CHRISTIANE NORONHA FERNANDES BRUM

RNA-GUIDED SILENCING PATHWAYS IN *Coffea spp.*: GENOME-TRANSCRIPTOME-WIDE ANALYSES

LAVRAS – MG
2017

CHRISTIANE NORONHA FERNANDES BRUM

RNA-GUIDED SILENCING PATHWAYS IN *Coffea spp.*: GENOME-TRANSCRIPTOME-WIDE ANALYSES

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, para a obtenção do título de Doutor.

Professor PhD Antonio Chalfun Junior (UFLA)

Orientador

Professor Dr. Matheus de Souza Gomes (UFU)

Coorientador

LAVRAS – MG
2017

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Fernandes-Brum, Christiane Noronha.
RNA-guided silencing pathways in *Coffea spp.* : genome-transcriptome-wide analyses / Christiane Noronha Fernandes-Brum. - 2017.
162 p. : il.

Orientador(a): Antonio Chalfun Júnior.
Coorientador(a): Matheus de Souza Gomes.
Tese (doutorado) - Universidade Federal de Lavras, 2017.
Bibliografia.

1. café. 2. RNAseq. 3. microRNA. I. Júnior, Antonio Chalfun. II. Gomes, Matheus de Souza. III. Título.

CHRISTIANE NORONHA FERNANDES BRUM

RNA-GUIDED SILENCING PATHWAYS IN *Coffea spp.*: GENOME-TRANSCRIPTOME-WIDE ANALYSES

VIAS DE SILENCIAMENTO GUIADAS POR RNA EM *Coffea spp.*: AMPLAS ANÁLISES GENÔMICAS-TRANSCRIPTÔMICAS

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, para a obtenção do título de Doutor.

Aprovada em 24 de Fevereiro de 2017

Dr. Raphael Ricon de Oliveira

PhD Sandra Marisa Mathioni (Donald Danforth Plant Science Center)

Dra. Terezinha Aparecida Teixeira (UFU)

Professor PhD Antonio Chalfun Junior (UFLA)

Orientador

Professor Dr. Matheus de Souza Gomes (UFU)

Coorientador

LAVRAS - MG
2017

A Deus, a meus pais José Antonio e Vanilce, às minhas irmãs Juliana e a Amanda, aos meus sobrinhos Manuela e Heitor, e ao meu marido Thiago,

Dedico

AGRADECIMENTOS

Agradeço primeiramente a Deus pela vida e pela oportunidade de estudá-la. Agradeço aos meus pais pelo apoio e incentivo aos estudos e pelo exemplo de vida. Às minhas irmãs e minhas melhores amigas, Juliana e Amanda, pelo ombro amigo e todo amor (e pelos sobrinhos). Ao meu marido Thiago pelo companheirismo, apoio, suporte, amor e paciência durante todos esses anos juntos. Obrigada por acreditarem em mim mais do que eu mesma.

Agradeço aos meus orientadores, Antonio Chalfun Junior e Matheus de Souza Gomes, pela confiança, pelos ensinamentos e pela oportunidade de trabalharmos juntos. Obrigada pela liberdade e abertura com que conduzem a orientação.

Agradeço aos meus amigos do Laboratório de Fisiologia Molecular de Plantas, mesmo aqueles que não estão mais no grupo, por poder trabalhar e aprender com cada um. Agradeço a todos que me ajudaram em todas as etapas desse projeto. Agradeço ao Cecílio por me ajudar a pensar no projeto, me acompanhar nas análises de campo e pelas conversas regadas a muito café. Agradeço ao Bruno e demais estagiários pelas coletas e análises em campo, sob chuva ou sol, finais de semana ou madrugada. Agradeço imensamente à Pâmela por todo o trabalho e dedicação nas análises computacionais, assim como o Thales, que nos ajudou muito na finalização das análises. Agradeço também à Andressa, Gabriel, Matheus Daúde, Kauanne, José Diogo e Luiza por ajudar e muito na mineração dos dados e/ou nas extrações de RNA. Agradeço à Thaís Cardoso pela colaboração nas análises dos miRNAs.

Agradeço ao INCT-café pelo apoio financeiro para o sequenciamento das bibliotecas, e ao Prof. Blake Meyers, à Dra Sandra Mathioni e o Dr Atul Kakrana pela imensa colaboração no projeto de RNAseq. Agradeço à UFLA pela

oportunidade de construir minha formação acadêmica dentro dessa grande instituição, ao Programa de Pós-Graduação em Biotecnologia Vegeal, e à CAPES pela concessão da bolsa.

Agradeço também aos amigos da ABU, por poder compartilhar de momentos de comunhão com vocês e aprender a ser um pouco mais parecida com Cristo.

RESUMO

O Brasil é o maior produtor e exportador de café, a segunda commodity mais comercializada no mundo. O café é uma cultura bianual cuja fenologia é fortemente influenciada por fatores ambientais, como fotoperíodo, temperatura e distribuição de chuvas. A planta do cafeeiro apresenta desenvolvimento floral sequencial e vários eventos de florescimento. Esses eventos levam à maturação desigual dos frutos e, portanto, a uma baixa qualidade da bebida e a um aumento nos custos de produção. Dentro do panorama das mudanças climáticas, são extremamente necessários estudos que visem esclarecer os mecanismos de regulação do desenvolvimento para o melhoramento genético. Os microRNAs (miRNAs) são uma classe de RNAs pequenos não codificantes (sRNAs), com 21 nt de comprimento, em média, que regulam a expressão de mRNA em plantas inibindo a sua tradução com ou sem clivagem do alvo. O objetivo desta tese foi fornecer uma ampla análise de miRNAs de café e desvendar as vias de silenciamento guiadas por RNA nesta cultura. Para isso, o genoma de *C. canephora* foi pesquisado para a identificação e caracterização dos miRNAs e dos componentes das vias de silenciamento de RNA. Além disso, foram construídas bibliotecas de RNAseq pequenos RNAs a partir de gemas florais de *C. arabica* em diferentes estádios de desenvolvimento, visando identificar genes miRNAs conservados e de gênero específico relacionados ao desenvolvimento floral no café. Estes dois estudos representam um trabalho pioneiro que oferece um passo significativo para uma melhor compreensão da regulação transcricional e pós-transcricional desta grande cultura. A identificação e caracterização dos componentes das vias de silenciamento guiadas por RNA nesta importante cultura não somente proporcionam conhecimento da biologia destas plantas, mas também fornecem base para o melhoramento vegetal através de ferramentas biotecnológicas para lidar com as restrições dessa cultura.

Palavras-chave: café, *C. arabica*, *C. canephora*, RNAseq, smallRNA, microRNA

ABSTRACT

Brazil is the major producer and exporter of coffee beans, which are the second most traded commodity worldwide. Coffee is a biannual crop whose phenology is strongly influenced by environmental factors, such as photoperiod, temperature and rainfall distribution. The coffee tree presents sequential flower development and several blossoming events. These events lead to unequal maturation of fruits and, therefore, impair the quality of the beverage and increase the production costs. Within the panorama of climate changes, studies aiming to clarify the regulation of developmental mechanisms for crop breeding are extremely required. microRNAs (miRNAs) are a class of non-coding small RNAs (sRNAs), 21 nt long on average, that regulate mRNA expression in plants inhibiting their translation with or without cleavage of the target. The aim of this dissertation was to provide a wide analysis of coffee miRNAs and to unravel the RNA-guided silencing pathways. To do that, the *Coffea canephora* genome was surveyed for the identification and characterization of the miRNAs and the components of RNA-silencing pathways. Furthermore, small-RNAseq libraries were constructed from floral buds in different development stages, aiming to identify conserved and genus-specific miRNAs related to floral development in coffee. These two studies represent a pioneering work that offers a significant step towards a better understanding of the transcriptional and post-transcriptional regulation of floral development in this major crop. The identification and characterization of the components in the RNA-guided silencing pathways in this important crop provide not only knowledge of the plant biology, but also basis for further enhancement through biotechnological tools to address its constraints.

Keywords: coffee, *C. arabica*, *C. canephora*, RNAseq, smallRNA, microRNA

SUMÁRIO

PART 1 11
1. INTRODUCTION 12
2. LITERATURE REVIEW 14
2.1 Coffee economic aspects and botany 14
2.2 *Coffea arabica* phenology 17
2.3 *Coffea arabica* cultivars: productivity vs environment 19
2.4 *Coffea arabica* flowering 22
2.5 Molecular regulation of flowering: an overview 25
2.6 microRNAs in flowering 30
2.7 Small RNAs biogenesis and processing 32
2.8 miRNAs in coffee: state of the art 39
Hypothesis 41
Aims 41
General aims: 41
Specific aims: 41
3. CONCLUSION 43
3.1 General Conclusions 43
3.2 Future Perspectives 45
REFERENCES 46
PART 2: Articles 60
ARTICLE 1 - A genome-wide analysis of the RNA-guided silencing pathway in coffee reveals insights into its regulatory mechanisms 61
ARTICLE 2 - Characterization and Profiling of the small RNAs transcriptome of two phases of flowering from two cultivars of Coffea arabica 108

PART 1

1. INTRODUCTION

Brazil is the major producer and exporter of coffee, the second most traded commodity worldwide (DAMATTA et al., 2007). *Coffea canephora* (Robusta) and *Coffea arabica* (Arabica) are the only economically important species of coffee and are produced and consumed worldwide (ICO, 2016a;2016b). The Brazilian production in 2016 was of approximately 52 mi 60kg bags, 41.29 mi of these corresponding to *C. arabica* (CONAB, 2016). The estimated production for coffee (Arabica and Robusta) in 2017 is of 47 mi bags (CONAB, 2016).

Coffee is a biannual crop whose phenology is strongly influenced by environmental factors, such as photoperiod, temperature and rainfall distribution. The coffee tree presents sequential flower development and several blossoming events. These events lead to unequal maturation of fruits and, therefore, to a low quality of the beverage and an increase on production costs. It is known that a water deficit period followed by rainfall or irrigation is required for blossoming (CAMARGO, A., 1985). Accordingly, a main blossoming event may be induced after a slightly stronger water deficit, mitigating unequal maturation (RENA; MAESTRI, 1985), however, the relationship between water deficit and flowering remains unclear. Moreover, severe drought and heat periods can damage the plants during the reproductive development (CAMARGO, A.; CAMARGO, 2001).

Global warming and other climate change events are expected to occur in the coming years (IPCC, 2014), expecting increased temperature annual rates and longer periods of drought. These predictions raises worries regarding crops production, including the coffee production (BUNN et al., 2015; CAMARGO, M., 2010; DAVIS et al., 2012), since the coffee tree phenology is highly influenced by the weather (DAMATTA et al., 2007).

Within this panorama, studies aiming to clarify the regulation of developmental regulation mechanisms for crop breeding are extremely required. In this way, the study of regulatory molecules, such as the microRNAs (miRNAs), can provide insights into the coffee species for further development of biotechnological tools.

miRNAs are a class of non-coding small-RNAs (sRNAs), 21 nt long on average, that regulate mRNA expression in plants inhibiting their translation with or without cleavage of the target (VAUCHERET, 2006). These molecules were only recently discovered, and are involved in the control of several regulatory networks, including flowering and drought tolerance in plants (DING; TAO; ZHU, 2013; YAMAGUCHI; ABE, 2012).

Due to the high conservation of miRNAs sequence among species, miRNAs have been identified in an increasing number of species, by computational tools for identification based on homology with other species (CHAVES et al., 2015; DE SOUSA CARDOSO et al., 2016; HUANG et al., 2014; SUN et al., 2014). Furthermore, next generation sequencing analyses have been crucial for identification of conserved and lineage-specific miRNAs (WANG, F. et al., 2013; WANG, T. et al., 2011).

Although some efforts have been made to identify coffee miRNAs from transcripts, genomic sequences and from the recently released Coffee Genome v1.0 (*C. canephora*) (AKTER et al., 2014; CHAVES et al., 2015; DENOEUD et al., 2014; LOSS-MORAIS et al., 2014; REBIJITH et al., 2013), no specific work for identification and characterization of miRNAs in the Coffee Genome has been conducted yet. Therefore, the number of miRNAs identified so far in coffee might have been underestimated. Furthermore, there are no previous works concerning the identification of the RNA-guided silencing pathway components in coffee. Additionally, small-RNAseq has never been made in *C. arabica* and, therefore,

differential expression and genus-specific miRNAs have never been analyzed before in this economically important species.

Therefore, the aim of this thesis was to provide a wide analysis of coffee miRNAs and to unravel the RNA-guided silencing pathways, intending to pave the way towards coffee breeding by making available new molecular biotechnological tools. To this end, the *C. canephora* genome was surveyed for the identification and characterization of the miRNAs and components of the RNA-silencing pathways components (Article 1). Furthermore, small-RNAseq libraries were constructed from floral buds in different development stages, aiming at to identify conserved and genus-specific miRNAs related to floral development in coffee (Article 2).

2. LITERATURE REVIEW

2.1 Coffee economic aspects and botany

Coffee is one of the most important crops in the world and the second worldwide most traded commodity, representing an important source of income in several countries. Brazil is the largest consumer and the world's largest coffee producer, responsible for 36% of coffee production in 2016 (ICO, 2016a). The country is also the main exporter, with 32% of all coffee coming from Brazil, totalizing 34 mi bags exported (ICO, 2016b). Only two coffee species are commercially important, *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta), representing 64% and 36% of the world's coffee production, respectively (ICO, 2016b). More than 84% of the Brazilian coffee production is from Arabica. The Brazilian production in 2016 was of around 52 mi 60 kg bags, 41.29 mi of these corresponding to *C. arabica* (CONAB, 2016). The estimated production for coffee (Arabica and Robusta) in 2017 is of 47 mi bags (CONAB, 2016).

