

Article

Genome-Wide Screening and Characterization of Non-Coding RNAs in *Coffea canephora*

Samara M. C. Lemos ^{1,2}, Luiz F. C. Fonçatti ¹, Romain Guyot ³ , Alexandre R. Paschoal ¹ 
and Douglas S. Domingues ^{2,*} 

¹ Department of Computer Science, Federal University of Technology-Parana, Cornelio Procopio 86300-000, Brazil; samaramclemos@gmail.com (S.M.C.L.); luifon@alunos.utfpr.edu.br (L.F.C.F.); paschoal@utfpr.edu.br (A.R.P.)

² Group of Genomics and Transcriptomes in Plants, São Paulo State University, Rio Claro, Sao Paulo 13506-900, Brazil

³ Institut de Recherche pour le Développement (IRD), University Montpellier, 34394 Montpellier, France; romain.guyot@ird.fr

* Correspondence: doug@rc.unesp.br; Tel.: +55-19-3526-4207

Received: 12 August 2020; Accepted: 8 September 2020; Published: 11 September 2020



Abstract: *Coffea canephora* grains are highly traded commodities worldwide. Non-coding RNAs (ncRNAs) are transcriptional products involved in genome regulation, environmental responses, and plant development. There is not an extensive genome-wide analysis that uncovers the ncRNA portion of the *C. canephora* genome. This study aimed to provide a curated characterization of six ncRNA classes in the *Coffea canephora* genome. For this purpose, we employed a combination of similarity-based and structural-based computational approaches with stringent curation. Candidate ncRNA loci had expression evidence analyzed using sRNA-seq libraries. We identified 7455 ncRNA loci (6976 with transcriptional evidence) in the *C. canephora* genome. This comprised of total 115 snRNAs, 1031 snoRNAs, 92 miRNA precursors, 602 tRNAs, 72 rRNAs, and 5064 lncRNAs. For miRNAs, we identified 159 putative high-confidence targets. This study was the most extensive genomic catalog of curated ncRNAs in the *Coffea* genus. This data might help elaborating more robust hypotheses in future comparative genomic studies as well as gene regulation and genome dynamics, helping to understand the molecular basis of domestication, environmental adaptation, resistance to pests and diseases, and coffee productivity.

Keywords: coffee; plant ncRNAs; curated annotation; genome-wide; bioinformatics; in silico

1. Introduction

Non-coding RNAs (ncRNAs) are functional regulatory molecules that mediate cellular processes, including chromatin remodeling, transcription, post-transcriptional modifications, and signal transduction [1,2]. They can be roughly divided into two categories: housekeeping and regulatory ncRNAs [3]. Housekeeping ncRNAs, such as ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs), are responsible for basic cellular functions. Regulatory ncRNAs, such as microRNAs (miRNAs), small interfering RNAs (siRNAs), and long non-coding RNAs (lncRNAs), are capable of moving between cells, performing cell-to-cell signaling in development or physiology, and mediating epigenetic inheritance [4].

In plants, most studies usually identify a single class of ncRNAs, such as miRNAs or long non-coding RNAs [4–8]. Although several studies have identified miRNAs in plants [9–11], the identification has not always been performed in the most reliable way [12]. Criteria to identify and annotate miRNAs have been in constant improvement over time [12–15].

Coffee is one of the world’s most popular beverages, and 80% of it is produced by 25 million smallholders. Around 125 million people worldwide depend on coffee for their livelihoods. *Coffea canephora* is one of the most important agricultural commodities, corresponding to nearly 40% of the world coffee production [16] and the first coffee species with a publicly available genome annotation [17,18]. The *Coffea canephora* genome annotation has identified 25,574 protein-coding genes, and approximately 50% of the genome is composed of transposable elements [18]. The repertoire of annotated non-coding RNAs in that study [18] is restricted to microRNAs (92 precursors). To date, information about ncRNAs for *Coffea canephora* is not clearly organized.

A total of seven studies have predicted miRNAs in *C. canephora* using distinct approaches. Five studies have focused only on prediction [18–22]. Two studies have used small RNA sequencing to confirm the transcriptional activity of miRNAs [23,24]. There is only one study [22] that has focused on a comprehensive analysis of the *C. canephora* sequenced genome to annotate miRNAs, but this study has not taken account of recent changes in miRNA annotation rules in plants. For example, the criteria concerning expression analysis and the curation of precursors have become more strict, and they have an impact on defining high-quality prediction of miRNA loci [12,15].

In this context, there is a gap in a highly accurate ncRNAs annotation for coffee. For that, we provided the most extensive and organized ncRNA loci catalog of the *Coffea canephora* genome. We characterized snRNAs, snoRNAs, miRNAs, tRNAs, rRNAs, lncRNAs and performed a manual standardization of the previously identified microRNAs in the species.

2. Results

2.1. Non-Coding RNA Loci Overview

A total of 6976 expressed ncRNA loci were identified in the *C. canephora* genome. About 64% of the *C. canephora* genomic sequence was assembled on 11 chromosomes [18]. Assembled contigs that were not assigned to any chromosome were grouped arbitrarily into a pseudomolecule named “Chromosome 0” [18]. All results from the pseudomolecule “Chromosome 0” are presented separately in Supplementary File S3. For practical reasons, we considered only assembled chromosomes for overall statistics. Among the assembled chromosomes, 3966 expressed ncRNA loci were identified. Chromosome 2 had the highest number of loci (626 ncRNA loci), and chromosome 3 showed the lowest number of ncRNA loci, with 203 (Table S1, Supplementary File S1). The distribution of ncRNAs among chromosomes is detailed in Table S2, Supplementary File S1, and Figure 1 using chromPlot [25]. We did not identify any preferential location of ncRNA loci.

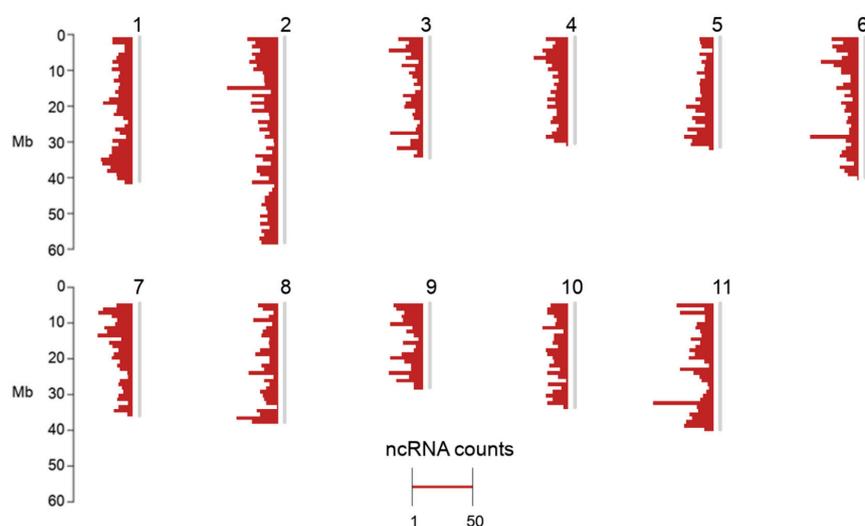


Figure 1. Distribution of ncRNAs in *Coffea canephora* genome.

