TITLE

Coffee microsporogenesis and related small interfering RNAs biosynthesis have a unique pattern among eudicots suggesting a sensitivity to climate changes.

AUTHORS

Kellen Kauanne Pimenta de Oliveira[1], Raphael Ricon de Oliveira[1], Gabriel de Campos Rume[1], Thales Henrique Cherubino Ribeiro[1], Christiane Noronha Fernandes-Brum[1], Atul Kakrana[2], Sandra Mathioni[2], Blake C. Meyers[2,3], Matheus de Souza Gomes, Antonio Chalfun-Junior[1*].

* CORRESPONDING AUTHORS

AFFILIATIONS

[1]Laboratory of Plant Molecular Physiology, Plant Physiology Sector, Institute of Biology, Federal University of Lavras, Lavras-MG, Brazil.

[2]Donald Danforth Plant Science Center, St. Louis, MO, 63132, USA.

[3]Division of Plant Science and Technology, University of Missouri-Columbia, Columbia, Missouri, USA.

[4]Institute of Genetics and Biochemistry (INGEB), Laboratory of Bioinformatics and Molecular Analysis (LBAM), Federal University of Uberlândia (UFU)-Campus Patos de Minas, Patos de Minas, Minas Gerais, Brasil.

ABSTRACT

Recently, the siRNAs pathways, and especially reproductive phasiRNAs, have attracted attention in eudicots since their biological roles are still unknown and their biogenesis took different evolutionary pathways compared to monocots. In this work, we used *Coffea arabica* L., a recently allotetraploid formed from the hybridization of *C. canephora* and *C. eugenioides* unreduced gametes, to explore microsporogenesis and small RNAs related pathways in a eudicot crop. First, we identified the microsporogenesis stages during anther development revealing

that pre-meiosis occurs in anthers of 1.5 mm inside floral buds (FBs), whereas meiosis between 1.5 and 4.2 mm FBs, and post-meiosis in FBs larger than 4.2 mm. These stages coincide with the Brazilian winter, a period of FBs reduced growth which suggests temperature sensitivity. Next, we identified and quantified the expression of reproductive 21- and 24-nt phasiRNAs during coffee anther development together with their canonical and novel miRNA triggers, and characterized the DCL and AGO families. Our results showed that the pattern of reproductive phasiRNA abundance in *C. arabica* is unique among described eudicots and the canonical trigger car-miR2275 is involved in the processing of both 21 and 24 nt phasiRNAs. Fourteen DCL genes were identified, but DCL5, related to phasiRNA biosynthesis in monocots, was not according to its specificity for monocots. Thus, our work explored the knowledge gap about microsporogenesis and related siRNAs pathways in coffee, contributing to the control of reproductive development and to the improvement of fertility in eudicots.

KEYWORDS

Coffee microsporogenesis; Small interfering RNAs (siRNAs); Phased small interfering RNAs (phasiRNAs); MicroRNAs (miRNAs); DICER-like (DCL) proteins; Reproductive development; Eudicot crops.

1 INTRODUCTION

Microsporogenesis is a fundamental process to produce pollen and guarantee fertilization of plants with sexual reproduction (Pradillo & Santos, 2018; Adhikari et al., 2020). This process is evolutionarily diverse in plants and largely governed by molecular pathways that include small interfering RNAs (siRNAs) and enzymes related to its biosynthesis (Ma et al., 2005; Albert et al., 2014; Åstrand et al., 2021). Recently, some studies have shed light on the function and biosynthesis of siRNA molecules, as phased secondary small interfering RNAs (phasiRNAs) and microRNAs (miRNAs) in monocot species (Nadot et al., 2008; Lee & Carroll, 2018; Dhaka et al., 2020). However, this knowledge is still poorly comprehended in eudicot species, especially crops with complex genomes and sexual reproduction cycles. Here, we explored this gap using *Coffea arabica* L. (Rubiaceae), a recent allotetraploid formed from the hybridization of *C.*

canephora and *C. eugenioides* unreduced gametes (Lashermes et al., 1999; Scalabrin et al., 2020).

The phasiRNAs are small RNAs found in plants, generated from a long RNA precursor at intervals of 21 to 26 nucleotides (nt), hypothesized to be a relevant group for reproductive regulation (Xia et al., 2019; Zhai et al., 2015), these phasiRNAs are produced from non-coding precursors ("PHAS" transcripts) and transcribed from non-repetitive loci (*PHAS* loci) (Axtell, 2013a; Komiya, 2017). For example, reproductive phasiRNAs have a spatiotemporal pattern in monocots in which hundreds of *PHAS* loci on all chromosomes produce 21-nt phasiRNAs abundant in pre-meiotic anthers and 24 nt-phasiRNAs enriched in meiotic stage anthers (Fan et al., 2016; Kakrana et al., 2018; Zhai et al., 2015). Disturbance of 21-*PHAS* loci underlies photoperiod-sensitive male sterility in rice, while mutants that do not produce 24-nt phasiRNA produce conditional male sterility in maize (Zhai et al., 2015). Moreover, reproductive phasiRNAs are also present in female reproductive organs and therefore may function in both male and female germinal development (Kakrana et al., 2018).

The phasiRNAs can be classified into two main groups, based on their origin: from *PHAS* loci located in noncoding or protein-coding regions (Fei et al., 2013; Liu et al., 2020). Systematic analysis of the main phylogenetic plant groups, encompassing algae, mosses, gymnosperms, basal angiosperms, monocots, and eudicots, identify abundant *PHAS* loci among protein-coding genes, suggesting that phasiRNAs predominantly regulate these genes from which they are originate, in contrast to the dominant trans-regulatory mode of microRNAs (miRNAs) (Liu et al., 2020)

In monocots, two pathways associated with germlines are well described and produce diverse and abundant 21-nt (premeiotic) and 24-nt (meiotic) phased siRNAs (Fei et al., 2016; Zhai et al., 2015). These phasiRNAs are produced from capped and polyadenylated non-coding precursors ("PHAS" transcripts) and transcribed by RNA polymerase II (Pol II) from non-repetitive loci. They are typically triggered by two 22-nt miRNAs, miR2118/482 for 21-nt phasiRNAs and miR2275 for 24-nt phasiRNAs. 3' mRNA fragments are converted to double-stranded RNA by RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) and

processed by DICER-like 4 (DCL4) and DICER-like 5 (DCL5) to produce 21-nt and 24-nt phasiRNAs, respectively. Then, phasiRNAs are loaded into ARGONAUTE (AGO) proteins to induce target RNA silencing (Kakrana et al., 2018; Komiya, 2017; Xia et al., 2019).

Notably, DCL5 is a Dicer protein found specifically in monocots and, so far, has not been reported in eudicots, despite the existence of 24-nt phasiRNAs and the conservation of miR2275, both broadly present in angiosperms (Xia et al., 2019). Moreover, genomic analyses showed divergence related to the miR2275/24-nt phasiRNA pathway in plants, since it is absent in legumes, *Arabidopsis,* and in *Solanaceae* species that produce 24-nt phasiRNAs but that lack the miR2275 trigger (Xia et al., 2019).

Evolutionary studies with non-grass monocots (garden asparagus, daylily, and lily) have suggested that the DCL5-dependent meiotic 24-nt phasiRNA pathway may have originated more than 117 MYA (KAKRANA et al., 2018). The DCL5 subfamily seems to have evolved from an ancient 'eudicot-type' DCL3 ancestor, able to produce 24-nt phasiRNAs, and DCL5 accumulated mutations specifying it for distinct substrates and ensuring the production of functional siRNAs (Chen et al., 2022). Thus, plants have diversified and optimized RNA silencing mechanisms to coordinate reproductive development during evolution (Chen et al., 2022). How functional differentiation of miRNA/DCL triggers and reproductive phasiRNAs was achieved, and their biological relevance, remains elusive, especially during the microsporogenesis of eudicot plants.

2 MATERIALS AND METHODS

Plant material.

Floral buds of *C. arabica* cv. Red Catuaí at different development stages (separated by length in millimeters) were collected in the experimental field of UFLA (-21.20549, -44.98021) to identify cells at pre-meiosis, meiosis, and post-meiosis. Tissues used for microscopy were fixed in Carnoy 3:1 and stored in the refrigerator at 4°C, whereas those used for RNA extractions were collected in liquid nitrogen and stored in an ultrafreezer at -80°C, until used.

Optical and Scanning Electron Microscopy (SEM)

For optical microscopy, slides were prepared according to (Gardner; Rattenbury, 1974) protocol and pictures taken with a 400x magnifying glass with more than 4X digital zoom (Fig. 1A to C). For Scanning Electron Microscopy (SEM), floral buds were removed from the primary fixative and transferred to microtubes containing 0.05 M cacodylate buffer for ten minutes. The buffer was changed three times. Samples were immersed in 1% osmium tetroxide solution in a cacodylate buffer for two hours at room temperature. Finally, they were washed with distilled water three times and dehydrated in series with acetone (25%, 50%, 75%, 90% - 10 minutes each, and 100% three times for 10 minutes). After this step, samples were completely dried utilizing a sputtering apparatus (Balzers CPD 030). After drying, samples were mounted on stubs and submitted to gold metallization in the Balzers SCD 050 vaporizer and observed in SEM (LEO EVO 40 XVP (Carl Zeiss) with Bruker X-ray microanalysis system (Quantax EDS) and cryosystem (Gatan).

