-LRR region. Plant Physiology. 2011; 156(2):779. https://doi.org/10.1104/pp.111.176230 PMID: 21482634

14. Van Ghelder C, Esmenjaud D, Callot C, Dubois E, Mazier M, Duval H. Ma Orthologous Genes in Prunus spp. Shed Light on a Noteworthy NBS-LRR Cluster Conferring Differential Resistance to Root-Knot Nematodes. Frontiers in Plant Science. 2018; 9:1269. https://doi.org/10.3389/fpls.2018.01269 PMID: 30254651

15. Saucet SB, Van Ghelder C, Abad P, Duval H, Esmenjaud D. Resistance to root-knot nematodes Meloidogyne spp. in woody plants. New Phytologist. 2016; 211(1):41–56. https://doi.org/10.1111/nph.13933 PMID: 27128375

16. Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Külheim C, et al. Progress in Myrtaceae genetics and genomics: Eucalyptus as the pivotal genus. Tree Genetics & Genomes. 2012; 8 (3):463–508. https://doi.org/10.1007/s11295-012-0491-x

17. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, et al. The genome of Eucalyptus grandis. Nature. 2014; 510(7505):356–62. https://doi.org/10.1038/nature13308 PMID: 24919147

18. Silva-Junior OB, Faria DA, Grattapaglia D. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 Eucalyptus tree genomes across 12 species. New Phytologist. 2015; 206(4):1527–40. https://doi.org/10.1111/nph.13322 PMID: 25684350

19. Coser SM, Ferreira MFdS, Ferreira A, Mitre LK, Carvalho CR, Clarindo WR. Assessment of genetic diversity in Psidium guajava L. using different approaches. Scientia Horticulturae. 2012; 148:223–9. https://doi.org/10.1016/j.scienta.2012.09.030

20. Tuler AC, Carrijo TT, Peixoto AL, Garbin ML, da Silva Ferreira MF, Carvalho CR, et al. Diversification and geographical distribution of Psidium (Myrtaceae) species with distinct ploidy levels. Trees. 2019; 33 (4):1101–10. https://doi.org/10.1007/s00468-019-01845-2

21. Feng C, Feng C, Lin X, Liu S, Li Y, Kang M. A chromosome-level genome assembly provides insights into ascorbic acid accumulation and fruit softening in guava (Psidium guajava). Plant Biotechnology Journal. 2020; 19(4):717–30. https://doi.org/10.1111/pbi.13498 PMID: 33098334

22. Kumar C, Kumar R, Singh SK, Goswami AK, Nagaraja A, Paliwal R, et al. Development of novel g-SSR markers in guava (Psidium guajava L.) cv. Allahabad Safeda and their application in genetic diversity, population structure and cross species transferability studies. PLOS ONE. 2020; 15(8):e0237538. https://doi.org/10.1371/journal.pone.0237538 PMID: 32804981

23. Risterucci AM, Nansot G, Grangeon R, Lepitre V, de Reeper A, Argout X, et al. Development of Guava Microsatellite (SSR) Markers Using the SAT Software. In: Rohde W, Fermin G, editors. II International Symposium on Guava and Other Myrtaceae. Acta Horticulturae. 849. Leuven 1: Int Soc Horticultural Science; 2010. p. 113–9.

24. Diaz-Garcia L, Padilla-Ramirez JS. Development of single nucleotide polymorphism markers and genetic diversity in guava (Psidium guajava L.). Plants People Planet. https://doi.org/10.1002/ppp3.10295

25. Grossi LL, Fernandes M, Silva MA, de Oliveira Bernardes C, Tuler AC, dos Santos PHD, et al. DArTseq-derived SNPs for the genus Psidium reveal the high diversity of native species. Tree Genetics & Genomes. 2021; 17(2):23. https://doi.org/10.1007/s11295-021-01505-y

26. Myles S. Improving fruit and wine: what does genomics have to offer? Trends in Genetics. 2013; 29 (4):190–6. https://doi.org/10.1016/j.tig.2013.01.006 PMID: 23428114