Ninety-three snRNAs, 778 snoRNAs, 427 tRNAs, 51 rRNAs, 88 miRNAs, and 2529 lncRNAs were determined as high-confidence ncRNAs. The average length of small ncRNAs (snRNAs, snoRNAs, tRNAs, rRNAs, and miRNA precursors) was 126 nt, smaller than the mean length of protein-coding genes CDS (1206 nt) and introns (483 nt) [18]. The number of ncRNA loci in *Coffea canephora* was comparable to other curated ncRNA annotations in other plant species (Table 1). In summary, we noticed that tRNA, rRNA, and snRNA loci were in the range of most analyzed species (Table 1), even considering variation. For miRNA, we expected this result since we used a more stringent criterium of miRNA annotation, following several steps of filtering to avoid high false-positives. In the case of snoRNAs, we used a novel tool (SnoReport 2), which seemed to explain differences in discovery. Finally, we had applied the most updated pipeline for lncRNA prediction, which brought us a close number against other species except *Clonorchis sinensis*. Although *C. canephora* genome size is similar to the size of *Proteus vulgaris* and *Berberis vulgaris* genomes, evolutionary differences and applied methodologies should be taken into account before any comparison.

Table 1. Comparative analysis of *C. canephora* genome features.

NcRNA Class (Genome Size)	<i>C. canephora</i> (~569 Mb)	<i>A. thaliana</i> (~135 Mb)	<i>C. sinensis</i> (~3.2 Gb)	<i>B. vulgaris</i> (~567 Mb)	<i>P. vulgaris</i> (~550 Mb)
tRNA	602	689	700	1043	712
rRNA	72	15	2860	231	145
miRNA	88	325	233	258	309
snRNA	115	82	223	121	165
snoRNA	1031	287	454	218	356
lncRNA	5064	3559	42,951	2546	1033
Total	6976	4957	47,421	4417	2720

2.2. Small Nuclear RNAs

snRNAs bind to specific proteins and compose the spliceosome, a large ribonucleoprotein complex responsible for the alternative splicing process [26], which is crucial since it increases diversity. In plants, the extension of the alternative splicing process can reach over 60% of intron-containing genes [27].

We identified 93 high-confidence snRNA regions, ranging from 90 to 199 nucleotides and belonging to eight spliceosomal U snRNA families (Table S8, Supplementary File S2). There are nine spliceosomal U snRNAs: (U1, U2, U4, U4 atac, U5, U6, U6 atac, U11, and U12) [26]. The U snRNAs loci identified in the *Coffea canephora* genome were from families U1 (21), U2 (16), U4 (6), U5 (9), U6 (36), U6atac (2), U11 (2), and U12 (1). U4 atac snRNAs were not identified in *C. canephora*. These are not usually found in plants, and they have been only identified in *Arabidopsis thaliana* and *Beta vulgaris* [28].

C. canephora snRNA genes were distributed mostly on chromosome 8 with 12 snRNAs, followed by chromosome 2 with 11 loci and chromosome 1 with 10 loci (Table S1, Supplementary File S1). A total of three loci did not have transcriptional evidence (Table S7, Supplementary File S2).

Twenty-two high-confidence snRNA regions were identified in the pseudomolecule “Chromosome 0”. The snRNAs that were not assigned to any chromosome had sizes ranging from 92 to 203 nucleotides and belonged to five spliceosomal U snRNA families: U1 (4), U2 (7), U4 (3), U5 (2), and U6 (6) (Table S19, Supplementary File S3).

2.3. Small Nucleolar RNAs

snoRNAs accumulate in the nucleolus and play a role in site-specific RNA modifications [29]. C/D box snoRNAs have one or more C/D box motifs in their sequence and form complex ribonucleoproteins (snoRNPs) that guides 2'-O-ribose methylation of ribonucleotides like

pre-rRNAs [26]. H/ACA box snoRNAs have H/ACA box motifs in their sequence and compose small nucleolar ribonucleoproteins (snoRNPs) that guide the pseudouridylation of specific nucleotides [26].

We identified 778 high-confidence snoRNA loci, ranging from 71 to 197 nucleotides in the *C. canephora* genome (Table S9, Supplementary File S2). Most loci had C/D box motifs (662), and 116 loci had H/ACA motifs. The *C. canephora* snoRNAs were mostly on chromosome 2, with 104 loci, followed by chromosome 4 with 85 loci and chromosome 11 with 84 loci (Table S1 and Supplementary File S1).

Two hundred and fifty-three high-confidence snoRNA regions were identified in the pseudomolecule “Chromosome 0”. The snoRNAs that were not assigned to any chromosome had sizes ranging from 76 to 212 nucleotides. Two hundred and forty-five loci had C/D box motifs, and eight loci had H/ACA motifs (Table S20, Supplementary File S3).

Some snoRNAs were organized in clusters: we found 53 genes in 22 clusters (Table S2, Supplementary File S1). snoRNA clusters are encoded or positioned closely in the same chromosomal region [29]. They have been identified in other plants: 43 in *A. thaliana* and 68 in *Oryza sativa* [30,31]. There are at least 110 snoRNA conserved families in the plant kingdom [29]. In our data, we found 56 families conserved in plants (Table S3, Supplementary File S1). The most identified C/D snoRNA family was R71 with 28 loci, also identified in *A. thaliana* and *Brassica rapa* [29]. The most identified H/ACA snoRNA family was R74 with six loci (Table S3, Supplementary File S1), also identified in *Vitis vinifera*, *C. sinensis*, and other species [29]. We observed seven loci with no transcriptional evidence (Table S7, Supplementary File S2).

2.4. Transfer RNAs

Transfer RNAs (tRNAs) carry amino acids and act in protein translation machinery [32]. We identified 427 high-confidence tRNA loci, ranging from 70 to 92 nucleotides (Table S10, Supplementary File S2). Most tRNA genes were on chromosome 2 with 91 genes, followed by chromosome 7 with 55 genes (Table S1, Supplementary File S1). tRNA^{Met} was the most frequent tRNA gene, with 61 loci. The less frequent was tRNA^{His} with 15 genes. tRNA genes copy numbers could vary among species (Table S4, Supplementary File S1).

One hundred and seventy-five high-confidence tRNA regions were identified in the pseudomolecule “Chromosome 0”. The tRNAs that were not assigned to any chromosome had sizes ranging from 70 to 92 nucleotides (Table S21, Supplementary File S3).

For each of the 20 essential amino acids, there are at least five possible anti-codons/isoacceptors totalizing 61 in the genetic code [32]. In our analyses, we identified 54 anti-codons/isoacceptors. Similar analyses in eudicots and monocots have presented anti-codons/isoacceptors, ranging from 45 to 52 [32,33]. tRNA^{Leu} was the most abundant with six isoacceptors [AAG (6) CAA (16) CAG (7) TAA (5) TAG (7), GAG (1)]; all isoacceptors are organized in Table S5, Supplementary File S1. In counterpart, tRNA^{Asp}, tRNA^{Met}, tRNA^{Cys}, tRNA^{Phe}, tRNA^{Trp}, and tRNA^{Tyr} had only one isoacceptor.

Using tRNAscan-SE [34], one putative gene for tRNA^{Sec} was also identified, which is responsible for selenocysteine biosynthesis. Although higher plants do not possess selenoproteins due to the absence of the selenocysteine insertion machinery, tRNA^{Sec} genes were searched in the *C. canephora* genome using SecMarker [35], and it was concluded that this locus was a false-positive. Previous studies have already reported that tRNAscan-SE can identify false positives for tRNA^{Sec} in plants [36].

2.5. Ribosomal RNAs

Ribosomal RNAs (rRNAs) associate with a set of proteins to form a small or a large subunit of ribosomes [37]. Fifty-one high-confidence rRNA loci were identified, ranging from 82 to 1275 nucleotides, after merging results and excluding redundancies (Table S11, Supplementary File S2). A total of 14 loci did not have transcriptional evidence (Table S7, Supplementary File S2). Units 5S (34), 5.8S (8), two fragments from a large ribosomal subunit, five fragments from a small ribosomal subunit were identified, while two were indeterminate.

Twenty-one high-confidence rRNA regions were identified in the pseudomolecule “Chromosome 0”. The rRNAs that were not assigned to any chromosome had sizes ranging from 76 to 3992 nucleotides (Table S22, Supplementary File S3).