Small RNA libraries construction (Illumina sequencing)

Anthers from floral buds at different developmental stages were extracted and macerated in liquid nitrogen. Total RNA was extracted from 100 mg of tissue using the PureLink® RNA Mini Kit (Invitrogen) according to the manufacturer's protocol. RNA samples were treated with DNase I using the RNase-free DNase Set (Qiagen) to eliminate residual DNA contamination. RNA integrity was visually analyzed in 1% agarose gel, and RNA content, as well as quality, were obtained by spectrophotometry (OD260/280 and OD260/230 > 1.8) (NanoVue GE Healthcare, Munich, Germany). RNA integrity number (RIN) was obtained with the Agilent 2100 Bioanalyzer system (Agilent Technologies) according to the manufacturer's recommendations for plant RNA. Only samples with RIN greater than 7 were used in the following steps. Size selection for 20-30 nucleotide (nt) small RNAs was performed in denaturing Urea-PAGE gels, and libraries were constructed using the TruSeq Small RNA Library Preparation Kit (Illumina, cat # RS-200-0024) (Mathioni et al., 2017). All libraries were single-end sequenced with 51 cycles on an Illumina HiSeq 2500 instrument.

A total of 104.303.112 small-RNA-seq single-end reads were inspected for adapters with the Minion tool (Davis et al., 2013). The identified adapter

sequences were removed with Trimmomatic v. 0.39 (Bolger et al., 2014). Only reads with more than 17 nucleotides were selected for further analysis after quality control. After sequencing and data cleanup steps, the small RNA libraries were aligned against *C. arabica* and *C. canephora* genomes databases, in which approximately 95% of the reads aligned in the first and 80% in the second. Details of the small RNA libraries of *C. arabica* anthers in pre-meiosis, meiosis and post-meiosis, including the numbers of Sequence Read Archive database (SRAs), are available in Table S1.

Computational *de novo* identification of phased, siRNA-generating loci and their miRNA triggers.

A total of 102.985.416 reads from pre-meiotic, meiotic, and post-meiotic small RNA libraries were submitted to the PHASIS suite (Kakrana et al., 2017). The first step was the de novo prediction of *PHAS* loci with the phase detect tool. Subsequently, phase merge was applied with parameter "-mode" set to "merge" to create a unified list of the PHASIS predicted in samples from the different stages. Finally, "phastrigs" were applied to identify miRNAs triggers for the previously identified *PHAS* loci; to do so we supplied to "phastrigs" a fasta file containing all the mature miRNAs identified in the "Conserved and novel miRNA prediction from *Coffea* sp" section. These steps were run twice to identify both 21 or 24-PHAS and the minimum significance p-value was set to $1e^{-6}$. Custom R-scripts were developed to plot heatmaps of the inferred abundance of each predicted PHAS.

In silico and phylogenetic analysis of coffee DCL- and AGO-like proteins

To identify AGO and DCL homologs present in the coffee genome (available at National Center for Biotechnology Information's (NCBI) webpage: https://www.ncbi.nlm.nih.gov/assembly/GCF_003713225.1/#/def) and perform phylogenetic analysis, we used the methods described in Rume et al., 2023. Briefly, sequences of AGO and DCL proteins of different species (see Tables S1) were downloaded from Genbank (Benson et al., 2009), UniProt (The Uniprot Consortium, 2021), and the MaizeGDB (Woodhouse et al., 2021). These

sequences were utilized in separate alignments against the proteome of *C. arabica* using BLASTp v2.10.1 (Altschul et al., 1990) to retrieve similar sequences with an e-value of 10^{-3}. To verify the presence of characteristic conserved domains (Fig. S2) from each protein family, candidate sequences were then analyzed using Conserved domains (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), and filtered so that only one protein per gene locus remained. The phylogenetic analyses were made from the alignments of these identified coffee sequences together with those described in other species (Table S1).

Transcriptional analysis of *DCLs* and *AGOs* in RNAseq libraries of different coffee tissues

To analyze the transcriptional abundance of the DCL and AGO genes identified, we used published and novel RNA-seq libraries from *C. arabica* (details in Table S1). Libraries were subjected to quality analyses using the program FastQC v.0.11.9 (Robinson et al., 2010) and trimmed as necessary with Trimmomatic v0.39 (Bolger et al., 2014). STAR v.2.7.8a (Dobin et al., 2013) with standard parameters was used to carry out fragment alignments against the genome of *C. arabica*. The resulting libraries were sorted, read duplicates were removed with the Picard toolkit (http://broadinstitute.github.io/picard), and mapped fragments were counted with the htseq-count algorithm (Anders et al., 2015). Expression values were calculated in log (RPKM + 1) using the edgeR package (Robinson et al., 2010), from the Bioconductor R v3.13 project (Huber et al., 2015).

Conserved and novel miRNA prediction in *Coffea sp.*

To search for putative conserved miRNAs and their precursors, we applied an adapted algorithm previously described by (De Souza Gomes et al., 2011) to the genome and transcriptome databases of *C. canephora* (Denoeud et al., 2014). The second prediction of conserved and novel miRNAs was made from RNA-seq libraries of *C. arabica* (de Oliveira et al., 2020; Cardon et al., 2022), in which mapped sRNA reads were used as input to two different computational

pipelines for the discovery of miRNAs – a stringent pipeline for *de novo* identification and a relaxed pipeline for identification of conserved 'known' miRNAs (Jeong et al., 2013). Steps in both pipelines involved processing using Perl scripts as described earlier (Jeong et al., 2011), with a modified version of miREAP (https://sourceforge.net/projects/mireap/) and CentroidFold (Sato et al., 2009). In the 'stringent' criteria pipeline, sRNAs of length between 20 and 24 nt, with abundance ≥ 50 TP30M in at least one library, and total genome hits ≤ 20 were assessed for the potential pairing of miRNA and miRNA* using modified miREAP optimized for plant miRNA discovery with parameters –d 400 –f 25. Strand bias for precursors was computed as the ratio of all reads mapped to the sense strand against total reads mapped to both strands. In addition to strand bias, abundance bias was computed as the ratio of the two most abundant reads against all the reads mapped to the same precursor. Candidate precursors with strand bias ≥ 0.9 and abundance bias ≥ 0.7 were selected, and the foldback structure for the precursor was predicted using CentroidFold. Each precursor was manually inspected to match the criteria as described earlier (Jeong et al., 2013). All the miRNAs identified through this stringent pipeline were then annotated by matching mature sequences to miRBASE (version - 21), and those that did not match any known miRNA were considered lineage or species-specific. In the 'relaxed' criteria pipeline, which is implemented to maximize identification of 'known' miRNAs; relaxed filters were applied – sRNA between 20 and 24nt, with hits ≤ 20 and abundance ≥ 15 TP30M; and precursors with strand bias ≥ 0.7 and abundance bias ≥ 0.4. The stem-loop structure of candidate precursors was visually inspected, the same as the 'stringent' pipeline. Mature sequences of identified miRNAs were further matched with miRBASE entries (version-21), and those with total 'variance' (mismatches and overhangs) ≤ 4 were considered conserved miRNAs. Finally, sequences of the precursors of the two predictions were aligned and all parameters were compared. Redundant sequences were eliminated.

Expression evaluation of conserved and novel miRNAs identified in the *Coffea sp.* genomes

A total of 102,985,416 reads from pre-meiotic, meiotic, and post-meiotic small-RNAs libraries were aligned against the conserved and putative novel

miRNAs identified in the *Coffea sp*. The alignment was performed with ShortStack v. 3.8.5 (Axtell, 2013b) with the parameters "--mmap u --nohp --locifile". Custom R-scripts were developed to plot heatmaps of the inferred abundance of each putative novel miRNAs.

3 RESULTS

Cytogenetic analysis identified the stages of microsporogenesis in *Coffea arabica*

Coffee anthers in different stages of development were harvested from floral buds (FBs) ranging from 1 to 15 millimeters and analyzed using optical microscopy and SEM. Cytogenetic analysis showed that microsporocytes in pre-meiosis I, with cells in interphase (INT), appear in anthers of floral buds up to 1.5 mm (Fig.1 A). In this stage, the nucleus (NC) and nucleolus (NO) are visible, reflecting the chromosome organization to duplicate and form chromatids (Conagin, 1961; Ma, 2005). The cells in meiosis were observed in anthers from floral buds of 1.5 to 4.2 mm (Fig. 1B). In this stage, it was possible to observe cells in two stages of meiosis, prophase (PRO) and telophase (TPH). In prophase, the chromosomes have begun to condense and the nucleoli and nuclear envelope are no longer visible. In the cells in telophase, the last phase of meiosis, the chromosomes unwound and it is possible to see two nucleoli and the nuclear envelope reappears (Conagin, 1961; Ma, 2005).