The coffee trees belong to the *Rubiacea* family, in the genus *Coffea*, section *Eucoffea*, in which are comprised the main economically important species: *Coffea arabica* and *Coffea canephora*, among 103 species (GUERREIRO-FILHO et al., 2008). The *C. arabica* species originated in Africa, in the southwest of Ethiopia, southeast of Sudan and north of Kenya, over 1000 and 2000 m of altitude and annual temperature around 18 to 20ºC (GUERREIRO-FILHO et al., 2008). Worldwide dispersion was through Iemen and it arrived in Brazil during colonization, in 1727 (GUERREIRO-FILHO et al., 2008). *C. arabica* is an autogamous species (with frequent alogamy) and allotetraploid (2n=4x=44 chromosomes), probably originated from hybridization of non-reduced gametes of two diploid species, *C. eugenioides* and *C. canephora* (LASHERMES et al., 1999). Although a very strict genetic basis (CARVALHO, 1993), *C. arabica* presents highly variable cultivars due to interbreeding, mutations, agronomic practices and environmental effects (ALVES, 2008). *C. canephora* is a diploid species (2n=2x=22 chromosomes) and alogamous. This species originates from a broad region of hot, humid and low lands, that extended from Guinea-Bissau to Congo, from the west coast to the central region of the African continent, at low altitude lands and temperatures around 22 to 26ºC (GUERREIRO-FILHO et al., 2008).

The coffee trees are perennial shrubs of constant growth that can reach up to 4 meters, with branched dimorphisms regarding the growth pattern (ALVES, 2008). The branches that grow in the upright direction are called orthotropic, and the laterally grown branches are called plagiotropic. The growth pattern of these two types of branches confer a cylindrical shape to the coffee tree (ALVES, 2008), with an orthotropic branch as the main central branch and the plagiotropic branches originating from the denominated ‘head-of-series’ buds, because each head of series bud is above a group of 5-6 vegetative buds (serial buds) in the orthotropic branch leaves axils (ALVES, 2008). Secondary and tertiary

plagiotropic branches can originate from the primary plagiotropic branches (DO LIVRAMENTO, 2010).

The leaves are formed in the plagiotropic branches, in pairs and opposite to each other. The leaves are 12-24 cm long, lanceolate, or lance-shaped, and very dark green on the upper surface, but much lighter underneath (DO LIVRAMENTO, 2010). The coffee flowers are located in axillary groups, in a variable number of 2-19 per axil (ALVES, 2008; DO LIVRAMENTO, 2010). The inflorescences originate from buds disposed, generally, 4-5 per axil, in descendant series in the axil formed by the leaves with the branch. Each bud of the series develops in a short axis, that will result in an inflorescence with four flowers each (ALVES, 2008). Bracteoles are in number of two pairs for each flower set, being respectively lanceolate and triangular. The flower calyx is very rudimentary; the sepals resemble the leaves anatomy; the petals are united in a tube forming a salver-shaped corolla. Stamens are epipetalous; the anthers are two-celled, opening lengthwise. Pistil is represented by an inferior ovary, terminal style and two stigmatic branches (DEDECCA, 1957).

The coffee fruit is a drupe, containing normally two seeds. By abortion of an ovule one-seeded fruit may be formed (peaberry or Mocca). Ripe fruits have a fleshy and thick pericarp. Fruit development brings about a series of chemical and morphological modifications that lead to a reduction in the thickness of pericarp, from 1.5 mm in ripe fruits to 0.4 in dried fruits. Exocarp is represented by a single layer of hardened and lignified cells, with scattered stomata. Mesocarp is formed by lignified cells, the innermost of which are somewhat compressed and flattened. Endocarp constitutes in the ripe fruits the so-called "seed parchment" (DEDECCA, 1957).

The coffee seeds, popularly called coffee beans are elliptical or egg-shaped, plane-convex, possessing a longitudinal furrow on the plane surface. Seed coat is represented by the so-called "silver skin". The endosperm tissue seems to

present differences in the structure and chemical composition of its various layers. The endosperm contains water, protein, the alkaloids caffeine and coffearine, oil, sugar, dextrins, pentosans, cellulose, caffetannic acids, minerals, various acids and other minor constituents. The small embryo, localized at the bottom of the seed, on its convex surface is represented by an hypocotyl and two adherent cordiform cotyledons (DEDECCA, 1957).

2.2 *Coffea arabica* phenology

The complete phenological cycle of a coffee tree consists in a sequence of vegetative and reproductive phases that takes two years to be completed due to the specificities of growth and developments of this culture (Figure 1). According to Camargo and Camargo (2001), the phenological cycle of Arabica coffee under Brazilian climatic conditions totalizes six phases, from the vegetative stage to the fruits maturation and senescence of the branches (CAMARGO, A.; CAMARGO, 2001). The six phenological phases are: (1) vegetative growth; (2) flower bud induction and maturation; (3) Anthesis and fruit growth; (4) fruit filling (5) fruit ripening; (6) senescence (self-pruning). The series of events and time of the year of occurrence are summarized in Figure 1.

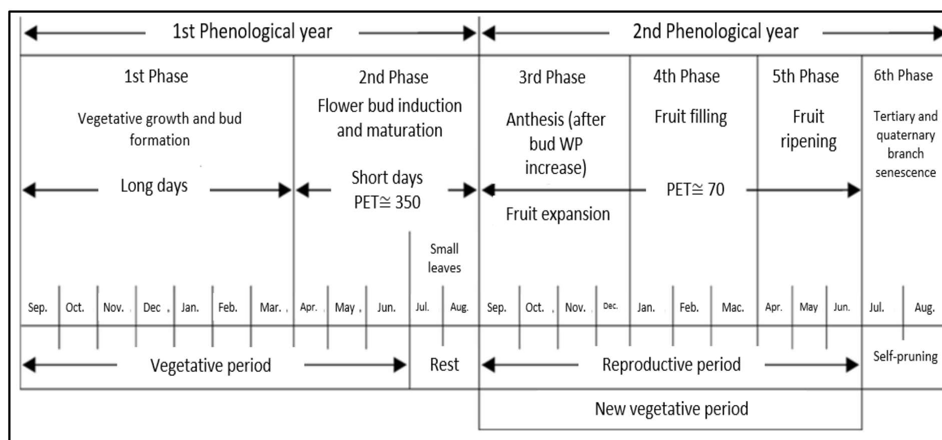
The first phenological year is comprised of vegetative growth (1st phase) in the months of September to March, characterized by long days (CAMARGO, A.; CAMARGO, 2001). This phase consists in the formation and development of nodes with axillary buds that will form the lateral branches for the next year's fruit production. This phase prepares physiologically prepares the plant to for the next year's production (CAMARGO, A.; CAMARGO, 2001).

Still in the first phenological year, occurs the flower bud induction and maturation (2nd phase) in the months of April to August, characterized by short days (CAMARGO, A.; CAMARGO, 2001). By the end of the second phase the plants enter a period of relative rest (dormancy), preparatory for the blossoming

as the buds become sensitive to the stimulus to regrow (DAMATTA et al., 2007). A moderated water deficit is needed in this phase, and the return of water availability provide the signal to regrow towards the anthesis (CAMARGO, A.; FRANCO, 1985).

The anthesis (3rd phase) defines the beginning of the second phenological year (CAMARGO, A.; CAMARGO, 2001; MORAIS et al., 2008; PEZZOPANE et al., 2003). The anthesis occurs 8-12 days after the "hydric shock" in the floral buds that were *ripe-to-flower* (CAMARGO, A.; FRANCO, 1985; DAMATTA et al., 2007). The 3rd phase in the second phenological year also comprises the beginning of fruit development and expansion, from September to December. The fourth phase, fruit filling, occurs in the summer period (January-March), when the grains are formed. Following, the fruit ripening occurs in the fifth phase, from April to June. The sixth and last phase comprises senescence of non-primary reproductive branches, occurring a self-pruning. It is worth notice that new vegetative branches are formed starting a new phenological year simultaneously to the reproductive period of the branches in the second phenological year (Figure 1).

Figure 1 – Representation of the six phenological phases between the first and second phenological years of Arabica coffee under Brazilian tropical climate conditions. PET stands for Potential Evapotranspiration.



Adapted from Camargo & Camargo (2001)

2.3 *Coffea arabica* cultivars: productivity vs environment

In addition to the hydric regime, the coffee tree development is also influenced by temperature. The coffee trees require mild temperatures in cultivation, with ideal average annual temperature of 18-23°C (CAMARGO, M., 2010). Under specific conditions of cultivation and irrigation some cultivars support averages of 24-25°C, such as the cultivars grown in the brazilian North and Northeast regions (CAMARGO, M., 2010). Extreme weather fluctuations, such as high temperatures and more frequent events of drought during the flowering period can affect the production with the abortion of flowers (CAMARGO, A.; CAMARGO, 2001).

With the perspective of climate changes in the recent years (IPCC, 2014) the coffee cultivation would be seriously affected by the annual mean temperature rise and irregular rainfall and drought events, which can be longer and/or more frequent (CAMARGO, M., 2010). Losses by around 50% in global area suitability for *C. arabica* and *C. canephora* production were predicted across different

scenarios of climate changes until 2050 (BUNN et al., 2015). Therefore, cultivars that can tolerate these adverse conditions are required. There are 130 arabica cultivars registered in the National Cultivars Registration Office (RNC, 2016). The coffee cultivars present distinct morphological characteristics, such as height and ripped fruit colors, along with variable physiological aspects, such as fruit maturation cycle (early or late), pathogen resistance, drought tolerance, productivity, among others. There are more suitable cultivars according to the region requirements, such as altitude, mean annual temperature, hydric regime, soil conditions, diseases occurrence, and others. Productivity is highly corelated to ideal growth conditions (DE CARVALHO et al., 2008) and, therefore, the coffee breeding programs shall consider the climate changes prospects.

The ‘Siriema’ cultivars are originated from the breeding of *C. racemosa* and the *C. arabica* cultivar ‘Blue Mountain’ to acquire resistance to the Leaf Miner, followed by two natural inbreeding events with *C. arabica* cultivar ‘Mundo Novo’ in the decade of 1970, and subsequent breeding with the cultivar ‘Catimor UFV 842’ (*C. canephora* background) to incorporate resistance to coffee rust. During the selection period in regions with low rainfall index, the plants showed drought tolerance (DE CARVALHO et al., 2008). The ‘Siriema’ plants are of low size, very early maturation cycle and beverage of good quality (DE CARVALHO et al., 2008). Two different selection strategies were chosen: a) to develop cultivars by crossing (seeds propagation); b) Perform cloning of good quality matrices (vegetative propagation). An F7 generation of the plants generated by seeds was commercially released in 2014 named as ‘Siriema AS1’ (MATIELLO et al., 2014; MATIELLO et al., 2015). The vegetative-propagated cultivar, ‘Siriema VC4’, is a more recently commercially released cultivar, in 2015 (MATIELLO et al., 2015).

The ‘Siriema VC4’ coffee plants showed differential response to water deficit when compared to the cultivar ‘Yellow Catuaí IAC 74’ when submitted to

up to 20 days of water deprivation (GRISI, 2006). The stomata closure to reduce water loss is observed in the 'Siriema VC4' plants, and these plants showed recovery after re-irrigation, even after reaching leaf water potentials as low as -3,5MPa (GRISI, 2006; GRISI et al., 2008). The 'Catuaí' plants present late stomata closure and are unable to recover the hydric status after 14 days under water deprivation, with the same water potential presented by the 'Siriema VC4' plants (-3.5 MPa). 'Siriema VC4' seedlings undertake up to 24 days without irrigation, to water potentials as low as -3.0 MPa, and are still able to recover after re-irrigation (FERNANDES-BRUM et al., 2013). The osmotic adjustment has been indicated as the main mechanism of tolerance displayed by the 'Siriema VC4' plants to cope with longer periods of drought (FERNANDES-BRUM et al., 2013; MELO et al., 2014), but further investigation is required.

'Catuaí' is a well established group of cultivars, being one of the main produced cultivars in Brazil (AGUIAR et al., 2004). The 'Catuaí' cultivar group was obtained from artificial crossing between productivity-selected coffee plants of 'Yellow Caturra (IAC 476-11)' and Mundo Novo (IAC 374-19) in 1949, and was commercially released in 1972. This cultivar presents good yeld and excellent beverage quality (DE CARVALHO et al., 2008). The small size of the plants allows denser planting, making harvesting and handling easier and cheaper (DE CARVALHO et al., 2008). There are several 'Catuaí' cultivars registered under different suffix numbers, such is the case of Red Catuaí IAC 144, which has presented good yield in most of the growing regions. This variety produces red-colored mature fruits of late maturation cycle (AGUIAR et al., 2004). The plants of the cultivars 'Catuaí' are susceptible to coffee rust and other diseases (DE CARVALHO et al., 2008).

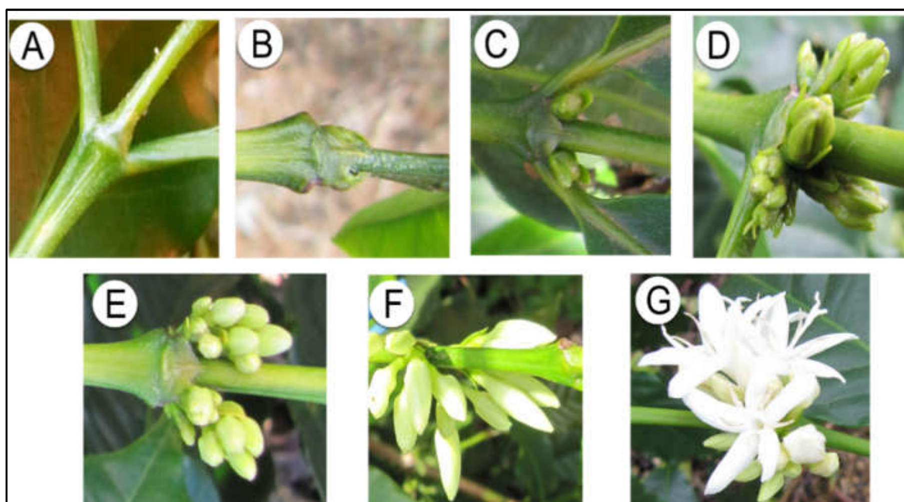
2.4 *Coffea arabica* flowering

The flowering period comprises the sequence of events since induction and floral differentiation until the anthesis (DO LIVRAMENTO, 2010). Every stage in floral development involves physiological, anatomical and morphological modifications, starting from the induction of buds from vegetative to reproductive stage with the inductive stimulus (MORAIS et al., 2008). This inductive stimulus occur in the months with short days, in the beginning of April in the Brazilian conditions (CAMARGO, A.; FRANCO, 1985). These series of events that culminates with flowering are affected by several factors such as temperature, light, soil and air water availability, carbon-to-nitrogen ratio, crop load and genotype (DAMATTA et al., 2007).