Ribosomal RNA genes were highly conserved and organized as repetitive units, supporting the different number of rRNA genes identified in chromosome 2 and chromosome 11 (Table S1, Supplementary File S1). Besides, their copy numbers might vary among species: *C. canephora* had a larger number of rRNA genes when compared to *Daucus carota* [38] and a smaller number when compared to 149 from *Phaseolus vulgaris* [39], 232 in *Beta vulgaris* [28], and 2860 from *Citrus sinensis* [40]. It is important to consider that different methods have been applied among these studies, and completeness of genomic assemblies have a crucial impact on finding those loci.

2.6. MicroRNAs Prediction

Recent studies have pointed out a large number of false-positive miRNAs identified in plants [12,15]. For example, in *Citrus sinensis*, only 98 of the 227 miRNA loci previously described have fulfilled all annotation criteria [12]. In our miRNA prediction, a list of 109 candidate regions was obtained, and a filter was applied in them to exclude overlapping cases. One hundred and eight miRNA precursors were evaluated using miRdup [41], and 216 putative mature miRNAs were predicted. Axtell and Meyers’s (2018) criteria were also applied in these candidates, and 11 precursors were selected (Table S12, Supplementary File S2).

The set of miRNAs consisted of 10 precursors in the assembled chromosomes plus 1 precursor that were not assigned to any chromosome (“Chromosome 0”), each one with a pair of mature sequences and belonging to five conserved plant miRNA families. The 22 mature miRNAs had sizes ranging between 21 and 22 nt and were identified in three independent sRNA libraries (Table S13, Supplementary File S2).

Comparing our results with previously identified *C. canephora* miRNAs [18–24], we identified seven new precursors and 18 new mature miRNAs for *C. canephora* (Tables S12 and S13, Supplementary File S2, Supplementary File S3).

2.7. MicroRNAs Curation

In order to have a highly curated and standardized miRNA atlas for *C. canephora*, we applied Axtell and Meyers (2018) criteria in 495 putative miRNA precursors and 672 putative mature miRNAs from previous studies [18–24].

In our analysis, we found 114 redundant sequences in the 495 retrieved precursors. Two hundred and eighty-two precursors had divergent mature sequences or did not have expression, 14 precursors had no hairpin or were longer than 300 nt, seven precursors presented only one mature sequence, and 186 failed in more than one criterion (Table S6, Supplementary File S1). Some studies have identified only one mature miRNA sequence, and those sequences were excluded from our analysis due to the “only one mature sequence” criterion. In Guedes et al. (2018) study, miR-398 and miR-408 were predicted with only one mature miRNA sequence but had their sequences validated by qPCR experiments, so we considered the results.

Therefore, we considered that 60 precursors and 120 mature sequences matched the curation criteria in published data (Supplementary File S4). In summary, the *C. canephora* set of high-standard miRNAs was composed of 72 precursors and 145 mature miRNAs belonging to 27 families (Tables S12 and S13, Supplementary File S2). Most of the identified miRNA families are variants of miR-156, miR-159, miR-162, miR-164, miR-166, miR-167, miR-168, miR-169, miR-171, miR-172, miR-319, miR-390, miR-393, miR-394, miR-395, miR-396, miR-397, miR-398, miR-399, miR-403, miR-408, miR-477, miR-482, miR-530, and miR-2111, all conserved in plants [15,42]. Families miR-171 and miR-399 were the most numerous, with 11 and six loci, respectively. miR-171 family is upregulated in *A. thaliana* under nitrogen deprivation and in *Sorghum bicolor* in zinc deprivation [43]. In *Coffea arabica*, miR-171 is also

upregulated under short-term nitrogen deprivation [44]. Family miR-399 is well known to be involved in phosphate homeostasis [45].

2.8. Phylogenetic Analysis of MicroRNAs

Among the 60 precursors, miRNAs from family miR-171 were the most numerous identified; therefore, we investigated their conservation with orthologs in *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Solanum lycopersicum*, *Theobroma cacao*, and *Vitis vinifera* (Figure 2).

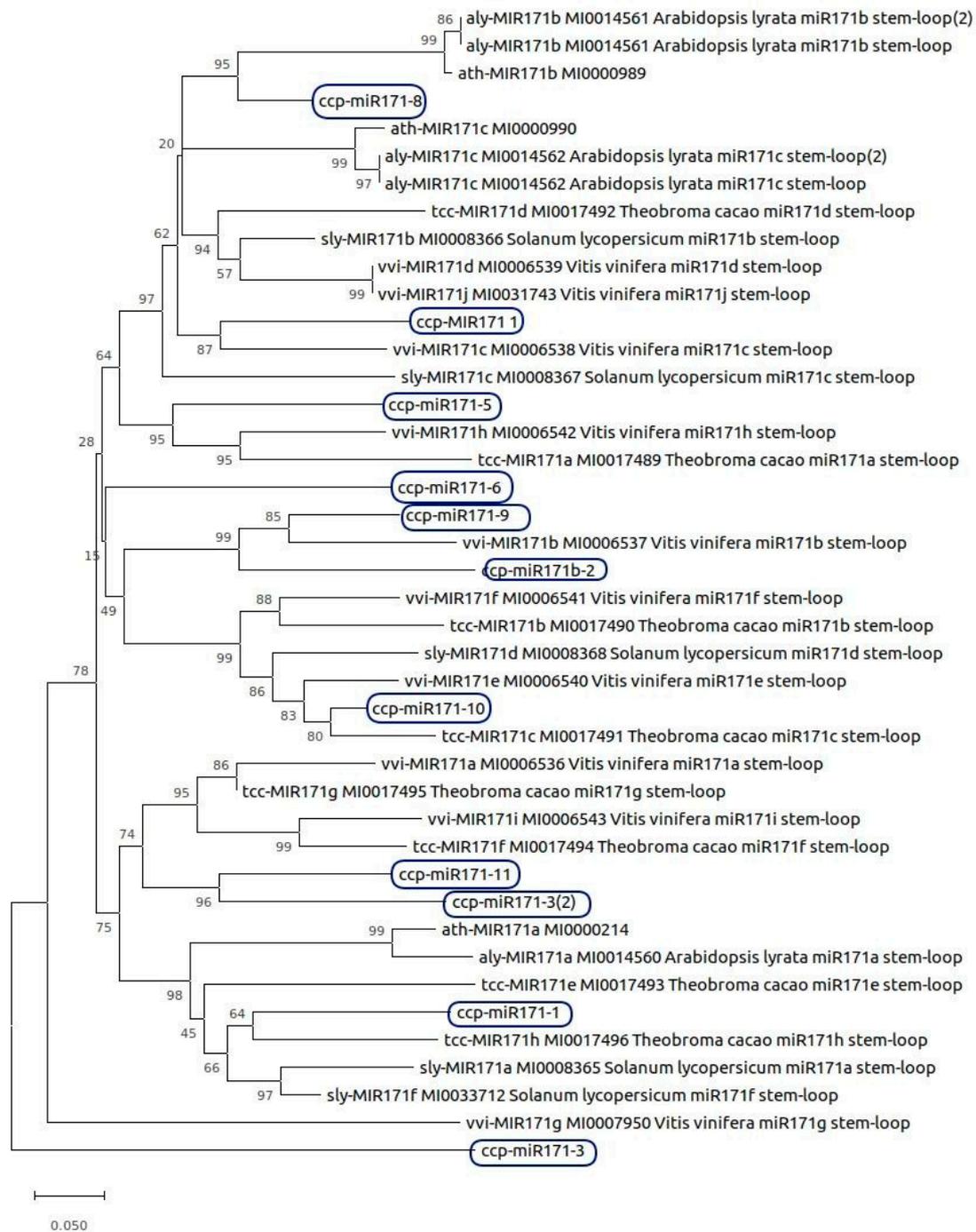


Figure 2. Phylogenetic analysis of miRNA family miR-171. The highlighted precursors showed high conservation among a total of 42 sequences, including their orthologs.