The post-meiosis stage was identified in anthers from floral buds larger than 4.2 mm, which presented mononucleated (VN) and vacuolated (VC) young microspores, with very thin cell walls, uniform, and clear cytoplasm, and a large, centralized nucleus (Fig. 1C). Binucleated microspores are also observed in post-meiosis, with a generative nucleus (GN) and a vegetative nucleus (VN). Soon after the post-meiotic mitosis, the generative nucleus can be seen attached to the microspore cell membrane; the nucleus then migrates to the center of the cell, already encased in its membrane system (Park & Twell, 2001). The external appearance of anthers in meiosis is shown in figure 1D, anthers with 4.2 mm or the G4 stage following Morais et al., (2008), in which no pollen grains are yet formed (Fig. 1E). The mature pollen occurs at the post-meiosis stage when

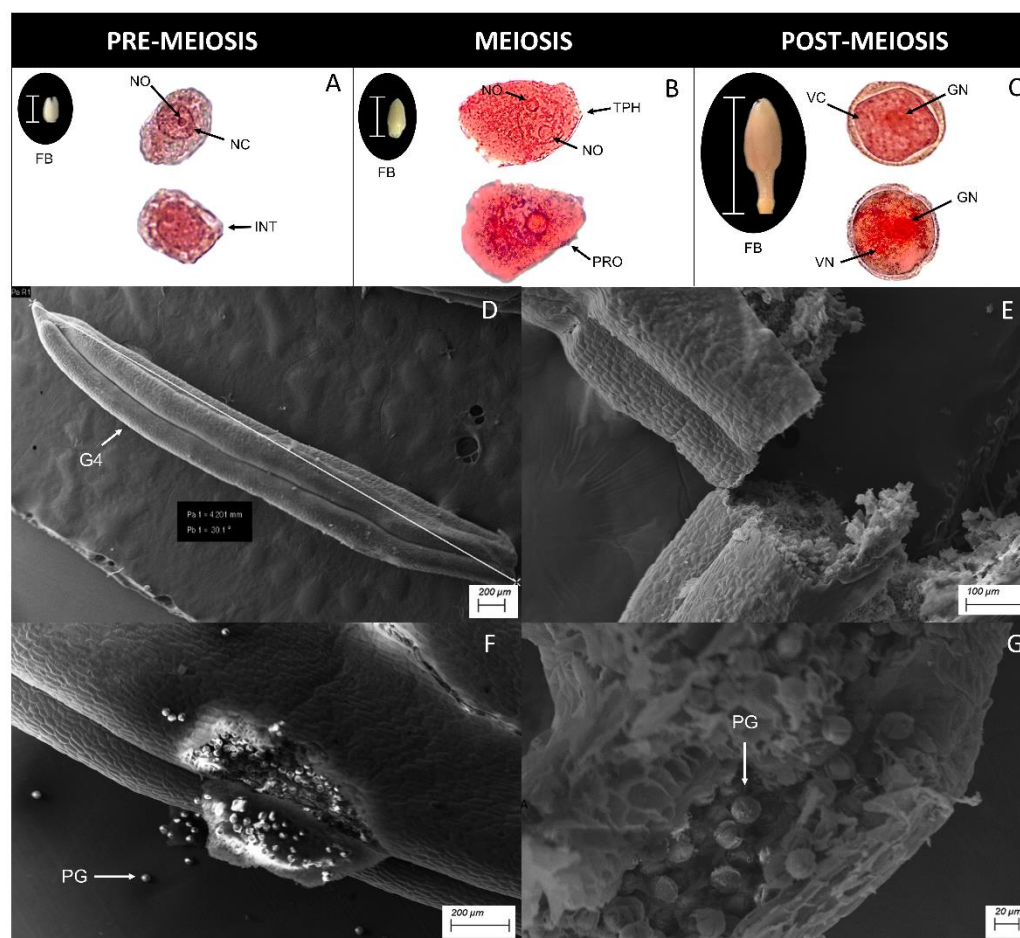anthers are larger than 4.2 mm and the tissue ruptures begin to appear (Fig. 1F and 1G).

Fig. 1 - The microsporogenesis in *Coffea* sp.: cytogenetic analysis by optical and scanning electron microscopy (SEM) of coffee anthers at different developmental stages. Figure A to C show the pre-meiosis, meiosis and post-meiosis stages of microsporogenesis identified in coffee floral buds (FBs) with 1.5 mm, 4.0 mm and 15 mm, respectively. Figure D to G are pictures of SEM showing details of coffee anther and the released mature pollen. A) FB with 1.5 mm (bar in the figure) together with two one-cell microsporocytes at the pre-meiosis stage retrieved from the anthers. The microsporocytes are in the interphase (INT) with a well-defined nucleus (NC) and nucleolus (NO). B) FB with 4.0 mm (bar in the figure) together with two one-cell microsporocytes at the meiosis, one in prophase (PRO) I that marks the beginning of meiosis and other in telophase (TPH), with two visible nucleoli. C) FB with 15 mm (bar in the figure) together with two one-cell microsporocytes at the post-meiosis stage, one young vacuolated (VC) and mononucleated microspore, and one binucleated microspore

with a generative nucleus (GN) and a vegetative nucleus (VN). D to G Pictures of anthers by SEM with mature pollen grains (PG) and size bars indicated in the figures.

Identification and quantification of reproductive phasiRNAs and miRNAs in sRNA-seq libraries

In our analyses, we found seven 21-nt phasiRNAs and ten 24-nt phasiRNAs differently expressed in *C. arabica* sRNA-seq libraries corresponding to the microsporogenesis stages (Fig. 2 and Table S2). The 21-nt phasiRNAs were found to be more enriched in the post-meiotic stage while the 24nt-phasiRNAs are more abundant in the pre-meiotic stage (Fig. 5 = 2 and Table S3 => S2), in contrast to previous reports in monocots (Fan et al., 2016; Zhai et al., 2015). However, a recent study reported that 24-nt phasiRNAs representing more than 50% of all 24-PHAS annotated loci were enriched in pre-meiotic wheat anthers (Bélanger et al., 2020), which suggests that the pattern of accumulation of reproductive phasiRNAs varies across species. The meiotic stage has a lower expression of phasiRNAs in both monocots and eudicots.
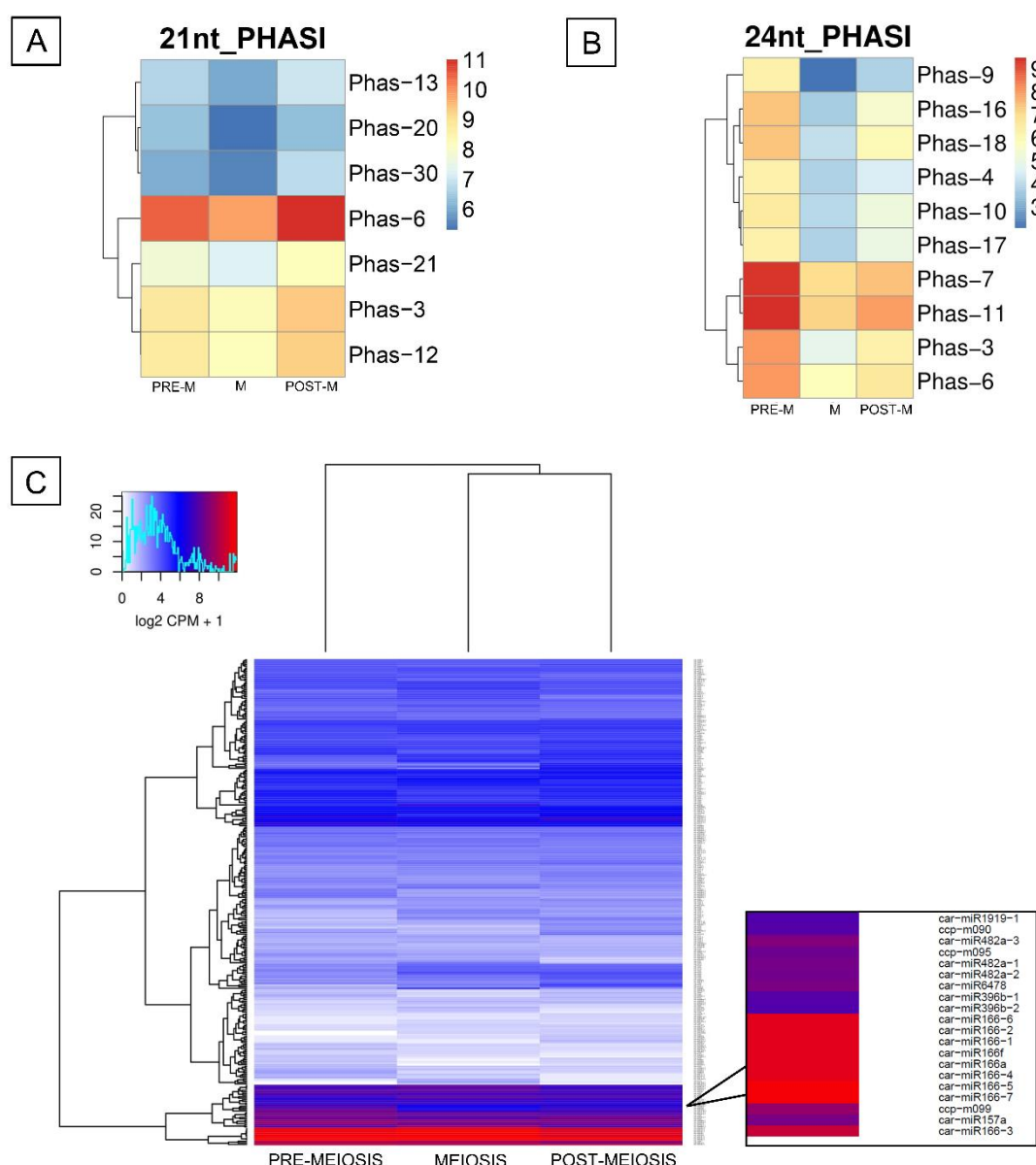
Fig. 2. Expression profile of phasi and miRNAs identified in *Coffea arabica* L. A: Heatmap of 21-nt phasiRNAs (lines) identified in libraries (columns) of anthers in pre-meiosis (PRE-M), meiosis (M), and post-meiosis (POST-M). B: Heatmap of 24-nt phasiRNAs (lines) identified in libraries (columns) of anthers in pre-meiosis (PRE-M), meiosis (M), and post (POST-M). C: Heatmap of conserved and novel miRNAs (lines) identified in libraries (columns) of anthers in pre-meiosis, meiosis, and post-meiosis. The dendrogram on top of the heatmaps reflects the relationship of different conditions while the dendrogram on the left reflects the relationship of different sRNAs across conditions. In figures A and B, the colors are coded as a function of logarithmic base 10 of expression values normalized in FPKM +1 (Fragments per million mapped reads), whereas in figure C colors are coded as a function of logarithmic base 2 of expression values normalized in CPM+1 (Counts per million mapped reads).

miR2118/482 is described as responsible for processing 21-nt phasiRNAs while miR2275 processes 24-nt phasiRNAs (Kakrana et al., 2018; Komiya, 2017; Xia et al., 2019). Thus, to evaluate the presence of these miRNA triggers and others probably related to microsporogenesis development, we identified the miRNA precursors in the *Coffea canephora* genome (Table S3). From this, 138 miRNA precursors and 149 gender-specific mature miRNAs, 80 of which were miRNA-5p and 69 miRNA-3p, were identified in the *Coffea canephora* genome (Table S3). We also identified 259 conserved miRNA precursors that gave rise to 349 mature miRNAs, including 98 miRNA-5p and 251 miRNA-3p sequences (Table S3).