The transition of the vegetative phase to the reproductive phase starts with the induction of the four buds in the leaf axils of the plagiotropic branches, the vegetative meristem (VM), into inflorescence meristems (IM) (MAJEROWICZ; SÖNDAHL, 2005; MOENS, 1963). A quick bud growth is observed in this stage, as well as the second pair of bracts development and initiation of a mucilage secretion by the colleters (OLIVEIRA et al., 2014). The mucilage secretion is thought to protect the buds against dehydration (MAYER; CARMELLO-GUERREIRO; MAZZAFERA, 2013). The production of four floral meristems (FMs) by the IM of each bud marks the beginning of the next stage (OLIVEIRA et al., 2014). Afterwards, differentiation of the floral organs occurs (OLIVEIRA et al., 2014). In *C. arabica*, this process is centripetal, which menas that cell divisions initiate in the periphery of the FM, which first generates the sepals, followed by the sequential differentiation of petals, stamens and carpels (OLIVEIRA et al., 2014). The flower development is finished with the formation of the inferior bilocular ovary (OLIVEIRA et al., 2014).

The main stages of coffee reproductive development were designated by MORAIS et al. (2008), and the floral development stages were named as G for buds and FL for flower (Figure 2). The G phases are: G1 – undifferentiated vegetative nodes; G2 – intumesced nodes; G3 – buds with up to 3mm; G4 – buds between 3,1 and 6mm; G5 – buds with 6,1 to 10mm (light green color); G6 – buds greater than 10mm (white color).

Figure 2 – Coffee flower development according to Morais et al. 2008



Legend: (A) Non-differentiated buds (G1 stage). (B) intumesced nodes (G2 stage). (C) Flower buds with up to 3 mm in length (G3 stage). (D) Flower buds ranging from 3.1 to 6 mm in length (G4) stage. (E) Flower buds ranging from 6.1 to 10 mm in length (light green color) (G5 stage). (F) Flower buds bigger than 10 mm in length (white color) (G6 stage). (G) Anthesis (FL stage). Adapted from LIMA, 2015

After the G6 phase, the anthesis (FL) is observed in the first hours of the day, flowers start to wilt on the next day, and fall on the third day after opening. At the G4 stage all floral whorls have been formed and flower differentiation is complete (MAJEROWICZ; SÖNDAHL, 2005; OLIVEIRA et al., 2014), and at this stage the buds stop growing, entering a dormancy period, in the months of June-August (CAMARGO, A.; CAMARGO, 2001). This is the winter time in

Brazil, when there is low rainfall incidence and low temperatures. A relatively short period of a severe water deficit, as long as predawn leaf water potential is below -0.8 MPa water deficit is required to induce coffee anthesis (CRISOSTO; GRANTZ; MEINZER, 1992). At this point, the buds acquire a sensitiveness to the stimulus to regrow, the *ripe-to-flower* stage (DAMATTA et al., 2007). The increase of water potential is considered a determinant factor for breaking the buds dormancy, the so called hydric shock, which occurs in September when either rainfall returns or with artificial irrigation (CAMARGO, A., 1985). Afterwards, the elongation of the buds is observed, allowing flowering to occur (OLIVEIRA et al., 2014). Anthesis is observed 8-12 days after rainfall or re-irrigation (CAMARGO, A.; FRANCO, 1985). In general, precipitations between 5 to 10 mm are considered to be sufficient for triggering anthesis (DE CARVALHO, 2008).

Sequential flowering is a common problem in coffee, which leads to unequal maturation of fruits within the same plant, interfering on beverage quality and increasing the costs due to several harvesting events (DAMATTA et al., 2007). Throughout the floral buds development period it is possible to find buds in different stages within the same inflorescence, among inflorescences in the same node, among different nodes within the same branch, and also among branches and among plants (QUEIROZ VOLTAN; IRINEU FAHL; CARVALHO CARELLI, 2011). Due to that, fruits in different maturation stages (green, yellow-green, cherry, raisin and dry fruits) are observed in the same node and among branches and plants. In fact, the asynchronous development is observed in vegetative buds, even before flower induction (MAJEROWICZ; SÖNDAHL, 2005; OLIVEIRA et al., 2014).

It is noticed that, when a slightly accentuated water deficit occurs during the floral buds dormancy period, the main flower opening event is very defined after re-irrigation, diminishing unequal development (RENA; MAESTRI, 1985).

It is required to reach a certain stage of development for the hydric shock to have influence over the blossoming, i.e., buds might have secondary growth of the conducting tissues, which allows the response to the return of irrigation (CRISOSTO et al., 1992). It is suggested that buds in earlier stages can reach the ideal stage of development and enter the dormancy point during the water deficit period, increasing the synchrony of anthesis after the return of water availability (DRINNAN; MENZEL, 1994; MES, 1957; REDDY, 1979).

Although there is a clear stimulus to resume growth by the buds after re-irrigation towards anthesis, little is known about what is the basis of the mechanism. LIMA (2015) suggests that ethylene precursors accumulate in the roots and shoots of seedlings under drought conditions, and re-irrigation upregulates the expression of ethylene biosynthesis genes in the shoots. A similar outcome is observed when an ethylene action inhibitor, 1-methylcyclopropene (1-MCP), which is thought to promote an ethylene burst, according to gene expression analysis, suggesting that ethylene may be one of the signals involved in coffee flowering induction upon rain or irrigation. Accordingly, SANTOS (2016) observed upregulation of ethylene biosynthesis genes in leaves and G4-buds after 1-MCP application to field plants, and further anthesis was observed in the absence of irrigation, corroborating to this finding.

2.5 Molecular regulation of flowering: an overview

Floral evocation is characterized by the series of events that occur in the shoot apex and determinate flowers formation by the apical meristems (TAIZ ; ZEIGER, 2016). There are several endogenous and exogenous factors that influence the transition from vegetative to reproductive phase, inducing floral evocation (SPANUDAKIS; JACKSON, 2014). Among exogenous factors are light and temperature (photoperiod and vernalization), and among endogenous factors are gibberellin (GA) hormonal signaling pathway, the autonomous and

plant age pathway (POSE; YANT; SCHMID, 2012). Each one of these cues produces complex signaling networks that crosstalk with each other to form an integrated regulatory network either inducing or repressing other genes, and have been recently reviewed (BECKER; EHLERS, 2016; O'MAOILEIDIGH; GRACIET; WELLMER, 2014; TEOTIA; TANG, 2015; THOMSON et al., 2017).

In a general manner, the signals originated by each of the cues are integrated and lead to the conversion of the shoot apical meristem (SAM) to the inflorescence meristem (IM) and subsequent flowers formation (O'MAOILEIDIGH et al., 2014). The integration of these signals is mediated by the *FLOWERING LOCUS T (FT*) gene and *SUPRESSOR OF OVEREXPRESSION OF CO1 (SOC1)* which, in a complex with FD and AGAMOUS-LIKE 24 (AGL24), respectively, promotes directly or indirectly the expression of the floral meristem identity genes (POSE et al., 2012; YAMAGUCHI; ABE, 2012). The integration of flowering signals is tightly controlled by a repressor complex that consists of two MADS-box transcription factors, *FLOWERING LOCUS C (FLC)* and *SHORT VEGETATIVE PHASE (SVP)* (LIU, C.; THONG; YU, 2009).

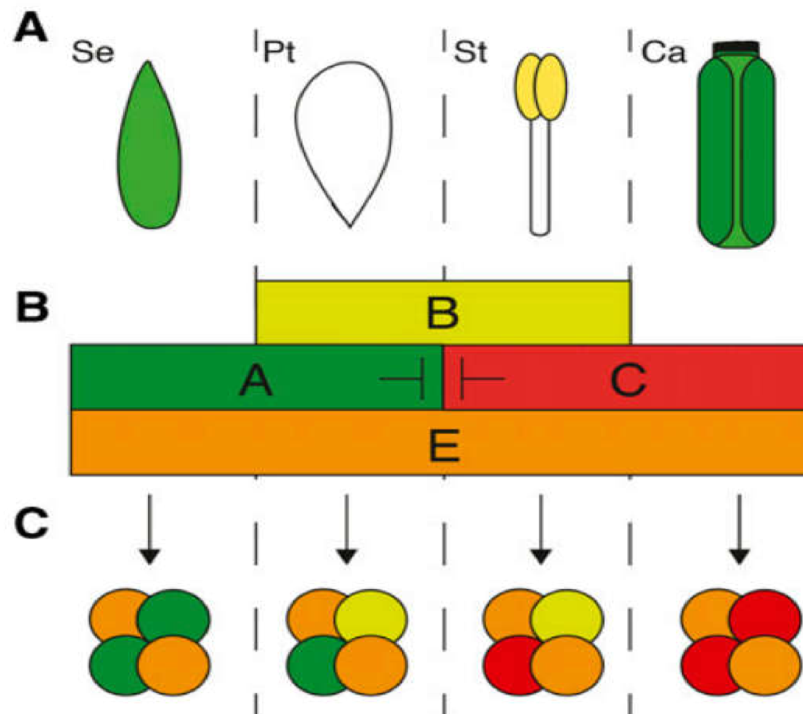
The result of the up or downregulation of the above cited above result in the activation of the floral meristem identity genes, which control the transition to the establishment of Floral Meristems (FM) (LIU, C. et al., 2009; O'MAOILEIDIGH et al., 2014). The FM emerge in the periphery regions of the IM (O'MAOILEIDIGH et al., 2014; REINHARDT; MANDEL; KUHLEMEIER, 2000; REINHARDT et al., 2003). The main floral meristem identity genes are *LEAFY (LFY)* and *APETALA1 (AP1)*, which control the activation of floral organs identity genes (O'MAOILEIDIGH et al., 2014). The main organ identity genes are transcription factors belonging to the MADS-box family (COEN; MEYEROWITZ, 1991). The genes belonging to this family are

involved not only in the floral development, but also flowering time control and fruit development (GU et al., 1998), additionally to other plant tissues such as growth of secondary roots and trichomes (ALVAREZ-BUYLLA et al., 2000).

Floral organ identity is explained by the model ABCE, which in 'ABC' each letter represents a group of genes responsible for the floral whorls identity, and 'E' stands for the genes that co-participate on identity determination and development in all the whorls (Figure 3A and 3B) (COEN; MEYEROWITZ, 1991; KRIZEK; FLETCHER, 2005; THEISSEN; SAEDLER, 2001; THOMSON et al., 2017). In *A. thaliana*, the A function genes, *AP1* and *APETALA2 (AP2)*, specify sepal identity in the first whorl (COEN; MEYEROWITZ, 1991). The combination of the A-function genes with the B-function genes, *APETALA3 (AP3),* and *PISTILLATA (PI)*, is responsible for petals specification in the second whorl. The B class genes combined with the C-function genes, *AGAMOUS (AG)*, specify stamen identity in the third whorl, and C class genes are responsible for carpels identity (COEN; MEYEROWITZ, 1991). The E function genes, *SEPALLATA1-4 (SEP1-4)*, are essential for floral organs identity specification in each whorl with the ABC genes (THEISSEN; SAEDLER, 2001). Except for AP2, all the ABCE-model genes cited above belong to the MADS-box family (COEN; MEYEROWITZ, 1991).

The molecular basis of the ABCE model is explained by the 'quartet model' (THEISSEN; SAEDLER, 2001 - Figure 3C), according to which the identity of the floral organs — sepals, petals, stamens and carpels — is determined by four different combinations of MADS-box proteins (THEISSEN; SAEDLER, 2001).

Figure 3 – Molecular aspects of the ABCE model.



Legend: A, Illustrations of the organs found in the outer to inner (left to right) whorls of the flower. Se, Sepal; Pt, petal; St, stamen; Ca, carpels. B, The ABCE model of flower development (THEISSEN; SAEDLER, 2001). Specific classes of floral organ identity genes are active within each floral whorl. The A class genes specify sepals in the first whorl; The A and B class genes specify petals within the second whorl; The B and C class genes specify stamens within the third whorl; The C class gene function specifies carpel identity within the fourth whorl. The E class genes are active within all four whorls. C, Combinatorial interactions of floral organ identity factors within each whorl form dimeric (not shown) and higher-order tetrameric complexes. Adapted from Thomson et al. (2017).

In coffee, the MADS-box ortholog genes involved in flowering were identified by *in silico* analyses (BARRETO et al., 2011; DE OLIVEIRA et al., 2010) in the coffee Expressed Sequence Tags (ESTs) database of the coffee genome project (CAFEST) (VIEIRA et al., 2006). Furthermore, a thorough identification of 23 coffee MADS-box was conducted and the expression pattern analyses in several vegetative and reproductive tissues were determined of for 18 of them (OLIVEIRA et al., 2014). Peculiar expression patterns of MADS-box family members are thought to be responsible for some coffee specificities regarding floral development. For instance, the expression of *CaAP1* and *CaPI* in bracts and colleters, respectively, along in the inflorescence meristem itself, suggests that these genes could be part of the mechanism triggering the beginning of floral development and acting, directly or indirectly, in the formation of colleters and/or in the activation of mucilage secretion (OLIVEIRA et al., 2014). Furthermore, the lack of the A-function gene *CaAP1* expression in the petal primordium could point toward a stamen-derived origin of petals in coffee and, therefore, connected organs (epipetalous stamen) at the end (OLIVEIRA et al., 2014). Additionally, the expression of flowering repressors, *CaFLC* and *CaSVP-1*, during flower development, specially of *CaFLC* in later stages, suggests involvement in the dormancy of floral buds, another particular feature of coffee flower development (OLIVEIRA et al., 2014).