2.9. MicroRNA Targets Prediction

Understanding miRNA-made regulation is possible through the identification of their target genes [43]. We identified 10,714 potential targets for 176 miRNAs, with an average of 60.9 targets per miRNA (Table S14, Supplementary File S2). Family miR-396 had a larger number of targets (>9 targets for each family member), a fact also observed in *Manihot esculenta* [46] and *Jatropha curcas* [47].

Annotation data of potential targets were imported from Plaza 4.0 database [48]. A total of 157 high-confidence targets, based on the expectation score, were classified according to their gene ontology classes (GO terms), InterPro results, and similarity with *A. thaliana* genes (Tables S14–S16, Supplementary File S2). A total of 605 GO terms were identified, and the most overrepresented categories were DNA binding (204 hits), protein binding (159 hits), and nucleus (118 hits).

2.10. MicroRNA Expression under Drought Stress

In order to check if our in silico analysis has support in wet-lab data, we performed differential expression analysis of identified miRNAs using small RNA-Seq data from a study that evaluated the effects of multiple drought cycles on *C. canephora* miRNA expression [24]. In our analysis, miR398 (Table S13, Supplementary File S2) was up-regulated in plants submitted to one drought-stress cycle. This result was also confirmed by stem-loop qPCR analysis [24], confirming that our approach was able to validate miRNA analysis in coffee.

2.11. Long Non-Coding RNAs

We identified 2384 high-confidence lncRNA loci, ranging from 200 to 6097 nucleotides (Table S17, Supplementary File S2). A total of 232 loci did not have transcriptional evidence (Table S17, Supplementary File S2). A total of 2531 high-confidence lncRNA regions were identified in the pseudomolecule “Chromosome 0”. The lncRNAs that were not assigned to any chromosome had sizes ranging from 200 to 8536 nucleotides (Table S24, Supplementary File S3).

Long non-coding RNAs are longer than 200 nucleotides and share some characteristics with messenger RNAs but have none or little protein-coding action [49]. Long non-coding RNAs are classified according to their function and position regarding protein-coding genes [8]. When overlapping protein-coding exons in the same direction, lncRNAs are classified as sense. When overlapping protein-coding exons in the opposite direction, they are antisense lncRNAs. If the lncRNA is inside an intron of a protein-coding gene, it is classified as an intronic lncRNA, and if it is between genes, it is classified as an intergenic lncRNA (lincRNA).

In *C. canephora*, 600 loci were identified as sense lncRNAs, 92 as antisense, 235 loci were overlapping another gene region, except exons, therefore, were classified as gene overlap lncRNAs, and 4220 did not overlap with genes.

3. Discussion

The results of the homology strategies show that ncRNAs searches require multiple combinations of computational strategies to detect the diversity of structure of the RNA families. In the case of snoRNA detection, the use of a specific tool for this kind of structure allowed a six-fold increase in the number of hits.

Several curation steps also improved the identification of miRNAs in the *C. canephora* genome. Due to the high number of false-positive miRNAs identified in plants, we chose to follow a strict set of rules to select high-confidence candidates in our prediction and previous studies. We noticed that more than half of miRNAs previously predicted in *C. canephora* could be false-positives, and even among the precursors that matched curation criteria, there were redundant sequences. miR-408 prediction and validation [24] did not match all curation criteria, but it was included in the final dataset due to its experimental validation. Thus, it is important to emphasize that wet-lab approaches will still be

necessary to biologically validate miRNA families when “big data analysis” is not able to confirm their existence.

The application of clear and rigorous criteria is an important contribution to define miRNA families when data mostly relies on high throughput approaches, but it will probably underestimate miRNA families in genomes. Among the 72 miRNA precursors that matched curation criteria, 70 were variants of highly conserved plant miRNA families (detailed in Table S12, Supplementary File S2, and Supplementary File S3). The remaining two precursors belonged to family miR-157, involved in regulating vegetative phase change in *A. thaliana* [50], and family miR-3627, known for regulating plant metabolism and disease resistance [51].

At least four miRNA families (miR-160, miR-479, miR-827, and miR-1446), considered as highly conserved in plants, were previously identified in *C. canephora*, but they were excluded in our analysis. MiR-160 has been identified in two studies [22,24], but it was excluded because of the mature miRNA size and lack of expression. MiR-160 regulates the auxin response factors ARF10, -16, and -17 in plants [24,52] negatively. MiR-479 has been previously identified [23], but it was excluded because of the mature miRNA size. MiR-479 regulates plant metabolism and disease resistance, as in *V. vinifera* [51]. MiR-827 has been previously identified [22], but it was excluded in our analysis because the precursor size was greater than 300 nt and the unpaired position of mature miRNAs in the precursor. MiR-827 negatively regulates the expression of NITROGEN LIMITATION ADAPTATION (NLA), a ubiquitin E3 ligase gene (AT1G02860) in *A. thaliana* [53]. MiR-1446 has been previously identified [22], but it was excluded because of the lack of expression. MiR-1446 regulates disease resistance, as in *S. lycopersicum* [54].

In summary, we delivered here a consolidated and a highly-curated annotation for the ncRNA complement of *C. canephora* genome, unifying, revising, and expanding previous analyses.

4. Materials and Methods

4.1. Sequence Datasets

We downloaded the *C. canephora* genome in FASTA format at the Coffee Genome Hub [17] and ncRNAs sequences from Ensembl Plants version 34 [55]. Species available at Ensembl are described in Supplementary File S1. All Ensembl files were merged into one single FASTA file. Additional steps are illustrated in Figure S1, Supplementary File S1.

4.2. Similarity Search

The similarity search was done using BLASTN version 2.2.31 [56]. Ensembl Plants ncRNAs were used as a query against *C. canephora* genomic sequence with parameters word_size = 7, dust, and e-value = 10^{-5} . These search results were filtered using an in-house Python script, maintaining sequences with query coverage and identity values over 90%. The script also compared the genomic positions of each sequence, considering strands. When overlapping cases were identified, the selection was made based on the highest query coverage, highest identity, lowest e-value, and highest bitscore value. Sequences with the size smaller than 60 nt were removed, and only tRNA, rRNA, miRNA, snRNA, and snoRNA hits were retrieved for further analyses.

4.3. Structural Search

ncRNAs were identified by structural search using Infernal version 1.1.2 [57] with 2582 family models from Rfam version 12.2 [58]. The cmsearch command was used with the cut_ga parameter. Results were then filtered using an in-house Python script, which compared the genomic positions of each sequence. In overlapping cases, candidates were selected with smaller e-values and higher bit-score values. Sequences smaller than 60 nt were excluded, and the sequences assigned as tRNA, rRNA, miRNA, snRNA, and snoRNA were retrieved. Additionally, ncRNA identification was performed using tools for specific ncRNAs classes.

4.4. Small Nuclear RNAs Filtering

All sequences identified as snRNAs from the similarity and structural searches were merged. Overlapping cases were manually analyzed based on the highest query coverage, highest identity, lowest e-value, and highest bitscore value. Most snRNAs can be categorized in spliceosomal U snRNAs, small nucleolar RNAs (snoRNA), and small Cajal body-specific RNAs (scaRNA), depending on their sub-nuclear localization, function, and structure [26]. We selected sequences from U snRNA spliceosomal families with sizes ranging between 60 and 220 nucleotides.

4.5. Small Nucleolar RNAs

We used SnoReport version 2.0 (UNB, Brasília, DF, Brazil) [59] with standard parameters to predict snoRNAs. We merged SnoReport2 results with previous approaches for snoRNA prediction (similarity and structural search). Overlapping cases were manually analyzed, and candidates were selected based on higher score values and lower e-values.