The main miRNA families were found together with conserved and lineage-specific miRNAs, and were quantified in the sRNA libraries; for example, miR157, miR396, and miR482 families were some of the most abundant in all the stages of microsporogenesis in *C. arabica* (Fig. 2C). The miR166 was also found to be notably abundant (Fig. 2C), this abundance of miR166 in Illumina is a common phenomenon, as noted in Baldrich, et al. (2015). In our results, car-miR482 is one of the main triggers for 21-nt phasiRNAs, whereas car-miR2275 appears as a trigger for car-miR2275 both 21-nt and 24-nt phasiRNAs (Table S4). We have not identified miR2118 in our data, so in addition to car-miR2275, results suggest that other miRNAs can cleave 24-nt phasiRNAs.

Characterization of DCL and AGO proteins in the *Coffea arabica* genome

DICER-like (DCL) and ARGONAUTE proteins (AGO) are two families of proteins related to the production of siRNAs and phasiRNAs, which are generated by specific and divergent members in plants (Baumberger & Baulcombe, 2005; Liu et al., 2009). Thus, to identify DCLs and AGOs and the possibility to generate phasiRNAs during microsporogenesis, we characterized the two families in the *C. arabica* genome. As a result, we found fourteen DCL and twenty AGO coffee proteins that we further classified through phylogenetic analyses (Figs. 3 and 4), comparing them with homologous sequences of other monocot and eudicot plants (Table S1). We found three copies of CaDCL1, in which CaDCL1.1 and CaDCL1.2 were highly conserved and considered the same, but derived from a

different parental subgenome, while CaDCL1.3 was truncated (Table S1 and Fig. S1); The DCL2 clade was the group with the highest number of CaDCLs homologs, eight, but only CaDCL2.1 presented the complete main domains (Table S1 and Fig. S1); two DCL3 copies were found to be similar in sequence but from different parental subgenomes; and finally, only one CaDCL4 copy was found, from the *C. eugenioides*' subgenome (Table S1). Compared to other works, here we found five more DCLs than previously described by Mosharaf et al. (2019) and eight more than those published by Noronha Fernandes-Brum et al. (2017). Details of all the sequences are available in Table S1.
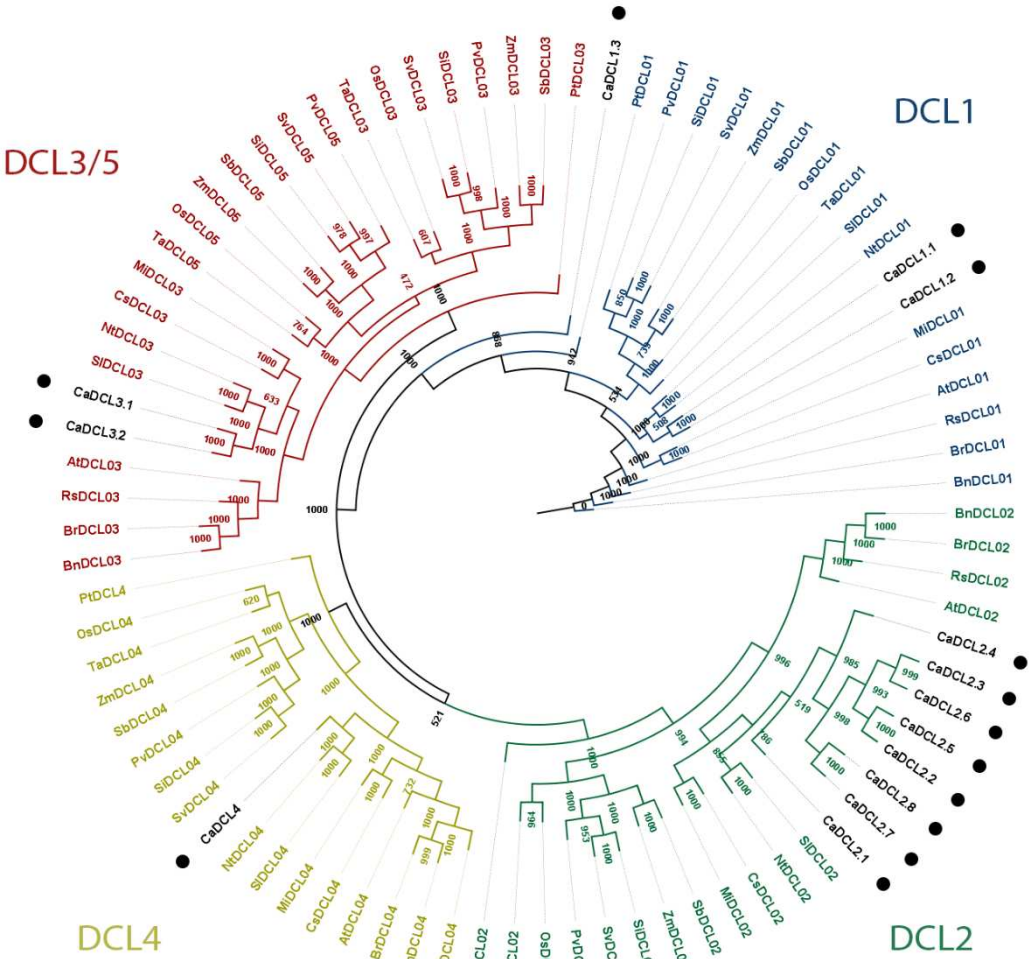
Fig. 3 - Phylogenetic analysis of coffee Dicer-like (DCL) proteins. The tree was generated through the neighbor-joining method containing DCL protein sequences identified in *C. arabica* (black dots) together with sequences of monocots, eudicots and gymnosperms. Details of all the sequences used are described in Table S1. The four major clades of DCLs were retrieved from sequences and indicated around the tree, being the group DCL3/DCL5 joined with DCL5, found only in monocots, due its evolutionary relationship (CHEN et al., 2022). A total of fourteen *C. arabica* DCL

sequences were identified, three DCL1 homologs, eight DCL2, two DCL3 and one DCL4 (six complete sequences and eight truncated: CaDCL1.3, CaDCL2.2, CaDCL2.3, CaDCL2.4, CaDCL2.5, CaDCL2.6, CaDCL2.7, CaDCL2.8). Legends: *Arabidopsis thaliana* (At), Mangifera indica (Mi), Brassica napus (Bn), Raphanus sativus (Rs), Citrus sinensis (Cs), Brassica rapa (Br), Nicotiana tabacum (Nt), Solanum lycopersicum (Sl), Vitis vinifera (Vv), Fragaria vesca (Fv), Zea mays (Zm), Sorghum bicolor (Sb), Setaria viridis (Sv), Setaria italica (Si), Panicum virgatum (Pv), Oryza sativa (Os), Triticum aestivum (Ta), Pinus tabuliformis (Pt).

Dicer and DCL proteins are a unique class of ribonucleases that generally contain six types of domains, including the DEAD-box, helicase-C, DUF283, PAZ, RNase III, and dsRBD domains. However, for some species, one or more of these domains may be absent (Liu et al., 2009; Margis et al., 2006). All identified CaDCLs presented Ribonuclease III domains, which have been considered one of the key domains for DCL proteins (Nicholson, 2014), except for CaDCL1.3 which only presented two (Fig. S1). In addition, CaDCL1.1, CaDCL1.2, CaDCL2.1, CaDCL3.1, CaDCL3.2, and CaDCL4 had the highest number of domains in common, Ribonuclease III, DEAD/RES III, Helicase_C, Dicer_dimer, PAZ, and DSRM (Fig. S1). Last instead being a divergent double-stranded RNA-binding domain coinciding with the DUF283 of Dicer (Dlakić, 2006).

Flowering plants have ten or more AGO proteins (Zhai et al., 2015). In total, 20 AGO-like proteins were found in the genome of *C. arabica* and distributed in three distinct clades of the phylogenetic tree (Fig 4), one CaAGO1, four CaAGO2, five CaAGO4, two CaAGO5, two CaAGO6, two CaAGO7, and four CaAGO10 were identified (Table S1). CaAGO1, CaAGO5, and CaAGO10 proteins were clustered into clade 1 which comprises the AGO1/5/10 proteins (Fig 4). Proteins CaAGO2 and CaAGO7 were grouped into clade II, which comprises AGO2/3/7 proteins. CaAGO4 and CaAGO 6 proteins were clustered into clade III, which comprises the AGO4/6/8/9 proteins. Among the AGO family members identified, all had copies in both subgenomes of *C. arabica*, except for CaAGO1, for which we identified only one copy on chromosome 4 from *C. canephora* (Table S1). In addition to the AGO sequences previously identified in the genome of *C. canephora* (Noronha Fernandes-Brum et al., 2017), we identified two more copies of AGO2, one copy of AGO5, and three AGO4 from