27. Lowry DB, Hoban S, Kelley J., L., Lotterhos K., E., Reed L., K., Antolin M., F., et al. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Molecular Ecology Resources. 2016; 17(2):142–52. https://doi.org/10.1111/1755-0998.12635 PMID: 27860289

28. Kaloshian I, Teixeira M. Advances in plant–nematode interactions with emphasis on the notorious nematode genus *Meloidogyne*. Phytopathology. 2019; 109(12):1988–96. https://doi.org/10.1094/PHYTO-05-19-0163-IA PMID: 31613704

29. Hussey RS, Janssen GJW. Root-knot nematode: Meloidogyne species. In: Starr JI, Cook R, Bridge J, editors. Plant Resistance to Parasitic Nematodes. Wallingford, UK: CAB International; 2002. p. 43–70.

30. Oostenbrink M. Major characteristics of the relation between nematodes and plants. Wageningen: Veenman. Report no. 66–4 46 p., 1966.

31. Inglis PW, Pappas MdCR, Resende LV, Grattapaglia D. Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. PLosOne. 2018; 13(10):e0206085. https://doi.org/10.1371/journal.pone.0206085 PMID: 30335843

32. Shah TS, Liu JZ, Floyd JAB, Morris JA, Wirth N, Barrett JC, et al. optiCall: a robust genotype-calling algorithm for rare, low-frequency and common variants. Bioinformatics. 2012; 28(12):1598–603. Epub 2012/04/12. https://doi.org/10.1093/bioinformatics/bts180 PMID: 22500001.

33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81 (3):559–75. Epub 2007/07/25. https://doi.org/10.1086/519795 PMID: 17701901.

34. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007; 23(19):2633–5. https://doi.org/10.1093/bioinformatics/btm308 PMID: 17586829

35. Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nature Methods. 2016; 13(12):1050–4. https://doi.org/10.1038/nmeth.4035 PMID: 27749838

36. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research. 2012; 40(D1):D1178–D86. https://doi.org/10.1093/nar/gkr944 PMID: 22110026

37. Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature. 2020; 587(7833):246–51. https://doi.org/10.1038/s41586-020-2871-y PMID: 33177663

38. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008; 24(5):637–44. https://doi.org/10.1093/bioinformatics/btn013 PMID: 18218656

39. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Molecular Biology and Evolution. 2018; 35(3):543–8. https://doi.org/10.1093/molbev/msx319 PMID: 29220515

40. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. In: M. K, editor. Gene Prediction Methods in Molecular Biology. 1962. New York: Humana; 2019. p. 227–45.

41. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014; 30(9):1312–3. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

42. Hickey G, Paten B, Earl D, Zerbino D, Haussler D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. Bioinformatics. 2013; 29(10):1341–2. https://doi.org/10.1093/bioinformatics/btt128 PMID: 23505295

43. Kolmogorov M, Armstrong J, Raney BJ, Streeter I, Dunn M, Yang F, et al. Chromosome assembly of large and complex genomes using multiple references. Genome Research. 2018; 28(11):1720–32. https://doi.org/10.1101/gr.236273.118 PMID: 30341161

44. Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom Kate R, Raney BJ, et al. The UCSC Genome Browser database: 2021 update. Nucleic Acids Research. 2021; 49(D1):D1046–D57. https://doi.org/10.1093/nar/gkaa1070 PMID: 33221922

45. Fiddes IT, Armstrong J, Diekhans M, Nachtweide S, Kronenberg ZN, Underwood JG, et al. Comparative Annotation Toolkit (CAT)—simultaneous clade and personal genome annotation. Genome Research. 2018; 28(7):1029–38. https://doi.org/10.1101/gr.233460.117 PMID: 29884752

46. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. 2012; 6(2):80–92. https://doi.org/10.4161/fly.19695 PMID: 22728672

47. Costa SR, Santos CAF, da Cunha e Castro JM. Inheritance of resistance to Meloidogyne enterolobii in Psidium guajava x P. guineense hybrid. European Journal of Plant Pathology. 2017; 148(2):405–11. https://doi.org/10.1007/s10658-016-1098-0

48. Scheben A, Batley J, Edwards D. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnology Journal. 2017; 15 (2):149–61. https://doi.org/10.1111/pbi.12645 PMID: 27696619