Vegetative and reproductive bud dormancy is a common feature in perennial tree species native to temperate and boreal regions as a strategy for surviving the cold winter (COOKE; ERIKSSON; JUNTTILA, 2012). The mechanism resembles the vernalization in *Arabidopsis* and cereals. DORMANCY-ASSOCIATED MADS-box factors (DAM) are homologues to the transcription factors SVP and AGL24 in *Arabidopsis* (COOKE et al., 2012). In Arabidopsis, SVP inhibits flowering through negative regulation of FT, while

AGL24 promotes flowering through positive regulation of LFY (HARTMANN et al., 2000; MICHAELS et al., 2003). The expression of DAM-related genes has been found to be upregulated in dormant buds of numerous plant species, such as poplar (ROHDE et al., 2007; RUTTINK et al., 2007), raspberry (MAZZITELLI et al., 2007), leafy spurge (*Euphorbia esula L.*) (HORVATH et al., 2008), Japanese apricot (*Prunus persica*) (ZHONG et al., 2013), Chinese cherry (*Prunus pseudocerasus*) (ZHU et al., 2015), among others. As demonstrated in leaf spurge (*Euphorbia esula L.*), DAM genes are believed to control the expression of *FT* genes (HAO et al., 2015), which are central developmental regulators that have several roles in plant development – including bud dormancy (BOHLENIUS et al., 2006; DANILEVSKAYA et al., 2011; HSU et al., 2011)

The transcriptome of Japanese Pear (*Pyrus pyrifolia* Nakai) flower buds transitioning through endodormancy indicated the involvement of phytohormones in endodormancy release (BAI et al., 2013). For instance, in the ethylene pathway, 1-Aminocyclopropane-1-Carboxylate Synthase (ACS), a gene encoding the rate-limiting enzyme for ethylene biosynthesis, was induced towards endodormancy release (BAI et al., 2013). Furthermore, the expression of DAM genes was down-regulated concomitant with endodormancy release (BAI et al., 2013). Epigenetic regulation such as chromatin remodeling and *de novo* DNA methylation are thought to play crucial roles in the regulation of floral bud dormancy (BAI et al., 2013; RÍOS et al., 2014). Furthermore, sRNAs are upregulated in endodormancy release, indicating a role in this pathway (BAI et al., 2016).

2.6 microRNAs in flowering

Micro RNAs (miRNAs) are involved in the reproductive development of plants (SPANUDAKIS; JACKSON, 2014; YAMAGUCHI; ABE, 2012) either in flowering time, flower development or other associated pathways

(YAMAGUCHI; ABE, 2012). At least nine conserved miRNA families, miR156, miR159, miR160, miR164, miR166/165, miR167, miR169, miR172, and miR319, have been reported to play key roles in flowering, which involves floral transition, floral patterning and floral organ development (LUO; GUO; LI, 2013). For instance, miR156 regulate genes involved in the control of the transition from the vegetative e to the floral phase, *SQUAMOSA PROMOTER BINDING PROTEIN LIKE (SPL)* gene family (ZHANG, T.; WANG; ZHOU, 2015; ZHOU, C.-M.; WANG, 2013). miR156 gradually decreases as plants age, and an increase in SPL promotes flowering through activating *FT*, MADS-box and *LFY* (WANG, J. W.; CZECH; WEIGEL, 2009). Antagonistically, miR172 increases as plants approach flowering, and also accumulates in leaves and floral buds (AUKERMAN; SAKAI, 2003). miR172 acts in the regulatory process of flowering time and also in the determination of floral organ identity, modulating the expression of the *APETALA2-like (AP2)* genes, which are repressors of flowering and A-function floral identity genes (AUKERMAN; SAKAI, 2003; WOLLMANN et al., 2010). miR172 is also activated in the temperature and photoperiod pathways (JUNG et al., 2007; LEE et al., 2010; YAMAGUCHI; ABE, 2012).

miR159 is involved in flowering time in the gibberellin (GA) pathway, and also regulates anther development by the modulation of a GA-specific transcriptional regulator, GAMYB (ACHARD et al., 2004). miR319 also relates to flowering time regulation, regulating TCP transcription factors (SCHOMMER et al., 2012; SPANUDAKIS; JACKSON, 2014). In addition to flowering time, miR159 and miR319 also modulates flower development (RUBIO-SOMOZA; WEIGEL, 2013). These are closely related miRNAs that both modulate miR167, which targets *AUXIN RESPONSIVE FACTOR* family members (*ARF6/ARF8*), and the interaction among these three miRNAs mediates sepal, petal and stamen development (RUBIO-SOMOZA; WEIGEL, 2013).

Recently, studies of the small RNA transcriptome in floral tissues have become more common in flower-associated pathways such as flowering time, flower development, flowers shape, flowers colors, flower opening, dormancy and other aspects (ACETO et al., 2014; BAI et al., 2016; BELLI KULLAN et al., 2015; LI, X. et al., 2015; ROY et al., 2016; WANG, TAO et al., 2014), identifying conserved and novel miRNAs, as well as novel and conserved targets. For instance, analyses have demonstrated that a MADS-box gene of B-function is a target of miR5179 in *Orchis italic*, demonstrating the importance of miRNAs in the diversification of the flower shape in orchids (ACETO et al., 2014). Therefore, the identification of sRNAs involved in flower associated pathways has become an important tool for the comprehension of the complex mechanisms triggering and regulating this developmental process.

2.7 Small RNAs biogenesis and processing

Plant small RNAs (sRNAs) are produced as a result of processing of double stranded duplexes from the helical regions of larger RNA precursors, and are classified according to the intra- or inter-molecular hybridization of the duplex (AXTELL, 2013). microRNAs (miRNAs) are derived from self-complementary hairpin structures, while small-interfering RNAs are derived from double-stranded RNA precursors (dsRNA) or hairpin-derived (BORGES; MARTIENSSEN, 2015; CHEN, 2009).

miRNAs are ~20-24 nucleotides long, usually 21nt, and control gene expression by negative regulation of their target genes, through sequence-specific degradation or translational repression (YAMAGUCHI; ABE, 2012). These molecules participate in several important regulatory processes, such as phase change and reproductive development, response to salinity and drought stresses (ELDEM et al., 2012; GENTILE et al., 2015; SHUAI et al., 2013; SPANUDAKIS; JACKSON, 2014; WANG, TIANZUO et al., 2011; XIE, F. et

al., 2014; YAMAGUCHI; ABE, 2012). Mutant plants for the proteins related to the miRNAs biogenesis present several developmental anomalies, evidencing the importance of miRNAs in regulatory processes in vegetative and reproductive development, in addition to stress response (VAUCHERET et al., 2004).

The miRNAs undergo several modification steps since transcription until the mature stage (~21nt), which can identify their respective target mRNA and guide slicing by the RISC complex. Plant miRNAs are different from other eukaryotes miRNAs (MEYERS et al., 2008; ROGERS, K; CHEN, 2013). The genes responsible for miRNA transcription (MIR) in plants are mostly within intergenic regions (REINHART et al., 2002; ROGERS, K; CHEN, 2013), and are transcribed by RNA Polymerase II (Pol II) (KIM, Y. J. et al., 2011).

The primary transcripts (pri-miRNAs), which size is similar to the protein coding pri-messenger RNAs (pri-mRNAs) (TANG, 2010), undergo capping and polyadenylation (XIE, Z. et al., 2005), and some of them undergo alternative splicing (BIELEWICZ et al., 2013; SCHWAB et al., 2013). Moreover, proteins involved in pri-mRNA splicing are also related to pri-miRNA splicing, including the protein SERRATE (SE) and the cap-binding complex (CBC) subunits CAP-BINDING PROTEIN 80 (CBP80) and CBP20 (KIM, S. et al., 2008; LOBBES et al., 2006). The SICKLE (SIC) protein is also implicated in splicing of pri-miRNAs, and in their absence, plants accumulate unspliced pri-miRNAs (ZHAN et al., 2012).

The pri-miRNA possess a hairpin structure, which is stabilized by the RNA-binding proteins DAWDLE (DDL), facilitating DCL1 to access or recognize pri-miRNAs. (YU et al., 2008). These molecules are then processed by the endonuclease activity of the DICER-LIKE1 (DCL1) (KURIHARA; TAKASHI; WATANABE, 2006) in miRNA precursors (pre-miRNAs), assisted by other proteins in the processing such as the dsRNAbinding proteins HYPONASTIC LEAVES 1 (HYL1) and SERRATE (SE) (DONG; HAN;

FEDOROFF, 2008; KURIHARA et al., 2006). The protein TOUGH (TGH), an RNA-binding protein, is a component of the DCL1-HYL1-SERRATE complex, the dicing body (REN; XIE; et al., 2012). MODIFIER OF SNC1 2 (MOS2) binds pri-miRNA and is involved in efficient processing of pri-miRNAs, but is not part of the dicing body (WU, X. et al., 2013). SIC, which is involved in pri-miRNA splicing, might also play a role in pri-miRNA processing (ZHAN et al., 2012). RECEPTOR FOR ACTIVATED C KINASE 1 (RACK1) and C-TERMINAL DOMAIN PHOSPHATASE-LIKE 1 (CPL1) are interactors with SE and are also implicated in the pri-miRNA processing (JEONG et al., 2013; SPETH et al., 2013). More recently, the protein REGULATOR OF CBF GENE EXPRESSION 3 (RCF3) has been described as a cofactor affecting miRNA biogenesis in specific plant tissues, interacting with CPL1, and also CPL2 (KARLSSON et al., 2015).

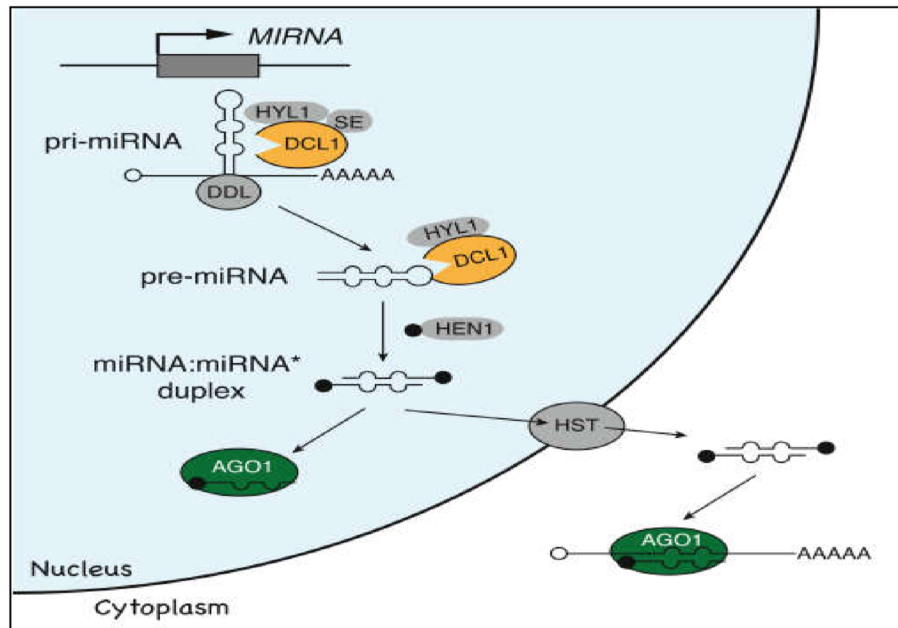
The pre-miRNAs in plants are of variable size, ranging from 49 to 900 nt in length (BOLOGNA; VOINNET, 2014), and undergo processing by DCL1, or, alternatively, by DCL2, DCL3 and DCL4, forming a 21-24 nt duplex structure with two 3' overhangs nucleotides projected laterally (MARGIS et al., 2006). miRNAs are in general 21nt long (DCL1 and DCL4), but the size varies depending on the DCL performing the cleavage, being 22 nt by DCL 2 and 24nt for DCL3 (ROGERS, K; CHEN, 2013). The distance between the RNase III and PAZ domains is suggested to be determinant factor in miRNA length (ROGERS, K; CHEN, 2013).

The duplex undergoes 3' methylation guided by a methyltransferase HUA ENHANCER1 (HEN1), for protection against further modification (LI, J. et al., 2005). 2′-O-methylation of the duplex by HEN1 is crucial for protecting the 3′ terminius from the action of exonucleases, such as small RNA-degrading nuclease (SDN) proteins (RAMACHANDRAN; CHEN, 2008), or 3'-oligouridylation by HESO1 to unmethylated miRNAs, leading to their degradation (REN; CHEN; YU, 2012; TU et al., 2015; ZHAO, Y. et al., 2012). The exportin HASTY (HST)

protein binds the duplex and export it from the nucleus to the cytoplasm (ZENG; CULLEN, 2004), but export in the absence of this protein is also possible, in a mechanism not fully elucidated yet (BOLOGNA; VOINNET, 2014).