C. eugenioides' chromosomes. We also identified one new AGO6 and two new copies of AGO10, one from each of the two subgenomes.
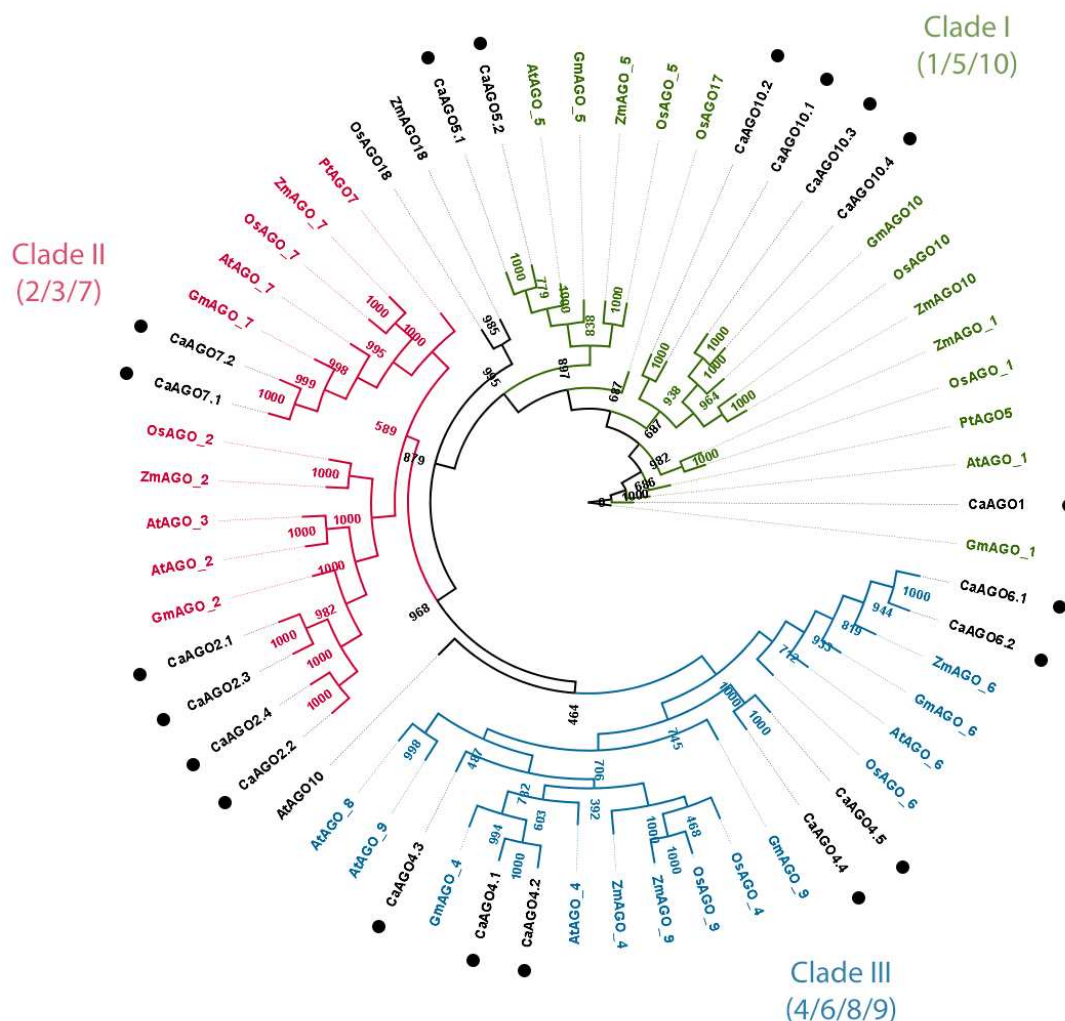

Fig. 4 - Phylogenetic analysis of coffee ARGONAUTA (AGO) proteins. The tree was generated through the neighbor-joining method containing AGO protein sequences identified in *C. arabica* (black dots) together with sequences of monocots, eudicots and gymnosperms. Details of all the sequences used are described in Table S1. The three major clades of AGOs were retrieved from sequences and indicated around the tree, clade I (AGO1/5/10), clade II (AGO2/3/7) and clade III (AGO4/6/8/9). A total of twenty *C. arabica* AGO sequences were identified, one AGO1 homolog, four AGO2, five AGO5, two AGO6, two AGO7 and ten AGO10. Legends: *Arabidopsis thaliana* (At), *Glycine max* (Gm), *Oryza sativa* (Os), *Zea mays* (Zm), *Pinus tabuliformis* (Pt).

AGO proteins are characterized by ArgoN (amino-terminal N), PAZ (PIWI–ARGONAUTE–ZWILLE), MID (middle), and PIWI domains (Meister, 2013). The conserved domains of the CaAGO proteins were analyzed, and all of them presented ArgoN, PAZ, PIWI, and ArgoL1 domains (Fig. S2). CaAGO1, CaAGO4.1, CaAGO4.2, CaAGO10.1, CaAGO10.2, CaAGO10.3, and

CaAGO10.4 had all 6 domains ArgoN, PAZ, Piwi, ArgoL1, ArgoL2 and ArgoMID. The AGO1 proteins have an additional N-glycine-rich region (Gly-rich Ago1) domain that was present both in the CaAGO1 identified in this work and previously (in the CcAGO1; Fernandes-Brum et al. 2017). The MID domain is present in eight identified CaAGOs (CaAGO1, CaAGO 4.1, CaAGO 4.2, CaAGO 5.1, CaAGO 10.1, CaAGO 10.2, CaAGO 10.3, CaAGO 10.4). The MID domain anchors the 5′ ends of the small RNA by providing a binding pocket in which the 5′ terminal base engages in stacking interactions with a conserved tyrosine. In addition, several hydrogen bonds coordinate correct 5′ end binding (Meister, 2013).

Expression analyses of DCL and AGO genes in different RNAseq libraries

To analyze the expression profile of CaDCLs and CaAGOs identified in this work in different tissues of the coffee tree, we aligned these sequences against RNA-seq libraries from *C. arabica* tissues under different conditions: leaves from plants under optimal and warm temperature conditions (cv. Acauã and Catuaí); FBs at G4 and G5 stages (cv. Catuaí and Siriema), embryos, stems, roots, meristems, floral buds, green drupes, yellow drupes, red drupes, fruits at multiple stages of development, fruits 30 days after flowering, and fruits 90 days after flowering (see details in Table S1).

All fourteen CaDCLs identified here were found to be expressed in at least one of these libraries (Fig. 5A). CaDCL2.1, CaDCL2.2, CaDCL3.1, CaDCL3.2 and CaDCL4 were the most highly expressed, with CaDCL3.2 more expressed in the reproductive libraries that correspond to FBs with 3.1 to 10 mm or the G4 and G5 stages described by Morais et al. (2008; G4_CA, G5_CA, G4_SE, and G5_SE in Fig. 5A). Of the twenty AGO proteins identified in this work, eighteen were expressed in the analyzed libraries (Fig. 5B). *CaAGO1* was the most expressed and appeared in all libraries, whereas CaAGO4.4 and CaAGO4.5 were not found expressed in any of the libraries.
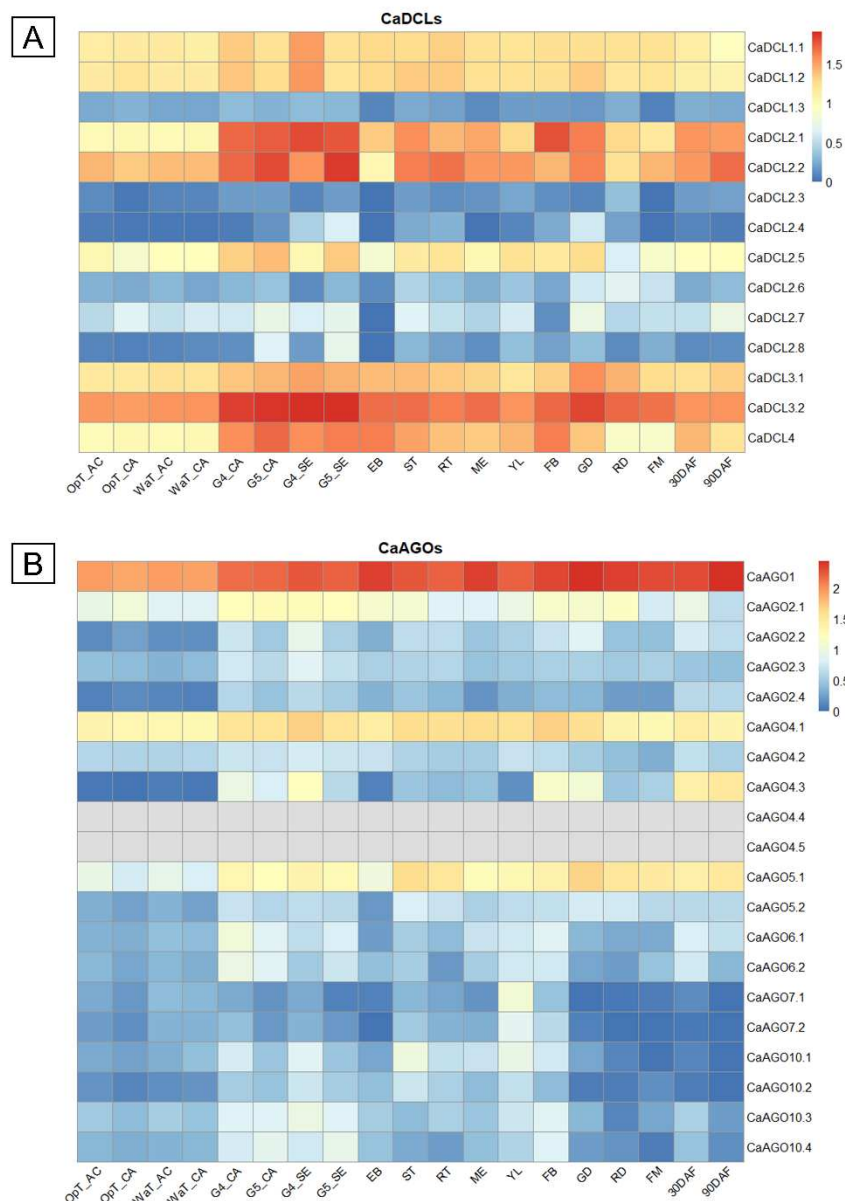
Fig. 5. Expression profile of DCLs (A) and AGOs (B) genes identified in *Coffea arabica* L. The coffee genes are indicated in the right column of each figure and the RNA-Seq libraries relative to different coffee tissues, in the bottom line. Colors are coded as a function of logarithmic base 10 of expression values normalized in FPKM +1 (Fragments per million mapped reads). After validation of RNA-Seq quality parameters (details in Table S1 and Material and Methods), the expression profiles were determined in 19 different coffee tissues: OpT_AC and OpT_CA, *C. arabica* cv. Acauã (AC) and Catuaí (CA) leaves from plants in optimal temperature conditions; WaT_AC and WaT_CA, *C. arabica* cv. Acauã (AC) and Catuaí (CA) leaves from plants in warm temperature conditions; G4_CA and SE: floral buds at G4 stage from *C. arabica* cv. Catuaí (CA) and Siriema (SE); G5_CA and SE: floral buds at G4 stage from *C. arabica*

cv. Catuaí (CA) and Siriema (SE); EB: embryos; ST: stems; RT: roots; ME: meristems; YL: yellow drupes; FB: floral buds; GD: green drupes; RD: red drupes; FM: fruits at multiple stages of development; 30DAF: fruits 30 days after flowering; 90DAF: fruits 90 days after flowering. The accession number of these RNAseq libraries and references are available at Table S1.