We defined as high-confidence C/D box snoRNA sequences with motifs RUGAUGA near the 5'-end for C box and CUGA near the 3'-end for D box. High-confidence H/ACA box snoRNAs were also identified for sequences with the motifs ACA box next 3'-end and an H box (ANANNA) located in the hinge region based on established criteria [29,60].

As most plant snoRNAs are organized in policistrons [29], we analyzed the genomic positions of identified snoRNA genes. Genomic regions harboring snoRNA genes within an interval smaller than 500 nt were considered snoRNA clusters [61].

4.6. Transfer RNAs

We used tRNAscan-SE version 1.3.1 (The Lowe Lab, University of California, Santa Cruz, CA, USA) [34] with standard parameters to identify tRNAs. These results were merged with data from similarity and structural searches. Sequences smaller than 70 nt and with more than 95 nt were excluded using a Python script, and the overlapping cases were manually analyzed. Final tRNA loci were defined based on higher score values and lower e-values. We also searched for tRNA^{Sec} genes on SecMarker pipeline [35], but none were found.

4.7. Ribosomal RNAs

rRNAs were searched using RNAmmer version 1.2 (DTU Bioinformatics, Lyngby, Denmark) [62] with standard parameters. RNAmmer results were merged with the structure and similarity searches. The overlapping cases were manually analyzed, and the final rRNA loci were defined based on higher score values and lower e-values.

4.8. MicroRNAs

All sequences identified as miRNA precursors from the similarity and structural searches were merged into one FASTA format file. This file was used as an input file on miRdup version 1.4 [41]. We trained miRdup with high-confidence plant miRNA precursor models and high-confidence plant miRNA mature models from miRBase release 21 [63], and then the sequences were confirmed as miRNA precursors. The positive cases were used as a new input file in miRdup to predict mature miRNAs. In this output, we applied Axtell and Meyers' (2018) criteria to determine miRNA candidates. Additional steps are illustrated in Figure S2 and Supplementary File S1.

These criteria were also applied to the set of previously identified miRNAs in *C. canephora* [21–24] for curation. All sequences were compared with each other to find which ones were redundant. The miRNA family with the largest number of identified precursors in the *C. canephora* genome was selected for phylogenetic analysis.

Sequences were aligned using ClustalW in MEGA X software (Temple University, Philadelphia, PA, USA) [64] for molecular phylogenetic analyses with the following alignment parameters:

gap opening 22.50; gap extension 0.83. A phylogenetic tree was inferred using the neighbor-joining method, and the sequence divergence was estimated using the p-distance method. The statistical reliability of the internal branches was assessed using 5000 bootstrap replicates.

C. canephora miRNA target genes were predicted using the *C. canephora* genes (CDS+UTR) sequence file from Coffee Genome Hub [17]. Targets were predicted using psRNATarget (Noble Research Institute, Ardmore, OK, USA) [65] with standard parameters. Results were classified as high-confidence candidates when the expectation value was lower than 2. The results were also classified according to their similarity with *A. thaliana* genes and the presence of protein domains. For this purpose, we used annotation data from PLAZA 4.0 [48] and InterPro 70 [66].

4.9. Long Non-Coding RNAs

For lncRNAs identification, we firstly analyzed a *C. canephora* transcriptome assembly [67] and *C. canephora* genes from Coffee Genome Hub [17]. First, sequences with more than 200 nucleotides were obtained using a Perl script from the RNAplonc tool version 1.0 (UTFPR, Cornélio Procópio, PR, Brazil) (200nt.pl) [68]. The remaining sequences were filtered using CD-HIT-EST [69] with 80% similarity cutoff. We then executed CPC version 2.0 [70] with standard parameters. Sequences classified as non-coding were considered as potential lncRNAs. This set of potential lncRNAs was analyzed against Uniref90 FASTA data [71] using the BLASTX version 2.2.31 [56] with an e-value of 10^{-5} and seg filter for low complexity regions. All 'no-hit' sequences were then considered lncRNAs. These lncRNA candidates had their coordinates determined in *C. canephora* genome using BLASTn version 2.2.31 (NCBI, Bethesda, MD, USA), with at least 95% of query coverage and identity, and applying in-house Python filters to select results and exclude redundancies.

Finally, we identified the coordinate intersection with other identified ncRNAs with BEDTools version 2.25.0 (University of Utah, Salt Lake City, UT, USA) [72], and the correspondent sequences were excluded. Additional steps are illustrated in Figure S3, Supplementary File S1. To classify lncRNA candidates, intersections were searched against the GFF3 data of *C. canephora* genes from Coffee Genome Hub [17] with BEDTools version 2.25.0 (University of Utah, Salt Lake City, UT, USA) [72]. The candidates were classified according to their distance from protein-coding genes: sense, for candidates that overlap exons on the same direction; antisense, for candidates that overlap exons on the opposite direction; gene overlap, for candidates overlapping with any another gene region except exons; no overlap, for candidates that do not overlap with genes.

4.10. Transcriptional Evidence Analysis

In order to have higher confidence in ncRNA loci, the transcriptional activity of ncRNAs was inferred by mapping small RNA-Seq data from three independent studies. We used data from NCBI accession GSE46617 [23], NCBI SRA accession PRJNA353111 [24], and we also performed a small RNA-seq run from *C. canephora* var C33 mature leaves (deposited in European Nucleotide Archive under accession PRJEB29660). The sRNA library was built and sequenced at the MGX platform (Institut de Génomique Fonctionnelle, Montpellier, France). The RNA library was constructed using the 'RNA-seq sample prep' kit (Illumina, San Diego, CA, USA). RNA was sequenced on a HiSeq 2500 (Illumina), in single-end, 50 bp sequencing. Raw data were analyzed using FastQC v0.11.7 (Babraham Bioinformatics, Cambridge, United Kingdom) [73], searching for low-quality reads and adapter sequences. Reads were then filtered using Trimmomatic version 0.36 (USADDELLAB, Aachen, Germany) [74], removing reads smaller than 15 nt and a quality score below 30, along with adapters removal.

Filtered data was then mapped to the public sequence of the *C. canephora* genome [18], available at the Coffee Genome Hub [17] using STAR version 2.6.0 (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA) [75] with standard parameters. ncRNA loci with mapped reads from at least two independent small RNA-seq studies were considered as expressed ncRNAs, and regions with >10 mapped reads were considered as high-confidence ncRNA loci.

For miRNAs, we also performed differential expression analysis using mapped data [24] since it is the sole sRNA study that is a classic “treatment x control” experiment. A BED file with the microRNAs candidates coordinates was intersected against BAM files of mapping data from plants submitted to one (C1) or three cycles (C3) of drought stress to retrieve expression data. Analyses were carried out with BEDTools version 2.25.0 (University of Utah, Salt Lake City, UT, USA) [72] command intersect with -s parameter. Count files were used as input files in IDEAMEX [76] to perform differential expression analyses.

The results were the intersection of results from four validated Bioconductor packages—NOISeq, limma-voom, DESeq2, and edgeR [77]. MicroRNA candidates from intersected results file with FDR adjusted-*p*-value < 0.05 were considered as differentially expressed genes (DEG).

5. Conclusions

Although many studies have pointed out plant ncRNAs playing key roles in developmental [47,78] and regulatory processes [79,80], it is still uncommon to find studies identifying more than one or two ncRNA classes. Nevertheless, our analysis showed that, with several curation steps, it is possible to better assign most of the expected “housekeeping ncRNAs” and predict regulatory ncRNAs with high-confidence. Besides, merging predicted miRNAs from this study with results from previous studies, we obtained the first highly curated miRNA dataset for *C. canephora*.