In the cytoplasm a strand of the duplex miRNA is chosen and incorporated to an ARGONAUTE (AGO) family protein, containing a PAZ and a PIWI domain, to form the RISC (RNA Induced Silencing Complex) system, assisted by HEAT-SHOCK PROTEIN 90 (HSP90) and SQUINT (SQN) (EARLEY et al., 2010; IKI et al., 2012; IKI et al., 2010). The selection of the guide strand is at least partially dependent upon the thermodynamic stabilities of the 5′ ends (BUDAK; AKPINAR, 2015). Most plant miRNAs possess a 5′ U (uridine), which are usually loaded to AGO1. The ARGONAUTE PIWI domain presents enconuclease activity, being able to cleave the miRNA targets (LIU, J. et al., 2005; ROGERS, K; CHEN, 2013). The miRNA guide the RISC complex to bind its target through sequence complementarity, directing either mRNA cleavage or translational inhibition (ROGERS, K.; CHEN, 2012). In the Figure 4, a general overview of the biogenesis and maturation of the miRNAs in plants is presented.

Figure 4 – General overview of the miRNA biogenesis in Arabidopsis.



Legend: MIR genes are transcribed by Pol II, then pri-miRNAs undergo processing by DCL1 to generate pre-miRNAs, which is further processed to generate the miRNA duplex. The duplex is 3'methylated by HEN1, to prevent further degradation and then exported to the cytoplasm. One of the strands of the duplex is loaded into an AGO protein in the cytoplasm forming the RISC complex. Adapted from Yamaguchi & Abe (2012).

The other major class of sRNAs, siRNAs, can act either at the transcription level, guiding DNA methylation, or at the post-transcriptional level, guiding cleavage and degradation of homologous cellular transcripts (BRODERSEN; VOINNET, 2006; MATZKE, M. A.; MOSHER, 2014). There are several classes of siRNAs, such as: hairpin-derived small-interfering RNAs (hp-siRNAs), imprecisely processed precursor hairpins that do not qualify as miRNAs; heterochromatic siRNAs (hc-RNA), produced mostly from intergenic and/or repetitive regions; secondary siRNAs, which can be phased (phasiRNA) and trans-acting siRNAs (tasiRNA); or natural antisense siRNA (natsiRNA),

which are produced from dsRNAs originating from overlapping transcription (cis-natsiRNA) or highly complementary transcripts originated from different loci (trans-natsiRNA) (AXTELL, 2013; BORGES; MARTIENSSEN, 2015).

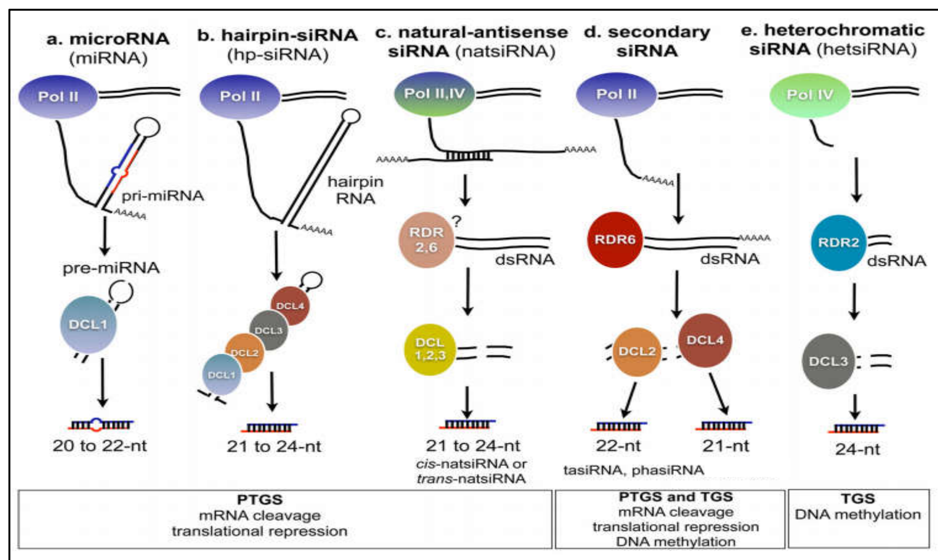
RNA-dependent RNA Polymerases (RDRs) play an important role in siRNA production, synthesizing a second-strand of RNA from RNA template, resulting in the formation of double-stranded RNA (dsRNA) (SCHIEBEL et al., 1993), with initial priming dependent or independent manners (MOISSIARD et al., 2007). The siRNA biogenesis shares the core mechanism with miRNAs, processed by a DCL protein (DCL2, DCL3 and DCL4), methylated by HEN1 and loaded into a member of the AGO family protein (AXTELL, 2013).

miRNA-mediated cleavage of particular target transcripts leads to the formation of dsRNA by RDR6 proteins, which is subsequently processed by DCL4 into secondary siRNAs phasiRNAs, which can act in trans (tasiRNA) targeting several families of genes (ALLEN et al., 2005; FEI; XIA; MEYERS, 2013). phasiRNAs, along with the class of natsiRNAs, which are suggested function mainly at the posttranscriptional level by either cleavage or translational suppression of target transcripts, and in some cases, they can direct DNA methylation (WU, L.; MAO; QI, 2012; WU, L. et al., 2010), but the specific function of phasiRNAs is still unknown.

Additionally, two plant specific DNA-dependent RNA Polymerases, Pol IV e Pol V, are involved in the biogenesis of 24 nt hc-siRNAs, which mediates RNA-dependent DNA Methylation (RdDM) through cytosine methylation (CG, CHG and CHH, where H = A, C or T) by the de novo methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) at target DNA loci (CAO; JACOBSEN, 2002; LAW; JACOBSEN, 2010). Pol IV transcribes heterochromatic regions which will form hc-siRNAs (ONODERA et al., 2005), followed by dsRNA synthesis by RDR2, processing by DCL3 and assembly of the resulting siRNA duplexes in AGO4-clade AGOs (LAW;

JACOBSEN, 2010). Pol V produces transcripts of the intergenic noncoding (IGN) at the loci that will be further methylated and is required for the recruitment of the RdDM machinery, including DRM2 and the AGO loaded with the hc-siRNA (WIERZBICKI, ANDRZEJ T. et al., 2012; ZHOU, M.; LAW, 2015). The recruitment occurs through interaction between protein-protein (Pol V-AGO) and nucleic acids, still unclear if siRNA:IGN or siRNA:DNA (MATZKE, MARJORI A.; KANNO; MATZKE, 2015; WIERZBICKI, A. T. et al., 2009). The biogenesis of the small RNA pathways are represented and summarized in Figure 5.

Figure 5 – Main pathways for biogenesis of endogenous small RNAs in plants



Legend: **a.** Genes encoding microRNAs (miRNAs; left) are transcribed by RNA Polymerase II (Pol II) and fold into hairpin-like structures called primary (pri)-miRNAs, which are processed by DICER-LIKE 1 (DCL1) into a shorter stem-loop structure called precursor (pre)-miRNAs. Pre-miRNAs are processed again by DCL1 into the mature miRNA duplex. During miRNA processing, DCL1 is assisted by several proteins. miRNAs are involved in post-transcriptional gene silencing (PTGS) by mediating mRNA cleavage or translational repression. **b**. Longer Pol II-derived hairpins, termed hairpin-derived small-interfering RNAs (hp-siRNAs; middle), might originate from inverted repeats, and are originally processed by all DCLs. **c**. Natural-antisense small-interfering RNAs (natsiRNA; right) are produced from dsRNAs originating from overlapping transcription (cis-natsiRNA) or highly complementary transcripts originated from

different loci (trans-natsiRNA). **d.** The precursors of secondary siRNAs are transcribed by Pol II, and may originate from non-coding loci, protein-coding genes and transposable elements. These transcripts are converted into double-stranded RNA (dsRNA) by RNA-DEPENDENT RNA POLYMERASE 6 (RDR6), and processed by DCL2 and DCL4 to produce siRNAs of 22- or 21-nucleotide (nt) in length, respectively. Secondary siRNAs are mostly involved in PTGS, but can also initiate RNA-directed DNA methylation (RdDM) at specific loci. They are subdivided into trans-acting siRNAs (tasiRNA) or phased siRNA (phasiRNA). **e.** Heterochromatic siRNAs (hc-siRNAs) are derived from transposable elements and repeats located at pericentromeric chromatin. Their biogenesis requires Pol IV transcription and the synthesis of dsRNA by RDR2, which is subsequently processed into 24-nucleotide long siRNAs by DCL3. These small RNAs are involved in maintaining RdDM-mediated transcriptional gene silencing (TGS). Adapted from Borges & Martienssen (2015).

2.8 miRNAs in coffee: state of the art

miRNA annotation has quickly developed in the past decade, since the creation of miRBase (http://www.mirbase.org/index.shtml), a database that provides access to all published miRNA sequences and guidelines on miRNA annotation (AMBROS et al., 2003; GRIFFITHS-JONES et al., 2006). The intensification of high throughput sequencing technology has increased considerably the annotation of miRNAs, and has also improved the precision of the parameters to annotate a sequence as a miRNA (KOZOMARA; GRIFFITHS-JONES, 2014).

In addition to model plants, miRNAs have been identified in crop species, for instance species used for food (ZHAO, M. et al., 2015), textiles (ZHANG, B. H. et al., 2007) and biofuel (GENTILE et al., 2015). Some previous studies have described miRNAs in coffee. Computer-based strategies have described 16 miRNA families in *C. arabica* (AKTER et al., 2014; REBIJITH et al., 2013). In addition, deep-sequencing libraries have been generated using both genomic (DENOEUD et al., 2014) and transcriptomic (LOSS-MORAIS et al., 2014) samples from *C. canephora,* and identified, in both cases miRNA genes belonging to 33 families.

CHAVES et al. (2015) presented a specific pipeline to search putative coffee miRNAs using ESTs and Genome Survey Sequences (GSS) databases for *C. arabica* and *C. canephora*. As a result, 36 microRNAs belonging to 26 families were identified, expanding the study of miRNAs and their target genes in coffee, and identifying miRNA families that had never been reported in coffee before. Furthermore, a total of 616 and 362 potential targets for *C. arabica* and *C. canephora*, respectively, were predicted. Moreover, the expression profiles of four miRNAs (miR172, miR167, miR171, miR390) were monitored by pulsed stem-loop RT-PCR in the tetraploid species *C. arabica* and the diploid species *C. canephora*. Finally, 5` RACE (Rapid Amplification of cDNA Ends) methodology was used to confirm the regulation of AUXIN RESPONSIVE FACTOR 8 (ARF 8) by miR167 in coffee plants (CHAVES et al., 2015).

Since the specific pipeline applied to EST and GSS (CHAVES et al., 2015) demonstrated to be very robust in the miRNAs identification when compared to previously reported studies (AKTER et al., 2014; REBIJITH et al., 2013), if applied to the *C. canephora* genome v1.0 more precursors of miRNAs could be identified than the initially 92 identified miRNAs (DENOEUD et al., 2014). Furthermore, the main proteins of the miRNA pathway such as DCL-like and AGO-like, and other proteins involved in the biogenesis, processing, function and turnover of sRNAs have never been described in coffee. With the *C. canephora* genome, it has become possible to survey these proteins and perform a thorough computational characterization. The identification and characterization of the RNA-guided silencing pathways components in this important crop will provide not only knowledge of the plant biology, but also basis for further enhancement trough biotechnological tools to address its constraints.

Hypothesis

We hypothesize that the sRNAs pathways are conserved and miRNAs are highly represented in the *C. canephora* genome, as well as genus-specific miRNAs are likely present in *C. canephora* and *C. arabica*. Furthermore, similarly to other plants, miRNAs might play a central role in the regulation of floral development, and, therefore, might be present and differentially expressed during the regulation of floral bud dormancy release towards flowering in *C. arabica* plants.

Aims

General aims:

The aim of this work was to deepen the study of sRNAs in *Coffea* by a thorough analysis in the *C. canephora* genome and in sRNAseq libraries from floral buds in different stages of development in two cultivars of *C. arabica.*

Specific aims:

- Identify the protein components of the sRNA pathways in the *C. canephora* genome, focusing on the miRNA pathway (Article 1);
- Identify and characterize the miRNAs in the *C. canephora* genome (Article 1);
- Identify the putative targets of miRNAs in the *C. canephora* genome (Article 1);
- Build small RNA-seq libraries of buds in different developmental stages - G4 and G5 - of the two cultivars of *C. arabica* ‘Siriema VC4’ and ‘Red Catuaí IAC 144’ (Article 2);

- Identify conserved and non-conserved miRNAs in the sRNAseq libraries (Article 2);
- Identify the putative targets of the miRNAs present in the sRNAseq libraries (Article 2);
- Identify the differentially expressed miRNAs among the stages and among the cultivars (Article 2).

3. CONCLUSION

3.1 General Conclusions

The study of the small RNA silencing pathways has become a hotspot in Genetics and Molecular Biology, and a rapid evolving knowledge has been created regarding the mechanisms and components involved in those pathways. Such RNA-based processes consist in sequence-specific inhibition of gene expression at transcription or post-transcriptional level through the action of small (20-26nt) homologous RNA sequences (BRODERSEN; VOINNET, 2006).

In the first study presented in this dissertation (Article 1 – page 61), a wide analysis of the *C. canephora* genome was presented. The proteins involved in the biosynthesis, function and turnover of sRNAs - eleven AGO, nine DCL-like and eight RDR proteins, as well as other 48 proteins implicated in the sRNA pathways, including HYL1, HST, HEN1, SE and TGH, were identified. Annotation of DCL1-like protein and indication of expansion of the locus related to the sRNA pathways in this species were also showed. Furthermore, validation was provided by expression analysis in RNAseq libraries.

Additionally, the MIR loci in *C. canephora* were investigated by homology based analysis with the MiRBase v.21 and a pipeline for prediction of putative hairpin structures (DE SOUZA GOMES et al., 2011). Several MIR loci previously unannotated by the coffee genome report (DENOEUD et al., 2014) were identified, totalizing 235 miRNA precursors producing 317 mature miRNAs, belonging to 113 MIR families. Characterization of three conserved MIR families (ccp-MIR156, ccp-MIR172, ccp-MIR390) showed overall conservation of structural, evolutionary and targets with their respective

orthologs. Thereafter, the putative miRNAs targets in the genome and showed that the main GO terms of the targeted genes coincide with the main terms of the whole genome, evidencing the importance of the miRNAs in *C. canephora*.