4 DISCUSSION

To comprehend the microsporogenesis and associated molecular pathways governing anther development in coffee, the cytological stages of meiosis were first determined. Here, we identified that the crucial steps of microsporogenesis, that is, pre-meiosis, meiosis, and post-meiosis, occur in FBs of length, respectively, less than 1.5 mm, 1.5 to 4.2 mm, and greater than 4.2 mm (Fig. 1A to C). Mature pollen grains are only observed after these stages (Fig. 1D to G). Compared to established classifications of coffee floral development, these meiosis stages are recognized in the G3 and G4 stages following Morais et al. (2008) and in Stage 4 following De Oliveira et al. (2014). Interestingly, coffee meiosis coincides with the Brazilian winter, marked by drought and low temperatures, a period of reduced growth of FBs, interpreted as dormant or latent (reviewed by López et al., 2021). A discussion about this terminology is provided by Considine & Considine (2016), but here we adopt the term "latent" because there is no evidence of the physiological basis of coffee dormancy (López et al. (2021)

Thus, here we show that the coffee meiosis occurs when FBs are apparently "dormant or latent" in the G3/G4 stage, since no external growth is visible but inside there is a high activity at the cellular level due to gametogenesis. This agrees with the reported post-meiosis observed in anthers of FBs with 6 to 8 mm, that contains microspores shortly after tetrad formation with the majority in mid- to late-uninucleate stage (Neuenschwander & Baumann, 1995; De Oliveira et al., 2014). Considering that microspore mitosis took place two to three days before anthesis in FBs of length 12-15 mm (Neuenschwander & Baumann, 1995), we now have the completed cycle of microsporogenesis and anthers development described.

Since the external growth at G3/G4 stages decreases, while an intense gametic development occurs within FBs, it is tempting to speculate that the plant's metabolic energy is shifted to gametogenesis. Moreover, it seems plausible to

associate this energetic control to temperature sensitivity during gametogenesis. Coffee is sensitive to warmth, changing its sugar content (De Oliveira et al., 2020), and some studies have reported that lower temperatures dramatically enhance pollen vigor while higher temperatures have the opposite effect (Wang et al., 2023). In accordance, pre-culturing FBs for two days at 4 °C derange the first division of the microspore nucleus (Neuenschwander & Baumann, 1995). Thus, while cold seems to have a beneficial role in coffee gametogenesis, the opposite seems to be true, with warming producing unreduced gametes, as observed in other species (Pécrix et al. 2011; Wang et al., 2017). Since a high temperature environment has the potential to increase gamete ploidy levels (Pécrix et al. 2011), this could explain the formation of diplogametes and the allotetraploid of *C. arabica* that originated in a period of global adverse conditions. Exploring these hypotheses and associated molecular mechanisms is necessary to make advances in the control of coffee fertility and ploidy for plant breeding programs.

Reproductive phasiRNAs are crucial for the microsporogenesis of many species, especially monocots (Liu & Wang, 2021; Liu et al., 2020). It was previously believed that only this group had 24-nt phasiRNAs, because their processing depends on a specialized DCL5 never reported in eudicots. However, recent studies have demonstrated the presence of these phasiRNAs in eudicots (Xia et al., 2019) and now we have also identified this class of molecules in *C. arabica*. Compared to monocots and other eudicots, we identified a smaller number of reproductive phasiRNAs, and a specific temporal pattern, with the 21 nt-phasiRNAs being more expressed in post-meiosis, and not in pre-meiosis, and the 24-nt phasiRNAs in pre-meiosis and not in meiosis (Kakrana et al., 2018; Pokhrel et al.,, 2021; Xia et al., 2019), consistent with the recent report in wheat by Bélanger, et al. (2020).

Another relevant element in the biogenesis of reproductive phasiRNAs is the miRNAs that act as their triggers. In several species, such as barley and wheat (Bélanger et al., 2020), and members of the family *Solanaceae* (Xia et al., 2019b), other triggers for 24-nt phasiRNAs have been described for in addition to the canonical miRNA miR2275. Interestingly, miR2118/482, known to trigger 21-nt phasiRNAs in eudicots and monocots, also has a role in generating 24-nt

phasiRNAs in the eudicots columbine and wild strawberry. (Pokhrel et al., 2021). This suggests that miR2118/482 members present in gymnosperms may initiate 24-nt phasiRNA expression in angiosperms/basal eudicots, but miR2275 may have arisen- and therefore assumed this role- later. In *C. arabica*, miR2118 seems to have been lost, while the miR482 family acts as a trigger for 21-nt phasiRNAs, possibly together with the miR2275 family that is supposed to have assumed a role in the processing of both reproductive phasiRNAs. Further experiments are necessary to investigate this hypothesis.

The recent discovery of 24-nt phasiRNAas in eudicots drew attention to which DCL could be involved in their biosynthesis process since there is no DCL5 reported yet (Liu et al., 2020; Pokhrel et al., 2021; Xia et al., 2019). The emergence of DCL5 in monocots is believed to be explained by a subfunctionalization of the ancestral DCL3 protein, which is suggested to function in the production of heterochromatic siRNAs (hc-siRNAs) and phasiRNAs (Liu et al., 2020). In rice, OsDCL3 and OsDCL5 have completely different substrate specificities, while the eudicot AtDCL3 has an intermediate preference for dsRNAs with an overhanging 5' triphosphate and 3' structure (Chen et al., 2022). This suggests that monocot DCL5 and DCL3 proteins were optimized for cognate substrates after duplication of the progenitor 'eudicot-like' DCL3 protein and this functional specialization appears to have been achieved through the accumulation of mutations in the PAZ domain (Chen et al., 2022). Therefore, it is hypothesized that DCL3 has dual functions in some eudicots, processing distinct substrates (Pol II vs Pol IV RNA) for the biogenesis of 24-nt phasiRNAs or 24-nt hc-siRNAs (Xia et al., 2019). Our results support this (Fig. 3 and S1), and future work testing protein activity should explore this possibility.

Interestingly, of the three *CaDCL* genes expressed in the tissues analyzed in this study, *CaDCL3.2* was more abundant in the libraries of FBs in the G4 and G5 stages (Fig. 5A). Such stages correspond to the meiosis and post-meiosis stages identified here, supporting the hypothesis of a DCL3 role in the biogenesis of siRNAs during coffee microsporogenesis. DCL4, described as responsible for cleaving 21-nt phasiRNAs, identified in this work, was expressed in all analyzed tissues, mainly in G4, G5 and embryos, which reinforces the involvement of this molecule in the process (Fig. 5A) (Kakrana et al., 2018; Komiya, 2017; Xia et al.,

2019). Each class of DCL has evolved to participate in a primary pathway, however, DCL2/DCL4 can also function as partial overlap, because defects in one class of DCL may be compensated by other classes in some cases (Fukudome & Fukuhara, 2017; Pokhrel et al., 2021).

In our study we identified twenty members of the CaAGO-like protein family (Fig. 4). The high number of AGO protein family members in flowering plants points to functional diversification, presumably reflecting regulatory pathways targeting the expanding sRNA class of molecules (Zhang et al., 2015). Members of the AGO4/6/8/9 clade have been described as effectors of 24 nt hc-siRNAs and are known to function in the regulation of transcription of genes (Zhang et al. 2015). In some eudicots, studies based on phylogenetic and expression data suggested that AGO5 and AGO1 could act as effector proteins for 21-nt phasiRNAs and AGO6 and AGO9 for 24-nt phasiRNAs (Pokhrel et al., 2021). In our data, at least one copy of each member of the *CaAGO*-like gene family was expressed in the libraries (Fig. 5B), with CaAGO4.2 and CaAGO4.3 being more abundant in the reproductive libraries, suggesting the role of the codified enzymes in the biogenesis of 24-nt phasiRNAs. *CaAGO6* was not expressed in the analyzed libraries and we did not identify an AGO9 homolog in our data.