49. Darrier B, Russell J, Milner SG, Hedley PE, Shaw PD, Macaulay M, et al. A Comparison of Mainstream Genotyping Platforms for the Evaluation and Use of Barley Genetic Resources. Frontiers in Plant Science. 2019; 10(544). https://doi.org/10.3389/fpls.2019.00544 PMID: 31105733

50. Decker JE, Pires JC, Conant GC, McKay SD, Heaton MP, Chen K, et al. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proceedings of the National Academy of Sciences. 2009; 106(44):18644. https://doi.org/10.1073/pnas.0904691106 PMID: 19846765

51. Miller JM, Kijas JW, Heaton MP, McEwan JC, Coltman DW. Consistent divergence times and allele sharing measured from cross-species application of SNP chips developed for three domestic species. Molecular Ecology Resources. 2012; 12(6):1145–50. ISI:000309739700019. https://doi.org/10.1111/1755-0998.12017 PMID: 22994965

52. Kimura M, Ota T. The average number of generations until extinction of an individual mutant gene in a finite population. Genetics. 1969; 63(3):701–9. https://doi.org/10.1093/genetics/63.3.701 PMID: 5399256.

53. Thornhill AH, Ho SYW, Külheim C, Crisp MD. Interpreting the modern distribution of Myrtaceae using a dated molecular phylogeny. Molecular Phylogenetics and Evolution. 2015; 93:29–43. https://doi.org/10.1016/j.ympev.2015.07.007 PMID: 26211451

54. Butler JB, Freeman JS, Vaillancourt RE, Potts BM, Glen M, Lee DJ, et al. Evidence for different QTL underlying the immune and hypersensitive responses of Eucalyptus globulus to the rust pathogen Puccinia psidii. Tree Genetics & Genomes. 2016; 12(3):1–13. https://doi.org/10.1007/s11295-016-0987-x

55. Freeman JS, Potts BM, Vaillancourt RE. Few Mendelian genes underlie the quantitative response of a forest tree, Eucalyptus globulus, to a natural fungal epidemic. Genetics. 2008; 178(1):563–71. ISI:000252742200044. https://doi.org/10.1534/genetics.107.081414 PMID: 18202395

56. Junghans DT, Alfenas AC, Brommonschenkel SH, Oda S, Mello EJ, Grattapaglia D. Resistance to rust (*Puccinia psidii* Winter) in *Eucalyptus*: mode of inheritance and mapping of a major gene with RAPD markers. Theor Appl Genet. 2003; 108(1):175–80. https://doi.org/10.1007/s00122-003-1415-9 PMID: 14504745.

57. Mamani EMC, Bueno NW, Faria DA, Guimaraes LMS, Lau D, Alfenas AC, et al. Positioning of the major locus for Puccinia psidii rust resistance (Ppr1) on the Eucalyptus reference map and its validation across unrelated pedigrees. Tree Genetics & Genomes. 2010; 6(6):953–62. https://doi.org/10.1007/s11295-010-0304-z

58. Resende RT, Resende MDV, Silva FF, Azevedo CF, Takahashi EK, Silva-Junior OB, et al. Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in Eucalyptus. New Phytologist. 2017; 213(3):1287–300. https://doi.org/10.1111/nph.14266 PMID: 28079935

59. Rosado CCG, da Silva Guimarães LM, Faria DA, de Resende MDV, Cruz CD, Grattapaglia D, et al. QTL mapping for resistance to Ceratocystis wilt in Eucalyptus. Tree Genetics & Genomes. 2016; 12(4):1–10. https://doi.org/10.1007/s11295-016-1029-4

60. Christie N, Tobias PA, Naidoo S, Kulheim C. The Eucalyptus grandis NBS-LRR Gene Family: Physical Clustering and Expression Hotspots. Frontiers in Plant Science. 2016; 6:1238. Artn 1238 https://doi.org/10.3389/fpls.2015.01238 PMID: 26793216

61. Van Bel M, Diels T, Vancaester E, Kreft L, Botzki A, Van de Peer Y, et al. PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. Nucleic Acids Research. 2018; 46(D1):D1190–D6. https://doi.org/10.1093/nar/gkx1002 PMID: 29069403