Applying rigorous criteria based on the most recent plant miRNA annotation recommendation [15], we concluded that over 70% of *C. canephora* predicted precursor miRNAs were possibly false-positives. This was the most extensive genomic catalog of curated ncRNAs in the *Coffea* genus. The annotation of the *Coffea canephora* non-coding RNAs provided initial steps for a better understanding of the small RNA system in plants. Furthermore, it provided valuable research to establish curated non-coding RNA annotations for other plant genomes.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2311-553X/6/3/39/s1>, Supplementary File S1: supplementary methods; Figures S1–S3; Tables S1–S6. Supplementary File S2: Tables S7–S17. Supplementary File S3: Tables S18–S24. Supplementary File S4: *C. canephora* microRNAs data. File Ccanephora_ncRNA.gff3 with *C. canephora* ncRNA coordinates. Part of these materials is available online at <https://doi.org/10.7910/DVN/IUVZFFV>.

Author Contributions: Conceptualization, S.M.C.L., R.G., A.R.P. and D.S.D.; Data curation, S.M.C.L., L.F.C.F., R.G., A.R.P. and D.S.D.; Formal analysis, S.M.C.L., L.F.C.F., A.R.P. and D.S.D.; Funding acquisition, R.G., A.R.P. and D.S.D.; Investigation, S.M.C.L., R.G., A.R.P. and D.S.D.; Methodology, S.M.C.L., L.F.C.F., A.R.P. and D.S.D.; Project administration, R.G., A.R.P. and D.S.D.; Resources, R.G. and A.R.P.; Software, A.R.P.; Supervision, R.G., A.R.P. and D.S.D.; Validation, L.F.C.F., A.R.P. and D.S.D.; Visualization, L.F.C.F.; Writing—original draft, S.M.C.L., R.G., A.R.P. and D.S.D.; Writing—review & editing, S.M.C.L., A.R.P. and D.S.D. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001.

Acknowledgments: We thank the valuable comments of Luiz Filipe Protasio Pereira and Alan Péricles Rodrigues Lorenzetti on the initial versions of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ariel, F.; Romero-Barrios, N.; Jégu, T.; Benhamed, M.; Crespi, M. Battles and hijacks: Noncoding transcription in plants. *Trends Plant Sci.* **2015**, *20*, 362–371. [[CrossRef](#)] [[PubMed](#)]
2. Tang, J.; Chu, C. MicroRNAs in crop improvement: Fine-tuners for complex traits. *Nat. Plants* **2017**, *3*, 17077. [[CrossRef](#)] [[PubMed](#)]
3. Liu, D.; Mewalal, R.; Hu, R.; Tuskan, G.A.; Yang, X. New technologies accelerate the exploration of non-coding RNAs in horticultural plants. *Hortic. Res.* **2017**, *4*, 17031. [[CrossRef](#)]
4. D’Ario, M.; Griffiths-Jones, S.; Kim, M. Small RNAs: Big Impact on Plant Development. *Trends Plant Sci.* **2017**, *22*, 1056–1068. [[CrossRef](#)]

5. Jones-Rhoades, M.W.; Bartel, D.P. Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. *Mol. Cell* **2004**, *14*, 787–799. [[CrossRef](#)] [[PubMed](#)]
6. Chekanova, J.A. Long non-coding RNAs and their functions in plants. *Curr. Opin. Plant Biol.* **2015**, *27*, 207–216. [[CrossRef](#)]
7. Diler, E.; Unver, T.; Karakülah, G. Differential Expression of Hyperhydricity Responsive Peach MicroRNAs. *J. Integr. Bioinform.* **2016**, *13*, 57–69. [[CrossRef](#)]
8. Rai, M.I.; Alam, M.; Lightfoot, D.; Gurha, P.; Afzal, A.J. Classification and experimental identification of plant long non-coding RNAs. *Genomics* **2019**, *111*, 997–1005. [[CrossRef](#)]
9. Zhang, B.; Pan, X.; Cannon, C.H.; Cobb, G.P.; Anderson, T.A. Conservation and divergence of plant microRNA genes. *Plant J.* **2006**, *46*, 243–259. [[CrossRef](#)]
10. Nithin, C.; Thomas, A.; Basak, J.; Bahadur, R.P. Genome-wide identification of miRNAs and lncRNAs in *Cajanus cajan*. *BMC Genom.* **2017**, *18*, 878. [[CrossRef](#)]
11. Lin, Z.; Li, Q.; Yin, Q.; Wang, J.; Zhang, B.; Gan, S.; Wu, A.M. Identification of novel miRNAs and their target genes in *Eucalyptus grandis*. *Tree Genet. Genomes* **2018**, *14*, 60. [[CrossRef](#)]
12. Taylor, R.; Tarver, J.E.; Foroozani, A.; Donoghue, P.C.J. MicroRNA annotation of plant genomes—Do it right or not at all. *BioEssays* **2017**, *39*, 1600113. [[CrossRef](#)] [[PubMed](#)]
13. Ambros, V.; Bartel, B.; Bartel, D.P.; Burge, C.B.; Carrington, J.C.; Chen, X.; Dreyfuss, G.; Eddy, S.R.; Griffiths-Jones, S.; Marshall, M.; et al. A uniform system for microRNA annotation. *RNA* **2003**, *9*, 277–279. [[CrossRef](#)] [[PubMed](#)]
14. Meyers, B.C.; Axtell, M.J.; Bartel, B.; Bartel, D.P.; Baulcombe, D.C.; Bowman, J.L.; Cao, X.; Carrington, J.C.; Chen, X.; Green, P.J.; et al. Criteria for Annotation of Plant MicroRNAs. *Plant Cell* **2008**, *20*, 3186–3190. [[CrossRef](#)]
15. Axtell, M.J.; Meyers, B.C. Revisiting Criteria for Plant MicroRNA Annotation in the Era of Big Data. *Plant Cell* **2018**, *30*, 272–284. [[CrossRef](#)]
16. ICO. International Coffee Organization Statistics. Available online: <http://www.ico.org> (accessed on 20 July 2019).
17. Dereeper, A.; Bocs, S.; Rouard, M.; Guignon, V.; Ravel, S.; Tranchant-Dubreuil, C.; Poncet, V.; Garsmeur, O.; Lashermes, P.; Droc, G. The coffee genome hub: A resource for coffee genomes. *Nucleic Acids Res.* **2015**, *43*, D1028–D1035. [[CrossRef](#)]
18. Denoed, F.; Carretero-Paulet, L.; Dereeper, A.; Droc, G.; Guyot, R.; Pietrella, M.; Zheng, C.; Alberti, A.; Anthony, F.; Aprea, G.; et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **2014**, *345*, 1181–1184. [[CrossRef](#)]
19. Bibi, F.; Barozai, M.Y.K.; Din, M. Bioinformatics profiling and characterization of potential microRNAs and their targets in the genus *Coffea*. *Turk. J. Agric. For.* **2017**, *41*, 191–200. [[CrossRef](#)]
20. Nellikunnumal, S.M.; Chandrashekar, A. Computational Identification of Conserved microRNA and their Targets in *Coffea canephora* by EST Analysis. *Dyn. Biochem. Process Biotechnol. Mol. Biol.* **2012**, *6*, 70–76.
21. Chaves, S.S.; Fernandes-Brum, C.N.; Silva, G.F.F.; Ferrara-Barbosa, B.C.; Paiva, L.V.; Nogueira, F.T.S.; Cardoso, T.C.S.; Amaral, L.R.; Gomes, M.D.S.; Chalfun-Junior, A. New Insights on *Coffea* miRNAs: Features and Evolutionary Conservation. *Appl. Biochem. Biotechnol.* **2015**, *177*, 879–908. [[CrossRef](#)]
22. Fernandes-Brum, C.N.; Rezende, P.M.; Ribeiro, T.H.C.; De Oliveira, R.R.; Cardoso, T.C.D.S.; Amaral, L.R.D.; Gomes, M.D.S.; Chalfun-Junior, A. A genome-wide analysis of the RNA-guided silencing pathway in coffee reveals insights into its regulatory mechanisms. *PLoS ONE* **2017**, *12*, e0176333. [[CrossRef](#)] [[PubMed](#)]
23. Loss-Morais, G.; Ferreira, D.C.; Margis, R.; Alves-Ferreira, M.; Correa, R.L. Identification of novel and conserved microRNAs in *Coffea canephora* and *Coffea arabica*. *Genet. Mol. Boil.* **2014**, *37*, 671–682. [[CrossRef](#)]
24. Guedes, F.; Nobres, P.; Ferreira, D.C.R.; Menezes-Silva, P.E.; Ribeiro-Alves, M.; Correa, R.L.; DaMatta, F.M.; Alves-Ferreira, M. Transcriptional memory contributes to drought tolerance in coffee (*Coffea canephora*) plants. *Environ. Exp. Bot.* **2018**, *147*, 220–233. [[CrossRef](#)]
25. Oróstica, K.Y.; Verdugo, R.A. chromPlot: Visualization of genomic data in chromosomal context. *Bioinformatics* **2016**, *32*, 2366–2368. [[CrossRef](#)] [[PubMed](#)]
26. Ohtani, M. Transcriptional regulation of snRNAs and its significance for plant development. *J. Plant Res.* **2016**, *130*, 57–66. [[CrossRef](#)] [[PubMed](#)]
27. Reddy, A.S.; Marquez, Y.; Kalyna, M.; Barta, A. Complexity of the alternative splicing landscape in plants. *Plant Cell* **2013**, *25*, 3657–3683. [[CrossRef](#)]