5 CONCLUSIONS

In this work, we used *Coffea arabica* L., a recent allotetraploid formed from the hybridization of *C. canephora* and *C. eugenioides* unreduced gametes, to explore microsporogenesis and small RNA related pathways in a eudicot crop. We identified the pre-meiosis, meiosis and post-meiosis stages during anther development and, together with other work, provided a more complete overview of coffee microsporogenesis. Such stages coincide with the cold period in Brazilian conditions, which was associated with a benefit for pollen formation; on the contrary, warm conditions could be related to the formation of unreduced gametes and ploidy disturbance. Next, to explore siRNAs pathways, we identified and quantified the expression of reproductive 21- and 24-nt phasiRNAs, related miRNAs, and DLCs and AGOs. The presence of reproductive phasiRNAs in *C. arabica* presents a temporal pattern different from most of the eudicots already

studied, with 21-phasi RNAs enriched in post-meiosis and 24-nt phasi-RNAs in pre-meiosis, which could reflect a diversification along the evolutionary history of woody eudicot species. The triggers identified here reinforce data that canonical microRNAs such as miR2275 may be involved in the processing of both 21 and 24-nt reproductive phasiRNAs. Despite fourteen DCLs being found here, DCL5 was not found in coffee; DCL5 is related to phasiRNA biosynthesis in monocots, and its absence here supports the hypothesis that this function could have evolved and been maintained by a closely related DCL3-like protein. Thus, our work contributes to highlighting the microsporogenesis and related reproductive siRNAs pathways in coffee and contributes towards the control of reproductive development and improvement of fertility in eudicots of economic interest.

AUTHORS' CONTRIBUTIONS

ACKNOWLEDGEMENTS

FUNDING

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY

Main data supporting the findings of this study are available within the paper and within its supplementary materials published online. The raw data used for analyses and figures are available from the corresponding author, Antonio Chalfun-Júnior, upon request. Raw data for the RNAseq runs can be obtained through BioProject ID PRJNA923233. Raw data for the smallRNAseq runs can be obtained through BioProject ID PRJNA925088. The respective SRA accessions are available in Supplemental Table S1.

REFERENCES

Adhikari, P. B., Liu, X., Wu, X., Zhu, S., & Kasahara, R. D. (2020). Fertilization in flowering plants: an odyssey of sperm cell delivery. *Plant molecular biology*, *103*, 9-32.https://doi.org/10.1007/s11103-020-00987-z

Albert, B., Toghranegar, Z., & Nadot, S. (2014). Diversity and evolution of microsporogenesis in Bromeliaceae. *Botanical Journal of the Linnean Society*, *176*(1), 36-45.https://doi.org/10.1111/boj.12191

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *bioinformatics*, *31*(2), 166-169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, S. (2010). FastQC A Quality Control tool for High Throughput Sequence Data [Online][WWW Document]. *URL http://www. bioinformatics. babraham. ac. uk/projects/fastqc/(accessed 11.19. 20).*

Åstrand, J., Knight, C., Robson, J., Talle, B., & Wilson, Z. A. (2021). Evolution and diversity of the angiosperm anther: trends in function and development. *Plant reproduction*, *34*(4), 307-319. https://doi.org/10.1007/s00497-021-00416-1

Axtell, M. J. (2013a). Classification and comparison of small RNAs from plants. *Annual review of plant biology*, *64*, 137-159. https://doi.org/10.1146/annurev-arplant-050312-120043

Axtell, M. J. (2013b). ShortStack: comprehensive annotation and quantification of small RNA genes. *Rna*, *19*(6), 740-751. http://www.rnajournal.org/cgi/doi/10.1261/rna.035279.112.

Baldrich, P., Campo, S., Wu, M. T., Liu, T. T., Hsing, Y. I. C., & Segundo, B. S. (2015). MicroRNA-mediated regulation of gene expression in the response of rice plants to fungal elicitors. *RNA biology*, *12*(8), 847-863.https://doi.org/10.1080/15476286.2015.1050577

Baumberger, N., & Baulcombe, D. C. (2005). Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences*, *102*(33), 11928-11933.https://doi.org/10.1073/pnas.0505461102

Bélanger, S., Pokhrel, S., Czymmek, K., & Meyers, B. C. (2020). Premeiotic, 24-nucleotide reproductive PhasiRNAs are abundant in anthers of wheat and barley but not rice and maize. *Plant physiology*, *184*(3), 1407-1423. https://doi.org/10.1104/pp.20.00816

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2009). GenBank. *Nucleic acids research*, *37*(suppl_1), D26-D31.https://doi.org/10.1093/nar/gkn723

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Chen, S., Liu, W., Naganuma, M., Tomari, Y., & Iwakawa, H. O. (2022). Functional specialization of monocot DCL3 and DCL5 proteins through the evolution of the PAZ domain. *Nucleic Acids Research*, *50*(8), 4669-4684. https://doi.org/10.1093/nar/gkac223

Conagin, C. H. (1961). Microsporogênese, incompatibilidade e esterilidade masculina em Coffea congensis Froehner. *Bragantia*, *20*, 669-677. https://doi.org/10.1590/S0006-87051961000100027

Considine, M. J., & Considine, J. A. (2016). On the language and physiology of dormancy and quiescence in plants. *Journal of Experimental Botany*, *67*(11), 3189-3203. https://doi.org/10.1093/jxb/erw138

DaMatta, F. M., Rahn, E., Läderach, P., Ghini, R., & Ramalho, J. C. (2019). Why could the coffee crop endure climate change and global warming to a greater extent than previously estimated?. *Climatic Change*, *152*, 167-178. https://doi.org/10.1007/s10584-018-2346-4

DaMatta, F. M., & Ramalho, J. D. C. (2006). Impacts of drought and temperature stress on coffee physiology and production: a review. *Brazilian journal of plant physiology*, *18*, 55-81. https://doi.org/10.1590/S1677-04202006000100006

Davis, M. P., van Dongen, S., Abreu-Goodger, C., Bartonicek, N., & Enright, A. J. (2013). Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods*, *63*(1), 41-49. https://doi.org/10.1016/j.ymeth.2013.06.027

Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M. & Lashermes, P. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *science*, *345*(6201), 1181-1184. https://doi.org/10.1126/science.1255274

De Oliveira, R. R., Ribeiro, T. H. C., Cardon, C. H., Fedenia, L., Maia, V. A., Barbosa, B. C. F., ... & Chalfun-Junior, A. (2020). Elevated temperatures impose transcriptional constraints and elicit intraspecific differences between coffee genotypes. *Frontiers in Plant Science*, *11*, 1113. https://doi.org/10.3389/fpls.2020.01113

De Oliveira, R. R., Cesarino, I., Mazzafera, P., & Dornelas, M. C. (2014). Flower development in Coffea arabica L.: new insights into MADS-box genes. *Plant reproduction*, *27*, 79-94. https://doi.org/10.1007/s00497-014-0242-2

De Souza Gomes, M., Muniyappa, M. K., Carvalho, S. G., Guerra-Sá, R., & Spillane, C. (2011). Genome-wide identification of novel microRNAs and their target genes in the human parasite Schistosoma mansoni. *Genomics*, *98*(2), 96-111. https://doi.org/10.1016/j.ygeno.2011.05.007

Dhaka, N., Sharma, S., Vashisht, I., Kandpal, M., Sharma, M. K., & Sharma, R. (2020). Small RNA profiling from meiotic and post-meiotic anthers reveals prospective miRNA-target modules for engineering male fertility in sorghum. *Genomics*, *112*(2), 1598-1610. https://doi.org/10.1016/j.ygeno.2019.09.009

Dlakić, M. (2006). DUF283 domain of Dicer proteins has a double-stranded RNA-binding fold. *Bioinformatics*, *22*(22), 2711-2714. https://doi.org/10.1093/bioinformatics/btl468

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. https://doi.org/10.1093/bioinformatics/bts635

Fan, Y., Yang, J., Mathioni, S. M., Yu, J., Shen, J., Yang, X., ... & Zhang, Q. (2016). PMS1T, producing phased small-interfering RNAs, regulates photoperiod-sensitive male sterility in rice. *Proceedings of the National Academy of Sciences*, *113*(52), 15144-15149. https://doi.org/10.1073/pnas.1619159114

Fei, Q., Yang, L., Liang, W., Zhang, D., & Meyers, B. C. (2016). Dynamic changes of small RNAs in rice spikelet development reveal specialized reproductive phasiRNA pathways. *Journal of Experimental Botany, 67*(21), 6037-6049. https://doi.org/10.1093/jxb/erw361

Fei, Q., Xia, R., & Meyers, B. C. (2013). Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *The Plant Cell*, *25*(7), 2400-2415. https://doi.org/10.1105/tpc.113.114652

Fukudome, A., & Fukuhara, T. (2017). Plant dicer-like proteins: double-stranded RNA-cleaving enzymes for small RNA biogenesis. *Journal of Plant Research*, *130*, 33-44. https://doi.org/10.1007/s10265-016-0877-1

Gardner, R. O., & Rattenbury, J. A. (1974). Staining Pollen Nuclei with Snows Alcoholic Hydrochloric Acid-Carmine. *Stain technology*, *49*(5), 257-259. https://doi.org/10.3109/10520297409116990

Jeong, D. H., Thatcher, S. R., Brown, R. S., Zhai, J., Park, S., Rymarquis, L. A., ... & Green, P. J. (2013). Comprehensive investigation of microRNAs enhanced by analysis of sequence variants, expression patterns, ARGONAUTE loading, and target cleavage. *Plant physiology*, *162*(3), 1225-1245. https://doi.org/10.1104/pp.113.219873

Jeong, D. H., Park, S., Zhai, J., Gurazada, S. G. R., De Paoli, E., Meyers, B. C., & Green, P. J. (2011). Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. *The Plant Cell*, *23*(12), 4185-4207. https://doi.org/10.1105/tpc.111.089045

Kakrana, A., Li, P., Patel, P., Hammond, R., Anand, D., Mathioni, S. M., & Meyers, B. C. (2017). PHASIS: a computational suite for de novo discovery and characterization of phased, siRNA-generating loci and their miRNA triggers. *BioRxiv*, 158832. https://doi.org/10.1101/158832

Kakrana, A., Mathioni, S. M., Huang, K., Hammond, R., Vandivier, L., Patel, P., ... & Meyers, B. C. (2018). Plant 24-nt reproductive phasiRNAs from intramolecular duplex mRNAs in diverse monocots. *Genome Research*, *28*(9), 1333-1344. http://www.genome.org/cgi/doi/10.1101/gr.228163.117.