In the second study presented in this dissertation (Article 2 – Page 108) , a sRNA transcriptome was generated from floral buds in two development stages (G4 and G5) from two cultivars of *Coffea arabica*, 'Siriema VC4' and 'Red Catuaí IAC 144'. A total of 155 mature miRNAs were identified, 49 previously known and 106 novel miRNAs, which were considered genus-specific. A total of 211 MIR loci were annotated in the *C. canephora* genome. Differential expression of 17 miRNAs was observed between G4 and G5 stages considering both Siriema and Catuaí cultivars replicates. These miRNAs might play a crucial role in flower development and resume of growth processes. The putative novel miRNAs might provide further insights into gene regulation of flower development processes in *Coffea* species, and further investigation is required.

Furthermore, the secondary structure of the precursors of the differentially expressed novel miRNAs and the putative targets of the differentially expressed miRNAs were predicted in the *C. canephora* genome, and Gene Ontology Enrichment analyses revealed that the predicted targets are involved mainly in regulatory processes.

Put together, these two studies represent a pioneering work that offers a significant step towards a better understanding of the transcriptional and post-transcriptional regulation of flower development in this major crop. The identification and characterization of the RNA-guided silencing pathways components provide not only knowledge of the plant biology, but also provides basis for further enhancement trough biotechnological tools to address this its constraints.

3.2 Future Perspectives

Considering that *C. canephora* is a diploid (2n=2x=22) species and a parental of the tetraploid (2n=4x=44) *C. arabica*, expansions of loci of the proteins related to the small RNA pathways and the MIR loci between these two species are highly expected. This work provides a great basis for the analyses of the soon to be released *Coffea arabica* genome. The results of the sRNA transcriptome can be increased, including the possibility of the discovery of species-specific miRNAs.

Moreover, the generation of a parallel analysis of RNA end (PARE) libraries (ZHAI et al., 2014) to validate cleavage of miRNA targets by the sequencing of 5′ ends of cleaved or uncapped mRNAs, in the same developmental stages (G4 and G5) of *C. arabica,* will provide complementary information to unravel the function of these identified miRNAs in the release of bud dormancy and resume of growth. The transcriptome from protein-coding sequences will also help to characterize the processes occurring during floral bud development, not only the sRNA- dependent, providing a wide overview in the coffee flower development process. These analyses are in progress and will be available soon.

Although previous works have investigated miRNAs in coffee (AKTER et al., 2014; CHAVES et al., 2015; LOSS-MORAIS et al., 2014; REBIJITH et al., 2013), even in the coffee genome report (DENOEUD et al., 2014), such deep and wide analyses were never conduced before in *Coffea* spp.. Therefore, it can be considered that a new investigation area has been established in this work, and several outcomes will be generated based on this diverse set of information.

REFERENCES

ACETO, S. et al. The Analysis of the Inflorescence miRNome of the Orchid *Orchis italica* Reveals a *DEF*-Like MADS-Box Gene as a New miRNA Target. **PLoS ONE,** v. 9, n. 5, p. e97839, 2014.

ACHARD, P. et al. Modulation of floral development by a gibberellin-regulated microRNA. **Development,** v. 131, n. 14, p. 3357-65, Jul 2004.

AGUIAR, A. T. D. E. et al. Caracterização de cultivares de Coffea arabica mediante utilização de descritores mínimos. **Bragantia,** v. 63, p. 179-192, 2004.

AKTER, A. et al. Computational identification of miRNA and targets from expressed sequence tags of coffee (Coffea arabica). **Saudi Journal of Biological Sciences,** v. 21, n. 1, p. 3-12, Jan 2014.

ALLEN, E. et al. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. **Cell,** v. 121, n. 2, p. 207-21, Apr 2005.

ALVAREZ-BUYLLA, E. R. et al. MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. **Plant J,** v. 24, n. 4, p. 457-66, Nov 2000.

ALVES, J. D. Morfologia do Cafeeiro. In: DE CARVALHO, C. H. S. (Ed.). **Cultivares de Café: Origem, características e recomendações**. Brasília: Embrapa, 2008. cap. 2, p.334.

AMBROS, V. et al. A uniform system for microRNA annotation. **Rna,** v. 9, n. 3, p. 277-9, Mar 2003.

AUKERMAN, M. J.; SAKAI, H. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. **Plant Cell,** v. 15, n. 11, p. 2730-41, Nov 2003.

AXTELL, M. J. Classification and comparison of small RNAs from plants. **Annu Rev Plant Biol,** v. 64, n. 1, p. 137-59, Apr 2013.

BAI, S. et al. Small RNA and PARE sequencing in flower bud reveal the involvement of sRNAs in endodormancy release of Japanese pear (Pyrus pyrifolia 'Kosui'). **BMC Genomics,** v. 17, p. 230, Mar 2016.

BAI, S. et al. Transcriptome Analysis of Japanese Pear (Pyrus pyrifolia Nakai) Flower Buds Transitioning Through Endodormancy. **Plant and Cell Physiology,** v. 54, n. 7, p. 1132-1151, Jul 2013.

BARRETO, H. G. et al. In Silico and Quantitative Analyses of the Putative FLC-like Homologue in Coffee (Coffea arabica L.). **Plant Molecular Biology Reporter,** v. 30, n. 1, p. 29-35, Feb 2011.

BECKER, A.; EHLERS, K. Arabidopsis flower development—of protein complexes, targets, and transport. **Protoplasma,** v. 253, n. 2, p. 219-230, 2016.

BELLI KULLAN, J. et al. miRVine: a microRNA expression atlas of grapevine based on small RNA sequencing. **BMC Genomics,** v. 16, n. 1, p. 1-23, 2015.

BIELEWICZ, D. et al. Introns of plant pri-miRNAs enhance miRNA biogenesis. **EMBO Rep,** v. 14, n. 7, p. 622-8, Jul 2013.

BOHLENIUS, H. et al. CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. **Science,** v. 312, n. 5776, p. 1040-3, May 2006.

BOLOGNA, N. G.; VOINNET, O. The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in Arabidopsis. **Annual Review of Plant Biology,** v. 65, n. 1, p. 473-503, Apr 2014.

BORGES, F.; MARTIENSSEN, R. A. The expanding world of small RNAs in plants. **Nat Rev Mol Cell Biol,** v. 16, n. 12, p. 727-41, Dec 2015.

BRODERSEN, P.; VOINNET, O. The diversity of RNA silencing pathways in plants. **Trends in Genetics,** v. 22, n. 5, p. 268-280, May 2006.

BUDAK, H.; AKPINAR, B. A. Plant miRNAs: biogenesis, organization and origins. **Functional & Integrative Genomics,** v. 15, n. 5, p. 523-531, Jan 2015.

BUNN, C. et al. A bitter cup: climate change profile of global production of Arabica and Robusta coffee. **Climatic Change,** v. 129, n. 1, p. 89-101, Mar 2015.

CAMARGO, A. Florescimento e frutificação de Café arabica nas diferentes regiões cafeeiras do Brasil. **Pesquisa Agropecuária Brasileira,** v. 20, n. 7, p. 831-839, Jul 1985.

CAMARGO, A.; CAMARGO, M. B. P. D. Definição e Esquematização das Fases Fenológicas do Cafeeiro Arábica nas Condições Tropicais do Brasil. **Bragantia,** v. 60, n. 1, p. 65-68, 2001.

CAMARGO, A.; FRANCO, C. **Clima e fenologia do cafeeiro**. Ministério da Agricultura e Comércio. Rio de Janeiro, p.21. 1985

CAMARGO, M. The impact of climatic variability and climate change on arabic coffee crop in Brazil. **Bragantia,** v. 69, n. 1, p. 239-247, 2010 2010.

CAO, X.; JACOBSEN, S. E. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. n. 0960-9822 (Print), 20020717 DCOM- 20030204 2002.

CARVALHO, A. **Histórico do desenvolvimento do cultivo de café do Brasil**. IAC. Campinas: IAC 1993.

CHAVES, S. S. et al. New Insights on Coffea miRNAs: Features and Evolutionary Conservation. **Appl Biochem Biotechnol,** v. 177, n. 4, p. 879-908, Oct 2015.

CHEN, X. Small RNAs and Their Roles in Plant Development. **Annual Review of Cell and Developmental Biology,** v. 25, n. 1, p. 21-44, Nov 2009.

COEN, E. S.; MEYEROWITZ, E. M. The war of the whorls: genetic interactions controlling flower development. **Nature,** v. 353, n. 6339, p. 31-37, May 1991.

CONAB. **Acompanhamento da Safra Brasileira, Café**. 3**:** 103 p. 2016.

COOKE, J. E. K.; ERIKSSON, M. E.; JUNTTILA, O. The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. **Plant, Cell & Environment,** v. 35, n. 10, p. 1707-1728, 2012.

CRISOSTO, C. H.; GRANTZ, D. A.; MEINZER, F. C. EFFECTS OF WATER DEFICIT ON FLOWER OPENING IN COFFEE (COFFEA-ARABICA L). **Tree Physiology,** v. 10, n. 2, p. 127-139, Mar 1992.

DAMATTA, F. M. et al. Ecophysiology of coffee growth and production. **Brazilian Journal of Plant Physiology,** v. 19, p. 485-510, 2007.

DANILEVSKAYA, O. N. et al. Beyond flowering time: pleiotropic function of the maize flowering hormone florigen. **Plant Signal Behav,** v. 6, n. 9, p. 1267-70, Sep 2011.

DE CARVALHO, C. H. S. **Cultivares de café: Origem, características e recomendações**. Brasília, DF: EMBRAPA Café, 2008.

DE CARVALHO, C. H. S. et al. Cultivares de Café Arábica de Porte baixo. In: DE CARVALHO, C. H. S. (Ed.). **Cultivares de Café: Origem, características e recomendações**. Brasília, DF: Embrapa, 2008. cap. 9,

DE OLIVEIRA, R. R. et al. In Silico and Quantitative Analyses of MADS-Box Genes in Coffea arabica. **Plant Molecular Biology Reporter,** v. 28, n. 3, p. 460-472, Jan 2010.

DEDECCA, D. M. Anatomia e desenvolvimento ontogenético de Coffea arabica L. var. typica Cramer. **Bragantia,** v. 16, p. 315-366, 1957.

DENOEUD, F. et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science,** v. 345, n. 6201, p. 1181-1184, 2014.

DO LIVRAMENTO, D. E. Morfologia e Fisiologia do Cafeeiro. In: REIS, P. R. e DA CUNHA, R. L. (Ed.). **Café Arabica: do plantio à colheira**. Lavras, MG: EPAMIG, v.1, 2010. cap. 2, p.894.

DONG, Z.; HAN, M. H.; FEDOROFF, N. The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. **Proc Natl Acad Sci U S A,** v. 105, n. 29, p. 9970-5, Jul 2008.

DRINNAN, J. E.; MENZEL, C. M. Synchronization of anthesis and enhancement of vegetative growth in coffee (Coffea arabica L.) following water stress during floral initiation. **Journal of Horticultural Science,** v. 69, n. 5, p. 841-849, Jan 1994.

EARLEY, K. et al. An endogenous F-box protein regulates ARGONAUTE1 in Arabidopsis thaliana. **Silence,** v. 1, n. 1, p. 15, 20100804 DCOM- 20110714 2010.

ELDEM, V. et al. Genome-Wide Identification of miRNAs Responsive to Drought in Peach (*Prunus persica*) by High-Throughput Deep Sequencing. **PLoS ONE,** v. 7, n. 12, p. e50298, 2012.

FEI, Q.; XIA, R.; MEYERS, B. C. Phased, Secondary, Small Interfering RNAs in Posttranscriptional Regulatory Networks. **The Plant Cell,** v. 25, n. 7, p. 2400-2415, Jul 2013.

FERNANDES-BRUM, C. N. et al. Modificações no metabolismo de carboidratos em mudas de (Coffea arabica L. cv. SIRIEMA) sob condições de déficit hídrico. **Coffee Science,** v. 8, n. 2, Dec 2013.

GENTILE, A. et al. MicroRNAs and drought responses in sugarcane. n. 1664-462X (Electronic), 20150310 DCOM- 20150310 2015.

GRIFFITHS-JONES, S. et al. miRBase: microRNA sequences, targets and gene nomenclature. **Nucleic Acids Res,** v. 34, n. Database issue, p. D140-4, Jan 1 2006.

GRISI, F. A. **RELAÇÕES HÍDRICAS, BIOQUÍMICAS E ANATÔMICAS DE MUDAS DE CAFÉ (Coffea arabica L.) 'CATUAÍ' E 'SIRIEMA' SUBMETIDAS A DÉFICIT HÍDRICO**. 2006. 60 (Mestrado). Universidade Federal de Lavras

GRISI, F. A. et al. Avaliações anatômicas foliares em mudas de café 'catuaí' e 'siriema' submetidasao estresse hídrico. **Ciência e Agrotecnologia,** v. 32, p. 1730-1736, 2008.

GU, Q. et al. The FRUITFULL MADS-box gene mediates cell differentiation during Arabidopsis fruit development. **Development,** v. 125, n. 8, p. 1509-17, Apr 1998.

GUERREIRO-FILHO, O. et al. Origem e Classificação Botânica do Cafeeiro. In: DE CARVALHO, C. H. S. (Ed.). **Cultivares de Café: Origem, características e recomendações**. Brasília, DF: Embrapa Café, 2008.

HAO, X. et al. Coordinated Expression of FLOWERING LOCUS T and DORMANCY ASSOCIATED MADS-BOX-Like Genes in Leafy Spurge. **PLoS ONE,** v. 10, n. 5, p. e0126030, Mar 2015

HARTMANN, U. et al. Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. **The Plant Journal,** v. 21, n. 4, p. 351-360, Feb 2000.