28. Dohm, J.C.; Minoche, A.E.; Holtgräwe, D.; Capella-Gutierrez, S.; Zakrzewski, F.; Tafer, H.; Rupp, O.; Sörensen, T.R.; Stracke, R.; Reinhardt, R.; et al. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* **2013**, *505*, 546–549. [[CrossRef](#)]
29. Bhattacharya, D.P.; Canzler, S.; Kehr, S.; Hertel, J.; Grosse, I.; Stadler, P.F. Phylogenetic distribution of plant snoRNA families. *BMC Genom.* **2016**, *17*, 969. [[CrossRef](#)]
30. Brown, J.W.; Clark, G.P.; Leader, D.J.; Simpson, C.G.; Lowe, T.M. Multiple snoRNA gene clusters from Arabidopsis. *RNA* **2001**, *7*, 1817–1832. [[CrossRef](#)]
31. Chen, C.L.; Liang, D.; Zhou, H.; Zhuo, M.; Chen, Y.; Qu, L.H. The high diversity of snoRNAs in plants: Identification and comparative study of 120 snoRNA genes from *Oryza sativa*. *Nucleic Acids Res.* **2003**, *31*, 2601–2613. [[CrossRef](#)]
32. Mohanta, T.K.; Bae, H. Analyses of Genomic tRNA Reveal Presence of Novel tRNAs in *Oryza sativa*. *Front. Genet.* **2017**, *8*, 90. [[CrossRef](#)]
33. Michaud, M.; Cognat, V.; Duchêne, A.-M.; Maréchal-Drouard, L. A global picture of tRNA genes in plant genomes. *Plant J.* **2011**, *66*, 80–93. [[CrossRef](#)] [[PubMed](#)]
34. Lowe, T.M.; Chan, P.P. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **2016**, *44*, W54–W57. [[CrossRef](#)]
35. Santessmasses, D.; Mariotti, M.; Guigo, R. Computational identification of the selenocysteine tRNA (tRNA^{Sec}) in genomes. *PLoS Comput. Boil.* **2017**, *13*, e1005383. [[CrossRef](#)] [[PubMed](#)]
36. Lobanov, A.V.; Hatfield, D.L.; Gladyshev, V.N. Eukaryotic selenoproteins and selenoproteomes. *Biochim. Biophys. Acta (BBA)* **2009**, *1790*, 1424–1428. [[CrossRef](#)]
37. Cech, T.R.; Steitz, J.A. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* **2014**, *157*, 77–94. [[CrossRef](#)] [[PubMed](#)]
38. Iorizzo, M.; Ellison, S.; Senalik, D.; Zeng, P.; Satapoomin, P.; Huang, J.; Bowman, M.J.; Iovene, M.; Sanseverino, W.; Cavagnaro, P.; et al. A high-quality carrot genome assembly provides new insights into carotenoid accumulation and asterid genome evolution. *Nat. Genet.* **2016**, *48*, 657–666. [[CrossRef](#)] [[PubMed](#)]
39. Vlasova, A.; Capella-Gutierrez, S.; Rendón-Anaya, M.; Hernández-Oñate, M.A.; Minoche, A.E.; Erb, I.; Câmara, F.; Barja, P.P.; Corvelo, A.; Sanseverino, W.; et al. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genome Boil.* **2016**, *17*, 32. [[CrossRef](#)] [[PubMed](#)]
40. Xia, E.H.; Zhang, H.B.; Sheng, J.; Li, K.; Zhang, Q.J.; Kim, C.; Zhang, Y.; Liu, Y.; Zhu, T.; Li, W.; et al. The Tea Tree Genome Provides Insights into Tea Flavor and Independent Evolution of Caffeine Biosynthesis. *Mol. Plant* **2017**, *10*, 866–877. [[CrossRef](#)]
41. Leclercq, M.; Diallo, A.B.; Blanchette, M. Computational prediction of the localization of microRNAs within their pre-miRNA. *Nucleic Acids Res.* **2013**, *41*, 7200–7211. [[CrossRef](#)]
42. Cuperus, J.T.; Fahlgren, N.; Carrington, J.C. Evolution and Functional Diversification of miRNA Genes. *Plant Cell* **2011**, *23*, 431–442. [[CrossRef](#)] [[PubMed](#)]
43. Paul, S.; Datta, S.K.; Datta, K. miRNA regulation of nutrient homeostasis in plants. *Front. Plant Sci.* **2015**, *6*, 232. [[CrossRef](#)] [[PubMed](#)]
44. Santos, T.B.; Soares, J.D.M.; Lima, J.E.; Silva, J.C.; Ivamoto-Suzuki, S.T.; Baba, V.Y.; Souza, S.G.H.; Lorenzetti, A.P.R.; Paschoal, A.R.; Meda, A.R.; et al. An integrated analysis of mRNA and sRNA transcriptional profiles in *Coffea arabica* L. roots: Insights on nitrogen starvation responses. *Funct. Integr. Genom.* **2018**, *19*, 151–169. [[CrossRef](#)]
45. Pant, B.D.; Buhtz, A.; Kehr, J.; Scheible, W. MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. *Plant J.* **2008**, *53*, 731–738. [[CrossRef](#)] [[PubMed](#)]
46. Rogans, S.J.; Rey, C. Unveiling the Micronome of Cassava (*Manihot esculenta* Crantz). *PLoS ONE* **2016**, *11*, e0147251. [[CrossRef](#)]
47. Yang, T.; Wang, Y.; Teotia, S.; Zhang, Z.; Tang, G. The Making of Leaves: How Small RNA Networks Modulate Leaf Development. *Front. Plant Sci.* **2018**, *9*, 824. [[CrossRef](#)]
48. Bel, M.V.; Diels, T.; Vancaester, E.; Kreft, L.; Botzki, A.; Van De Peer, Y.; Coppens, F.; Vandepoele, K. PLAZA 4.0: An integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* **2018**, *46*, D1190–D1196. [[CrossRef](#)]
49. Nejat, N.; Mantri, N. Emerging roles of long non-coding RNAs in plant response to biotic and abiotic stresses. *Crit. Rev. Biotechnol.* **2018**, *38*, 93–105. [[CrossRef](#)]