Komiya, R. (2017). Biogenesis of diverse plant phasiRNAs involves an miRNA-trigger and Dicer-processing. *Journal of plant research*, *130*, 17-23. https://doi.org/10.1007/s10265-016-0878-0

Lashermes, P., Combes, M. C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F., & Charrier, A. (1999). Molecular characterisation and origin of the Coffea arabica L. genome. *Molecular and General Genetics MGG*, *261*(2), 259-266. https://doi.org/10.1007/s004380050965

Lee, C. H., & Carroll, B. J. (2018). Evolution and diversification of small RNA pathways in flowering plants. *Plant and Cell Physiology*, *59*(11), 2169-2187. https://doi.org/10.1093/pcp/pcy167

Liu, L., & Wang, T. (2021). Male gametophyte development in flowering plants: A story of quarantine and sacrifice. *Journal of Plant Physiology*, *258*, 153365. https://doi.org/10.1016/j.jplph.2021.153365

Liu, Q., Feng, Y., & Zhu, Z. (2009). Dicer-like (DCL) proteins in plants. *Functional & integrative genomics*, *9*, 277-286. https://doi.org/10.1007/s10142-009-0111-5

Liu, Y., Teng, C., Xia, R., & Meyers, B. C. (2020). PhasiRNAs in plants: their biogenesis, genic sources, and roles in stress responses, development, and reproduction. *Plant Cell*, *32*(10), 3059-3080. https://doi.org/10.1105/tpc.20.00335

López, M. E., Santos, I. S., de Oliveira, R. R., Lima, A. A., Cardon, C. H., & Chalfun-Junior, A. (2021). An overview of the endogenous and environmental factors related to the Coffea arabica flowering process. *Beverage Plant Research*, *1*(1), 1-16. https://doi.org/10.48130/BPR-2021-0013

Ma, H. (2005). Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. *Annu. Rev. Plant Biol.*, *56*, 393-434. https://doi.org/10.1146/annurev.arplant.55.031903.141717

Margis, R., Fusaro, A. F., Smith, N. A., Curtin, S. J., Watson, J. M., Finnegan, E. J., & Waterhouse, P. M. (2006). The evolution and diversification of Dicers in plants. *FEBS letters*, *580*(10), 2442-2450. https://doi.org/10.1016/j.febslet.2006.03.072

Mathioni, S. M., Kakrana, A., & Meyers, B. C. (2017). Characterization of plant small RNAs by next generation sequencing. *Current Protocols in Plant Biology*, *2*(1), 39-63. https://doi.org/10.1002/cppb.20043

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nature Reviews Genetics*, *14*(7), 447-459. https://doi.org/10.1038/nrg3462

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., ... & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic acids research*, *49*(D1), D412-D419. https://doi.org/10.1093/nar/gkaa913

Morais, H., Caramori, P. H., Koguishi, M. S., & Ribeiro, A. M. D. A. (2008). Escala fenológica detalhada da fase reprodutiva de Cooffea arabica. *Bragantia*, *67*, 257-260. https://doi.org/10.1590/S0006-87052008000100031

Mosharaf, M. P., Akond, Z., Kabir, M. H., & Mollah, M. N. H. (2019). Genome-wide identification, characterization and phylogenetic analysis of Dicer-like (DCL) gene family in Coffea arabica. *Bioinformation*, *15*(11), 824. https://doi.org/10.6026%2F97320630015824

Nadot, S., Furness, C. A., Sannier, J., Penet, L., Triki-Teurtroy, S., Albert, B., & Ressayre, A. (2008). Phylogenetic comparative analysis of microsporogenesis in angiosperms with a focus on monocots. *American Journal of Botany*, *95*(11), 1426-1436. https://doi.org/10.3732/ajb.0800110

Neuenschwander, B., & Baumann, T. W. (1995). Increased frequency of dividing microspores and improved maintenance of multicellular microspores of Coffea arabica in medium with coconut milk. *Plant cell, tissue and organ culture*, *40*, 49-54. https://doi.org/10.1007/BF00041118

Noronha Fernandes-Brum, C., Marinho Rezende, P., Cherubino Ribeiro, T. H., Ricon de Oliveira, R., Cunha de Sousa Cardoso, T., Rodrigues do Amaral, L., ... & Chalfun-Junior, A. (2017). A genome-wide analysis of the RNA-guided silencing pathway in coffee reveals insights into its regulatory mechanisms. *PLoS One*, *12*(4), e0176333. https://doi.org/10.1371/journal.pone.0176333

Park, S. K., & Twell, D. (2001). Novel patterns of ectopic cell plate growth and lipid body distribution in the Arabidopsis gemini pollen1 mutant. *Plant physiology*, *126*(2), 899-909. https://doi.org/10.1104/pp.126.2.899

Pecrix, Y., Rallo, G., Folzer, H., Cigna, M., Gudin, S., & Le Bris, M. (2011). Polyploidization mechanisms: temperature environment can induce diploid gamete formation in Rosa sp. *Journal of experimental botany*, *62*(10), 3587-3597. https://doi.org/10.1093/jxb/err052

Pokhrel, S., Huang, K., & Meyers, B. C. (2021). Conserved and non-conserved triggers of 24-nucleotide reproductive phasiRNAs in eudicots. *The Plant Journal*, *107*(5), 1332-1345. https://doi.org/10.1111/tpj.15382

Pradillo, M., & Santos, J. L. (2018). Genes involved in miRNA biogenesis affect meiosis and fertility. *Chromosome Research*, *26*, 233-241. https://doi.org/10.1007/s10577-018-9588-x

Retief, J. D. (1999). Phylogenetic analysis using PHYLIP. *Bioinformatics methods and protocols*, 243-258. https://doi.org/10.1385/1-59259-192-2:243

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*, *26*(1), 139-140. https://doi.org/10.1093/bioinformatics/btp616

Rozewicki, J., Li, S., Amada, K. M., Standley, D. M., & Katoh, K. (2019). MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic acids research*, *47*(W1), W5-W10. https://doi.org/10.1093/nar/gkz342

De Campos Rume, G., de Oliveira, R. R., Ribeiro, T. H. C., & Chalfun-Júnior, A. (2023). Genome-wide and expression analyses of MADS-box genes in the tetraploid Coffea arabica L. and its diploid parental subgenomes. *Plant Gene*, *34*, 100413. https://doi.org/10.1016/j.plgene.2023.100413

Sato, K., Hamada, M., Asai, K., & Mituyama, T. (2009). CENTROIDFOLD: a web server for RNA secondary structure prediction. *Nucleic acids research*, *37*(suppl_2), W277-W280. https://doi.org/10.1093/nar/gkz342

Sela, I., Ashkenazy, H., Katoh, K., & Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic acids research*, *43*(W1), W7-W14. https://doi.org/10.1093/nar/gkv318

Sharma, D. K., Andersen, S. B., Ottosen, C. O., & Rosenqvist, E. (2015). Wheat cultivars selected for high Fv/Fm under heat stress maintain high photosynthesis, total chlorophyll, stomatal conductance, transpiration and dry matter. *Physiologia plantarum*, *153*(2), 284-298. https://doi.org/10.1111/ppl.12245

UniProt Consortium. (2021). Bateman A., Martin M. *J., Orchard S., Magrane M., Agivetova R., Ahmad S., Alpi E., Bowler-Barnett EH, Britto R., et al. UniProt: The Universal Protein Knowledgebase in*. https://doi.org/10.1093/nar/gkaa1100

Wang, J., Li, D., Shang, F., & Kang, X. (2017). High temperature-induced production of unreduced pollen and its cytological effects in Populus. *Scientific Reports*, *7*(1), 5281. https://doi.org/10.1038/s41598-017-05661-x

Wang, Z., Chen, M., Yang, H., Hu, Z., Yu, Y., Xu, H., ... & Li, J. (2023). A simple and highly efficient strategy to induce both paternal and maternal haploids through temperature manipulation. *Nature Plants*, *9*(5), 699-705. https://doi.org/10.1038/s41477-023-01389-x

Woodhouse, M. R., Cannon, E. K., Portwood, J. L., Harper, L. C., Gardiner, J. M., Schaeffer, M. L., & Andorf, C. M. (2021). A pan-genomic approach to genome databases using maize as a model system. *BMC plant biology*, *21*, 1-10. https://doi.org/10.1186/s12870-021-03173-5

Xia, R., Chen, C., Pokhrel, S., Ma, W., Huang, K., Patel, P., ... & Meyers, B. C. (2019). 24-nt reproductive phasiRNAs are broadly present in angiosperms.

Nature communications, *10*(1), 627. https://doi.org/10.1038/s41467-019-08543-0

Zhai, J., Zhang, H., Arikit, S., Huang, K., Nan, G. L., Walbot, V., & Meyers, B. C. (2015). Spatiotemporally dynamic, cell-type–dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proceedings of the National Academy of Sciences*, *112*(10), 3146-3151. https://doi.org/10.1073/pnas.1418918112

Zhang, H., Xia, R., Meyers, B. C., & Walbot, V. (2015). Evolution, functions, and mysteries of plant ARGONAUTE proteins. *Current opinion in plant biology*, *27*, 84-90. https://doi.org/10.1016/j.pbi.2015.06.011

Zheng, Y., Wang, Y., Wu, J., Ding, B., & Fei, Z. (2015). A dynamic evolutionary and functional landscape of plant phased small interfering RNAs. *BMC biology*, *13*(1), 1-15. https://doi.org/10.1186/s12915-015-0142-4