HORVATH, D. P. et al. Transcriptome analysis identifies novel responses and potential regulatory genes involved in seasonal dormancy transitions of leafy spurge (Euphorbia esula L.). **BMC Genomics,** v. 9, p. 536, 2008.

HSU, C. Y. et al. FLOWERING LOCUS T duplication coordinates reproductive and vegetative growth in perennial poplar. **Proc Natl Acad Sci U S A,** v. 108, n. 26, p. 10756-61, Jun 2011.

ICO. Historical Data on the Global Coffee Trade - Total exports by all exporting countries. 2016a. Disponível em: < http://www.ico.org/new_historical.asp >. Acesso em: 02/02/2017.

ICO. Historical Data on the Global Coffee Trade - Total production by all exporting countries. 2016b. Disponível em: < http://www.ico.org/new_historical.asp >. Acesso em: 02/02/2017.

IKI, T. et al. Cyclophilin 40 facilitates HSP90-mediated RISC assembly in plants. **EMBO J,** v. 31, n. 2, p. 267-78, Jan 2012.

IKI, T. et al. In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. **Mol Cell,** v. 39, n. 2, p. 282-91, Jul 2010.

IPCC. **Intergovernmental Panel on Climate Change**. 2014

JEONG, I. S. et al. Arabidopsis C-terminal domain phosphatase-like 1 functions in miRNA accumulation and DNA methylation. **PLoS One,** v. 8, n. 9, p. e74739, 20130923 DCOM- 20140616 2013.

JUNG, J. H. et al. The GIGANTEA-regulated microRNA172 mediates photoperiodic flowering independent of CONSTANS in Arabidopsis. **Plant Cell,** v. 19, n. 9, p. 2736-48, Sep 2007.

KARLSSON, P. et al. KH domain protein RCF3 is a tissue-biased regulator of the plant miRNA biogenesis cofactor HYL1. **Proceedings of the National Academy of Sciences,** v. 112, n. 45, p. 14096-14101, 2015.

KIM, S. et al. Two cap-binding proteins CBP20 and CBP80 are involved in processing primary MicroRNAs. **Plant Cell Physiol,** v. 49, n. 11, p. 1634-44, Nov 2008.

KIM, Y. J. et al. The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. **Embo j,** v. 30, n. 5, p. 814-22, Mar 2011.

KOZOMARA, A.; GRIFFITHS-JONES, S. miRBase: annotating high confidence microRNAs using deep sequencing data. **Nucleic Acids Res,** v. 42, n. Database issue, p. D68-73, Jan 2014.

KRIZEK, B. A.; FLETCHER, J. C. Molecular mechanisms of flower development: an armchair guide. **Nat Rev Genet,** v. 6, n. 9, p. 688-698, Sep 2005.

KURIHARA, Y.; TAKASHI, Y.; WATANABE, Y. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. **Rna,** v. 12, n. 2, p. 206-12, Feb 2006.

LASHERMES, P. et al. Molecular characterisation and origin of the Coffea arabica L. genome. **Mol Gen Genet,** v. 261, n. 2, p. 259-66, Mar 1999.

LAW, J. A.; JACOBSEN, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. **Nat Rev Genet,** v. 11, n. 3, p. 204-220, Feb 2010.

LEE, H. et al. Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in Arabidopsis. **Nucleic Acids Res,** v. 38, n. 9, p. 3081-93, May 2010.

LI, J. et al. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. **Curr Biol,** v. 15, n. 16, p. 1501-7, Aug 2005.

LI, X. et al. Characterization and comparative profiling of the small RNA transcriptomes in two phases of flowering in Cymbidium ensifolium. **BMC Genomics,** v. 16, p. 622, 2015.

LIMA, A. A. **Ethylene Regulation Under Different Watering Conditions and Its Possible Involvement in Coffee (*Coffea arabica* L.) Flowering**. 2015. 151 (PhD). Plant Physiology - Department of Plant Biology, Universidade Federal de Lavras

LIU, C.; THONG, Z.; YU, H. Coming into bloom: the specification of floral meristems. **Development,** v. 136, n. 20, p. 3379-91, Oct 2009.

LIU, J. et al. MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. **Nat Cell Biol,** v. 7, n. 7, p. 719-23, Jul 2005.

LOBBES, D. et al. SERRATE: a new player on the plant microRNA scene. **EMBO Rep,** v. 7, n. 10, p. 1052-8, Oct 2006.

LOSS-MORAIS, G. et al. Identification of novel and conserved microRNAs in Coffea canephora and Coffea arabica. **Genetics and Molecular Biology,** v. 37, n. 4, p. 671-682, Nov 2014

LUO, Y.; GUO, Z.; LI, L. Evolutionary conservation of microRNA regulatory programs in plant flower development. **Dev Biol,** v. 380, n. 2, p. 133-44, Aug 2013.

MAJEROWICZ, N.; SÖNDAHL, M. R. Induction and differentiation of reproductive buds in Coffea arabica L. **Brazilian Journal of Plant Physiology,** v. 17, p. 247-254, 2005.

MARGIS, R. et al. The evolution and diversification of Dicers in plants. **FEBS Letters,** v. 580, n. 10, p. 2442-2450, May 2006.

MATIELLO, J. B. et al. **SIRIEMA AS1, CULTIVAR DE CAFEEIROS COM RESISTÊNCIA MÚLTIPLA, À FERRUGEM E AO BICHO MINEIRO**. 40o Congresso de Pesquisas Cafeeiras. Serra Negra: Embrapa Café 2014.

MATIELLO, J. B. et al. **SIRIEMA VC 4, CULTIVAR CLONAL DE CAFEEIROS COM RESISTÊNCIA MÚLTIPLA, À FERRUGEM E AO BICHO MINEIRO**. 41o Congresso de Pesquisas Cafeeiras. Poços de Caldas: Embrapa Café 2015.

MATIELLO, J. B. et al. **SIRIEMA AS1, CULTIVAR DE CAFEEIRO COM RESISTÊNCIA À FERRUGEM E AO BICHO MINEIRO**. IX Simpósio de Pesquisa dos Cafés do Brasil. Curitiba 2015.

MATZKE, M. A.; KANNO, T.; MATZKE, A. J. M. RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. **Annual Review of Plant Biology,** v. 66, n. 1, p. 243-267, Apr 2015.

MATZKE, M. A.; MOSHER, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. **Nat Rev Genet,** v. 15, n. 6, p. 394-408, Jun 2014.

MAYER, J. L. S.; CARMELLO-GUERREIRO, S. M.; MAZZAFERA, P. A functional role for the colleters of coffee flowers. **AoB Plants,** v. 5, Jan 2013 2013.

MAZZITELLI, L. et al. Co-ordinated gene expression during phases of dormancy release in raspberry (Rubus idaeus L.) buds. **J Exp Bot,** v. 58, n. 5, p. 1035-45, 2007.

MELO, E. F. et al. Anatomic and physiological modifications in seedlings of Coffea arabica cultivar Siriema under drought conditions. **Ciência e Agrotecnologia,** v. 38, p. 25-33, 2014.

MES, M. G. Studies on the flowering of Coffea arabica L. III. Various phenomena associated with the dormancy of the coffee flower buds. **Portugaliae Acta Biologica,** v. 5, n. 1, p. 25-44, 1957.

MEYERS, B. C. et al. Criteria for annotation of plant MicroRNAs. **Plant Cell,** v. 20, n. 12, p. 3186-90, Dec 2008.

MICHAELS, S. D. et al. AGL24 acts as a promoter of flowering in Arabidopsis and is positively regulated by vernalization. **The Plant Journal,** v. 33, n. 5, p. 867-874, Feb 2003.

MOENS, P. Les bourgeons végétatifs et génératifs de Coffea canephora Pierre. (Étude morphologique et morphogénetique). **La Cellule** n. 63, p. 165-244, 1963.

MOISSIARD, G. et al. Transitivity in Arabidopsis can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins. **Rna,** v. 13, n. 8, p. 1268-78, Aug 2007.

MORAIS, H. et al. Escala fenológica detalhada da fase reprodutiva de Cooffea arabica. **Bragantia,** v. 67, n. 1, p. 257-260, 2008.

O'MAOILEIDIGH, D. S.; GRACIET, E.; WELLMER, F. Gene networks controlling Arabidopsis thaliana flower development. **New Phytol,** v. 201, n. 1, p. 16-30, Jan 2014.

OLIVEIRA, R. R. et al. Flower development in Coffea arabica L.: new insights into MADS-box genes. **Plant Reproduction,** v. 27, n. 2, p. 79-94, Jun 2014.

ONODERA, Y. et al. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. **Cell,** v. 120, n. 5, p. 613-22, Mar 2005.

PEZZOPANE, J. R. M. et al. Escala para avaliação de estádios fenológicos do cafeeiro arábica. **Bragantia,** v. 62, n. 3, p. 499-505, 2003.

POSE, D.; YANT, L.; SCHMID, M. The end of innocence: flowering networks explode in complexity. **Curr Opin Plant Biol,** v. 15, n. 1, p. 45-50, Feb 2012.

QUEIROZ VOLTAN, R. B.; IRINEU FAHL, J.; CARVALHO CARELLI, M. L. Diferenciação de gemas florais em cultivares de cafeeiro. **Coffee Science,** v. 6, n. 1, p. 36, Jan 2011.

RAMACHANDRAN, V.; CHEN, X. Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. n. 1095-9203 (Electronic), 20080912 DCOM-20080925 2008.

REBIJITH, K. B. et al. In silico mining of novel microRNAs from coffee (Coffea arabica) using expressed sequence tags. **Journal of Horticultural Science and Biotechnology** v. 88, n. 3, p. 325-337, May 2013.

REDDY, A. Quiescence of coffee flower buds and observations on the influence of temperature and humidity on its release. **Journal of Coffee Research** v. 9, n. 1, p. 1-13, 1979.

REINHARDT, D.; MANDEL, T.; KUHLEMEIER, C. Auxin regulates the initiation and radial position of plant lateral organs. **Plant Cell,** v. 12, n. 4, p. 507-18, Apr 2000.

REINHARDT, D. et al. Regulation of phyllotaxis by polar auxin transport. **Nature,** v. 426, n. 6964, p. 255-60, Nov 2003.

REINHART, B. J. et al. MicroRNAs in plants. **Genes Dev,** v. 16, n. 13, p. 1616-26, Jul 2002.

REN, G.; CHEN, X.; YU, B. Uridylation of miRNAs by hen1 suppressor1 in Arabidopsis. **Curr Biol,** v. 22, n. 8, p. 695-700, Apr 2012.

REN, G. et al. Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis. **Proc Natl Acad Sci U S A,** v. 109, n. 31, p. 12817-21, Jul 2012.

RENA, A. B.; MAESTRI, M. Fisiologia do Cafeeiro. **Informe Agropecuário,** v. 11, n. 126, p. 26-40, Jun 1985.

RÍOS, G. et al. Epigenetic regulation of bud dormancy events in perennial plants. **Frontiers in Plant Science,** v. 5, p. 247, Jun 2014

RNC. Registro Nacional de Cultivares. 2016. Disponível em: < http://www.agricultura.gov.br/vegetal/registros-autorizacoes/registro/registro-nacional-cultivares >. Acesso em: 20/06/2016.

ROGERS, K.; CHEN, X. microRNA biogenesis and turnover in plants. **Cold Spring Harb Symp Quant Biol,** v. 77, p. 183-94, 2012.

ROGERS, K.; CHEN, X. Biogenesis, Turnover, and Mode of Action of Plant MicroRNAs. **Plant Cell,** v. 25, n. 7, p. 2383-2399, Jul 2013.

ROHDE, A. et al. Gene expression during the induction, maintenance, and release of dormancy in apical buds of poplar. **J Exp Bot,** v. 58, n. 15-16, p. 4047-60, 2007.

ROY, S. et al. Identification and Expression Analyses of miRNAs from Two Contrasting Flower Color Cultivars of Canna by Deep Sequencing. **PLoS ONE,** v. 11, n. 1, p. e0147499, Jan 2016.

RUBIO-SOMOZA, I.; WEIGEL, D. Coordination of Flower Maturation by a Regulatory Circuit of Three MicroRNAs. **PLoS Genet,** v. 9, n. 3, p. e1003374, 2013.

RUTTINK, T. et al. A molecular timetable for apical bud formation and dormancy induction in poplar. **Plant Cell,** v. 19, n. 8, p. 2370-90, Aug 2007.

SANTOS, I. S. **Influência do Etileno na Regulação do Florescimento de *Coffea arabica L.*** 2016. 104 (Mestrado em Fisiologia Vegetal). Universidade Federal de Lavras

SCHIEBEL, W. et al. RNA-directed RNA polymerase from tomato leaves. II. Catalytic in vitro properties. **J Biol Chem,** v. 268, n. 16, p. 11858-67, Jun 1993.

SCHOMMER, C. et al. Role of microRNA miR319 in plant development. **Signaling and Communication in Plants,** v. 15, p. 29-47, 2012.

SCHWAB, R. et al. Enhanced microRNA accumulation through stemloop-adjacent introns. **EMBO Reports,** v. 14, n. 7, p. 615-621, Oct 2013.

SHUAI, P. et al. Identification of drought-responsive and novel Populus trichocarpa microRNAs by high-throughput sequencing and their targets using degradome analysis. **BMC Genomics,** v. 14, n. 1, p. 233, 2013.

SPANUDAKIS, E.; JACKSON, S. The role of microRNAs in the control of flowering time. **Journal of Experimental Botany,** v. 65, n. 2, p. 365-380, Feb 2014.

SPETH, C. et al. RACK1 scaffold proteins influence miRNA abundance in Arabidopsis. **Plant J,** v. 76, n. 3, p. 433-45, Nov 2013.

TAIZ , L.; ZEIGER, E. **Fisiologia Vegetal**. 5a 2016. 782

TANG, G. Plant microRNAs: an insight into their gene structures and evolution. **Semin Cell Dev Biol,** v. 21, n. 8, p. 782-9, Oct 2010.