50. He, J.; Xu, M.; Willmann, M.R.; McCormick, K.; Hu, T.; Yang, L.; Starker, C.G.; Voytas, D.F.; Meyers, B.C.; Poethig, R.S. Threshold-dependent repression of SPL gene expression by miR156/miR157 controls vegetative phase change in *Arabidopsis thaliana*. *PLoS Genet.* **2018**, *14*, e1007337. [[CrossRef](#)]
51. Xue, M.; Yi, H.; Wang, H. Identification of miRNAs involved in SO₂ preservation in *Vitis vinifera* L. by deep sequencing. *Environ. Exp. Bot.* **2018**, *153*, 218–228. [[CrossRef](#)]
52. Lin, Y.; Lai, Z.; Tian, Q.; Lin, L.; Lai, R.; Yang, M.; Zhang, D.; Chen, Y.; Zhang, Z. Endogenous target mimics down-regulate miR160 mediation of ARF10, -16, and -17 cleavage during somatic embryogenesis in *Dimocarpus longan* Lour. *Front. Plant Sci.* **2015**, *6*, 956. [[CrossRef](#)] [[PubMed](#)]
53. Hewezi, T.; Piya, S.; Qi, M.; Balasubramaniam, M.; Rice, J.H.; Baum, T.J. Arabidopsis miR827 mediates post-transcriptional gene silencing of its ubiquitin E3 ligase target gene in the syncytium of the cyst nematode *Heterodera schachtii* to enhance susceptibility. *Plant J.* **2016**, *88*, 179–192. [[CrossRef](#)] [[PubMed](#)]
54. Kaur, P.; Shukla, N.; Joshi, G.; Vijayakumar, C.; Jagannath, A.; Agarwal, M.; Goel, S.; Kumar, A. Genome-wide identification and characterization of miRNAome from tomato (*Solanum lycopersicum*) roots and root-knot nematode (*Meloidogyne incognita*) during susceptible interaction. *PLoS ONE* **2017**, *12*, e0175178. [[CrossRef](#)]
55. Yates, A.; Achuthan, P.; Akanni, W.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; Bhai, J.; et al. Ensembl 2020. *Nucleic Acids Res.* **2020**, *48*, D682–D688. [[CrossRef](#)]
56. Boratyn, G.M.; Camacho, C.; Cooper, P.S.; Coulouris, G.; Fong, A.; Ma, N.; Madden, T.L.; Matten, W.T.; McGinnis, S.D.; Merezhuk, Y.; et al. BLAST: A more efficient report with usability improvements. *Nucleic Acids Res.* **2013**, *41*, W29–W33. [[CrossRef](#)] [[PubMed](#)]
57. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **2013**, *29*, 2933–2935. [[CrossRef](#)]
58. Kalvari, I.; Nawrocki, E.P.; Argasinska, J.; Quinones-Olvera, N.; Finn, R.D.; Bateman, A.; Petrov, A.I. Non-Coding RNA Analysis Using the Rfam Database. *Curr. Protoc. Bioinform.* **2018**, *62*, e51. [[CrossRef](#)]
59. Oliveira, J.V.A.; Costa, F.; Backofen, R.; Stadler, P.F.; Walter, M.E.M.T.; Hertel, J. SnoReport 2.0: New features and a refined Support Vector Machine to improve snoRNA identification. *BMC Bioinform.* **2018**, *17* (Suppl. 18), 464–486. [[CrossRef](#)]
60. Kalinina, N.O.; Makarova, S.; Makhotenko, A.; Love, A.J.; Taliany, M.E. The Multiple Functions of the Nucleolus in Plant Development, Disease and Stress Responses. *Front. Plant Sci.* **2018**, *9*, 132. [[CrossRef](#)]
61. Liu, T.T.; Zhu, D.; Chen, W.; Deng, W.; He, H.; He, G.; Bai, B.; Qi, Y.; Chen, R.; Deng, X.W. A Global Identification and Analysis of Small Nucleolar RNAs and Possible Intermediate-Sized Non-Coding RNAs in *Oryza sativa*. *Mol. Plant* **2013**, *6*, 830–846. [[CrossRef](#)]
62. Lagesen, K.; Hallin, P.; Rødland, E.A.; Stærfeldt, H.H.; Rognes, T.; Ussery, D.W.; Staerfeldt, H.H. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **2007**, *35*, 3100–3108. [[CrossRef](#)] [[PubMed](#)]
63. Kozomara, A.; Birgaoanu, M.; Griffiths-Jones, S. miRBase: From microRNA sequences to function. *Nucleic Acids Res.* **2019**, *47*, D155–D162. [[CrossRef](#)] [[PubMed](#)]
64. Kumar, S.; Stecher, G.; Li, M.; Nnyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Boil. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)] [[PubMed](#)]
65. Dai, X.; Zhuang, Z.; Zhao, P.X. psRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **2018**, *46*, W49–W54. [[CrossRef](#)]
66. Mitchell, A.; Attwood, T.K.; Babbitt, P.C.; Blum, M.; Bork, P.; Bridge, A.; Brown, S.D.; Chang, H.Y.; El-Gebali, S.; Fraser, M.I.; et al. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **2018**, *47*, D351–D360. [[CrossRef](#)]
67. Mondego, J.M.C.; Vidal, R.O.; Carazzolle, M.F.; Tokuda, E.K.; Parizzi, L.P.; Costa, G.G.L.; Pereira, L.F.P.; Andrade, A.C.; Colombo, C.A.; Vieira, L.G.E.; et al. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. *BMC Plant Boil.* **2011**, *11*, 30. [[CrossRef](#)]
68. Negri, T.D.C.; Alves, W.A.L.; Bugatti, P.H.; Saito, P.T.; Domingues, D.S.; Paschoal, A.R. Pattern recognition analysis on long noncoding RNAs: A tool for prediction in plants. *Brief. Bioinform.* **2019**, *20*, 682–689. [[CrossRef](#)]
69. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [[CrossRef](#)]

70. Kang, Y.J.; Yang, D.C.; Kong, L.; Hou, M.; Meng, Y.Q.; Wei, L.; Gao, G. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **2017**, *45*, W12–W16. [[CrossRef](#)]
71. Suzek, B.E.; Wang, Y.; Huang, H.; McGarvey, P.B.; Wu, C.H.; UniProt Consortium. The UniProt Consortium UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **2015**, *31*, 926–932. [[CrossRef](#)]
72. Quinlan, A.R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinform.* **2014**, *47*, 11.12.1–11.12.34. [[CrossRef](#)] [[PubMed](#)]
73. Wingett, S.W.; Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **2018**, *7*, 1338. [[CrossRef](#)] [[PubMed](#)]
74. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
75. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)]
76. Jiménez-Jacinto, V.; Sanchez-Flores, A.; Vega-Alvarado, L. Integrative Differential Expression Analysis for Multiple EXperiments (IDEAMEX): A Web Server Tool for Integrated RNA-Seq Data Analysis. *Front. Genet.* **2019**, *10*, 279. [[CrossRef](#)]
77. Costa-Silva, J.; Domingues, D.S.; Lopes, F.M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* **2017**, *12*, e0190152. [[CrossRef](#)]
78. Mishra, A.; Bohra, A. Non-coding RNAs and plant male sterility: Current knowledge and future prospects. *Plant Cell Rep.* **2018**, *37*, 177–191. [[CrossRef](#)]
79. Shin, S.Y.; Shin, C. Regulatory non-coding RNAs in plants: Potential gene resources for the improvement of agricultural traits. *Plant Biotechnol. Rep.* **2016**, *10*, 35–47. [[CrossRef](#)]
80. Kim, N.H.; Xi, Y.; Sung, S. Modular function of long noncoding RNA, COLDAIR, in the vernalization response. *PLoS Genet.* **2017**, *13*, e1006939. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).