62. Potuschak T, Lechner E, Parmentier Y, Yanagisawa S, Grava S, Koncz C, et al. EIN3-Dependent Regulation of Plant Ethylene Hormone Signaling by Two *Arabidopsis* F Box Proteins: EBF1 and EBF2. Cell. 2003; 115(6):679–89. https://doi.org/10.1016/S0092-8674(03)00968-1

63. Hershko A, Ciechanover A. The Ubiquitin system. Annual Review of Biochemistry. 1998; 67(1):425–79. https://doi.org/10.1146/annurev.biochem.67.1.425 PMID: 9759494

64. Salazar-Iribe A, Zúñiga-Sánchez E, Mejía EZ, Gamboa-deBuen A. Cell Wall Localization of Two DUF642 Proteins, BIIDXI and TEEBE, during Meloidogyne incognita Early Inoculation. Plant Pathol J. 2017; 33(6):614–8. https://doi.org/10.5423/PPJ.NT.05.2017.0101 PMID: 29238286

65. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends in Biochemical Sciences. 2003; 28(2):69–74. https://doi.org/10.1016/S0968-0004(03)00004-5 PMID: 12575993

66. Brasil JN, Cabral LM, Eloy NB, Primo LMF, Barroso-Neto IL, Grangeiro LPP, et al. AIP1 is a novel Agenet/Tudor domain protein from Arabidopsis that interacts with regulators of DNA replication, transcription and chromatin remodeling. BMC Plant Biology. 2015; 15(1):270. https://doi.org/10.1186/s12870-015-0641-z PMID: 26538092

67. Chandler JW. The Hormonal Regulation of Flower Development. Journal of Plant Growth Regulation. 2011; 30(2):242–54. https://doi.org/10.1007/s00344-010-9180-x

68. Almeida EJ, Santos JM, Martins ABG. Influência do parasitismo pelo nematoide de galhas nos níveis de nutrientes em folhas e na fenologia de goiabeira 'Paluma'. Bragantia. 2011; 70:876–81.

69. Hurchanik D, Schmitt D.P., Hue NV, Sipes BS. Plant Nutrient Partitioning in Coffee Infected with Meloi-dogyne konaensis. The Journal of Nematology. 2004; 36(1):76–84. PMID: 19262790

70. Nasr TA, Ibrahim IKA, El-Azab EM, Hassan MWA. Effect of root-knot nematodes on the mineral, amino acid and carbohydrate concentrations of almond and peach rootstocks. Nematologica. 1980; 26 (2):133–8. https://doi.org/10.1163/187529280X00017

71. Li C, Xu J, Li J, Li Q, Yang H. Involvement of Arabidopsis Histone Acetyltransferase HAC Family Genes in the Ethylene Signaling Pathway. Plant and Cell Physiology. 2014; 55(2):426–35. https://doi.org/10.1093/pcp/pct180 PMID: 24287137

72. Gu W, Wang X, Zhai C, Xie X, Zhou T. Selection on Synonymous Sites for Increased Accessibility around miRNA Binding Sites in Plants. Molecular Biology and Evolution. 2012; 29(10):3037–44. https://doi.org/10.1093/molbev/mss109 PMID: 22490819

73. Iwamoto M, Tagiri A. MicroRNA-targeted transcription factor gene RDD1 promotes nutrient ion uptake and accumulation in rice. The Plant Journal. 2016; 85(4):466–77. https://doi.org/10.1111/tpj.13117 PMID: 26729506

74. Tang J, Chu C. MicroRNAs in crop improvement: fine-tuners for complex traits. Nature Plants. 2017; 3 (7):17077. https://doi.org/10.1038/nplants.2017.77 PMID: 28665396

75. Vidal EA, Araus V, Lu C, Parry G, Green PJ, Coruzzi GM, et al. Nitrate-responsive miR393/*AFB3* regu-latory module controls root system architecture in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences. 2010; 107(9):4477. https://doi.org/10.1073/pnas.0909571107 PMID: 20142497