TEOTIA, S.; TANG, G. To Bloom or Not to Bloom: Role of MicroRNAs in Plant Flowering. **Molecular Plant,** v. 8, n. 3, p. 359-377, 2015.

THEISSEN, G.; SAEDLER, H. Plant biology. Floral quartets. **Nature,** v. 409, n. 6819, p. 469-71, Jan 2001.

THOMSON, B.; ZHENG, B.; WELLMER, F. Floral Organogenesis: When Knowing Your ABCs Is Not Enough. **Plant Physiology,** v. 173, n. 1, p. 56-64, Jan 2017.

TU, B. et al. Distinct and cooperative activities of HESO1 and URT1 nucleotidyl transferases in microRNA turnover in Arabidopsis. **PLoS Genet,** v. 11, n. 4, p. e1005119, Apr 2015.

VAUCHERET, H. et al. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. **Genes Dev,** v. 18, n. 10, p. 1187-97, May 2004.

VIEIRA, L. G. E. et al. Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology,** v. 18, p. 95-108, 2006.

WANG, J. W.; CZECH, B.; WEIGEL, D. miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. **Cell,** v. 138, n. 4, p. 738-49, Aug 2009.

WANG, T. et al. Identification of drought-responsive microRNAs in Medicago truncatula by genome-wide high-throughput sequencing. **BMC Genomics,** v. 12, n. 1, p. 367, 2011.

WANG, T. et al. Identification and profiling of novel and conserved microRNAs during the flower opening process in Prunus mume via deep sequencing. **Molecular Genetics and Genomics,** v. 289, n. 2, p. 169-183, 2014.

WIERZBICKI, A. T. et al. Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. **Genes & Development,** v. 26, n. 16, p. 1825-1836, 2012.

WIERZBICKI, A. T. et al. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. **Nat Genet,** v. 41, n. 5, p. 630-4, May 2009.

WOLLMANN, H. et al. On reconciling the interactions between APETALA2, miR172 and AGAMOUS with the ABC model of flower development. **Development,** v. 137, n. 21, p. 3633-42, Nov 2010.

WU, L.; MAO, L.; QI, Y. Roles of dicer-like and argonaute proteins in TAS-derived small interfering RNA-triggered DNA methylation. **Plant Physiol,** v. 160, n. 2, p. 990-9, Oct 2012.

WU, L. et al. DNA methylation mediated by a microRNA pathway. **Mol Cell,** v. 38, n. 3, p. 465-75, May 2010.

WU, X. et al. A role for the RNA-binding protein MOS2 in microRNA maturation in Arabidopsis. **Cell Res,** v. 23, n. 5, p. 645-57, May 2013.

XIE, F. et al. High-throughput deep sequencing shows that microRNAs play important roles in switchgrass responses to drought and salinity stress. **Plant Biotechnology Journal,** v. 12, n. 3, p. 354-366, 2014.

XIE, Z. et al. Expression of Arabidopsis MIRNA Genes. **Plant Physiology,** v. 138, n. 4, p. 2145-2154, Aug 2005.

YAMAGUCHI, A.; ABE, M. Regulation of reproductive development by non-coding RNA in Arabidopsis: to flower or not to flower. **J Plant Res,** v. 125, n. 6, p. 693-704, Nov 2012.

YU, B. et al. The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. **Proc Natl Acad Sci U S A,** v. 105, n. 29, p. 10073-8, Jul 2008.

ZENG, Y.; CULLEN, B. R. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. **Nucleic Acids Res,** v. 32, n. 16, p. 4776-85, 2004.

ZHAI, J. et al. Rapid construction of parallel analysis of RNA end (PARE) libraries for Illumina sequencing. **Methods,** v. 67, n. 1, p. 84-90, May 2014.

ZHAN, X. et al. Arabidopsis proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. **Proc Natl Acad Sci U S A,** v. 109, n. 44, p. 18198-203, Oct 2012.

ZHANG, B. H. et al. Identification of cotton microRNAs and their targets. **Gene,** v. 397, n. 1-2, p. 26-37, Aug 2007.

ZHANG, T.; WANG, J.; ZHOU, C. The role of miR156 in developmental transitions in Nicotiana tabacum. **Sci China Life Sci,** v. 58, n. 3, p. 253-60, Mar 2015.

ZHAO, M. et al. Evolutionary Patterns and Coevolutionary Consequences of MIRNA Genes and MicroRNA Targets Triggered by Multiple Mechanisms of Genomic Duplications in Soybean. LID - tpc.15.00048 [pii]. n. 1532-298X (Electronic), Mar 2015.

ZHAO, Y. et al. The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. n. 1879-0445 (Electronic), 20120427 DCOM- 20121207 2012.

ZHONG, W. et al. Genome-wide expression profiles of seasonal bud dormancy at four critical stages in Japanese apricot. **Plant Mol Biol,** v. 83, n. 3, p. 247-64, Oct 2013.

ZHOU, C.-M.; WANG, J.-W. Regulation of Flowering Time by MicroRNAs. **Journal of Genetics and Genomics,** v. 40, n. 5, p. 211-215, May 2013.

ZHOU, M.; LAW, J. A. RNA Pol IV and V in gene silencing: Rebel polymerases evolving away from Pol II's rules. **Current Opinion in Plant Biology,** v. 27, p. 154-164, Oct 2015.

ZHU, Y. et al. RNA-Seq-based transcriptome analysis of dormant flower buds of Chinese cherry (Prunus pseudocerasus). **Gene,** v. 555, n. 2, p. 362-76, Jan 2015.

PART 2: Articles

ARTICLE 1 - A genome-wide analysis of the RNA-guided silencing pathway in coffee reveals insights into its regulatory mechanisms

This article has been submitted to the Journal PLOS ONE

A genome-wide analysis of the RNA-guided silencing pathway in coffee reveals insights into its regulatory mechanisms

Christiane Noronha Fernandes-Brum[1], Pâmela Marinho Rezende[1], Thales Henrique Cherubino Ribeiro[1], Raphael Ricon de Oliveira[2], Thaís Cunha de Sousa Cardoso[3], Laurence Rodrigues do Amaral[3], Matheus de Souza Gomes[3], Antonio Chalfun-Junior[1*]

[1]Department of Biology, Section of Plant Physiology, Laboratory of Plant Molecular Physiology (LFMP), Federal University of Lavras (UFLA), Lavras, Minas Gerais, Brazil;

[2]Departamento de Genética, Universidad de Córdoba (UCO), Córdoba, Spain;

[3] 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

*Corresponding author:

E-mail: chalfunjunior@dbi.ufla.br

Abstract

microRNAs (miRNAs) are derived from self-complementary hairpin structures, while small-interfering RNAs (siRNAs) are derived from double-stranded RNA (dsRNA) or hairpin precursors. The core mechanism of sRNA production involves DICER-like (DCL) in processing the smallRNAs (sRNAs) and ARGONAUTE (AGO) as effectors of silencing, and siRNA biogenesis also involves action of RNA-Dependent RNA Polymerase (RDR), Pol IV and Pol V in biogenesis. Several other proteins interact with the core proteins to guide sRNA biogenesis, action, and turnover. We aimed to unravel the components and functions of the RNA-guided silencing pathway in a non-model plant species of worldwide economic relevance. The sRNA-guided silencing complex members have been identified in the *Coffea canephora* genome, and they have been characterized at the structural, functional, and evolutionary levels by computational analyses. Eleven AGO proteins, nine DCL proteins (which include a DCL1-like protein that was not previously annotated), and eight RDR proteins were identified. Another 48 proteins implicated in smallRNA (sRNA) pathways were also identified. Furthermore, we identified 235 miRNA precursors and 317 mature miRNAs from 113 MIR families, and we characterized *ccp-MIR156*, *ccp-MIR172*, and ccp-*MIR390*. Target prediction and gene ontology analyses of 2239 putative targets showed that significant pathways in coffee are targeted by miRNAs. We provide evidence of the expansion of the loci related to sRNA pathways, insights into the activities of these proteins by domain and catalytic site analyses, and gene expression analysis. The number of MIR loci and their targeted pathways highlight the importance of miRNAs in coffee. We identified several roles of sRNAs in *C. canephora*, which offers substantial insight into better understanding the transcriptional and post-transcriptional regulation of this major crop.

Key words: *Coffea canephora*, miRNA, post-transcriptional gene silencing (PTGS), siRNA, sRNA, silencing

Introduction

Small RNA (sRNA) silencing pathways have attracted increasing interest in the fields of genetics and molecular biology, and our current knowledge regarding the mechanisms and components involved in these pathways has rapidly evolved. Such RNA-based processes consist of sequence-specific inhibition of gene expression at the transcriptional or translational level by the action of small (20-26 nt) homologous RNA sequences [1].

Plant sRNAs are produced by processing of double-stranded duplexes from the helical regions of larger RNA precursors and are classified according to the intra- or intermolecular hybridization of the duplex [2]. microRNAs (miRNAs) are derived from self-complementary hairpin structures, while small-interfering RNAs (siRNAs) are derived from double-stranded RNA (dsRNA) or hairpin precursors [3,4].

MIR genes are transcribed by RNA polymerase II (Pol II) [5] and undergo several modifications from transcription to maturity. Primary transcripts (pri-miRNAs) are similar to protein-coding RNA precursors (pre-mRNA) in size [6] but possess a hairpin structure that is stabilized by the RNA-binding protein DAWDLE (DDL) [7]. These molecules are processed by the endonuclease activity of DICER-LIKE 1 (DCL1) [8] into precursors (pre-miRNAs) assisted by additional enzymes, including HYPONASTIC LEAVES 1 (HYL1) [8], SERRATE (SE) [9,10], and TOUGH (TGH) [11]. The pre-miRNAs are then processed by the DCL complex to form a duplex structure containing two 3' nucleotide overhangs at each end. miRNAs are generally 21 nt long (DCL1 and DCL4), but their size varies depending on the DCL that induces cleavage, being

22 nt for DCL2 and 24 nt for DCL3 [12]. miRNAs negatively regulate their target genes through sequence-specific degradation or translational repression [13]. However, some miRNAs are also involved in DNA methylation [14].

The duplex is 3' methylated by the methyltransferase HUA ENHANCER 1 (HEN1), which protects it from further modification and degradation [15]. The exportin HASTY (HST) is responsible for binding the duplex and transporting it from the nucleus to the cytoplasm [16]. Exportation in the absence of this protein is also possible but occurs via an unknown mechanism [17]. In the cytoplasm, one strand of the duplex is loaded onto an ARGONAUTE (AGO) family protein containing the PAZ and PIWI domains to form the RISC (RNA-Induced Silencing Complex). The PIWI domain possesses endonuclease activity and cleaves the target mRNA, which is also recognized by nearly perfect complementarity with the miRNA [12,18].

The other major class of sRNAs, siRNAs, can act either at the transcriptional level by guiding DNA methylation or at the post-transcriptional level by guiding the cleavage and degradation of homologous cellular transcripts [1,19]. RNA-dependent RNA Polymerases (RDRs) play an important role in siRNA production, synthesizing a second-strand RNA from the RNA template and thus producing a double-stranded RNA (dsRNA) molecule [20] with initial priming-dependent or priming-independent characteristics [21]. The biogenesis of siRNA shares a core mechanism with miRNAs. siRNAs are processed by a DCL protein (DCL2, DCL3, and DCL4), methylated by HEN1, and loaded onto a protein of the AGO family [2].

Additionally, two plant-specific DNA-Dependent RNA Polymerases, Pol IV and Pol V, are involved in the biogenesis of 24-nt siRNAs, which mediate RNA-Dependent DNA Methylation (RdDM). RdDM occurs through cytosine methylation (CG, CHG, and CHH, where H = A, C, or T) by the *de novo* methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE 2

(DRM2) at the target DNA locus [22,23]. Pol IV transcribes heterochromatic regions, which code for siRNAs [24], followed by dsRNA synthesis by RDR2, processing by DCL3, and the assembly of the resulting siRNA duplexes in the AGO4 clade of AGOs [23]. Pol V produces transcripts from Intergenic Non-coding (IGN) regions at loci that will be further methylated and is required for the recruitment of RdDM machinery, including DRM2 and siRNA-loaded AGO [25,26]. This recruitment occurs by the interaction between protein-protein (Pol V-AGO) and nucleic acids, however, it remains unclear whether siRNA:IGN or siRNA:DNA. [27,28].

Along with the core mechanism of sRNA production described above, using DCL in processing and AGOs as effectors, and additional participation of the RDR, Pol IV and Pol V in siRNA biogenesis, several other proteins interact with these core proteins to guide sRNA biogenesis, action, and turnover. These proteins have been recently reviewed [17,19]. For instance, RECEPTOR FOR ACTIVATED C KINASE 1 (RACK1) and C-TERMINAL DOMAIN PHOSPHATASE-LIKE 1 (CPL1) interact with SE and have been implicated in pri-miRNA processing [29,30]. Due to their recent emergence, the sRNA silencing pathways have not been fully elucidated, and knowledge of these pathways is constantly evolving. More recently, the protein REGULATOR OF CBF GENE EXPRESSION 3 (RCF3) has been described as a cofactor affecting miRNA biogenesis in specific plant tissues by interacting with CPL1 and CPL2 [31].

Aiming to expand the knowledge from model plants, the silencing complex has been identified in native and cultivated species, including rice (*Oryza sativa*) [32], common bean (*Phaseolus vulgaris*) [33], sorghum (*Sorghum bicolor*), and soybean (*Glycine max*) [34]. In *Coffea arabica* and *Coffea canephora*, the main economically important species of coffee, one of the most important crops in the world and the second most traded global commodity, MIR

families have been identified based on Expressed Sequence Tags (EST), Genome Survey Sequences (GSS), and other transcript-based analyses [35-38].

With the release of the *C. canephora* genome, miRNAs were also identified [39]. However, the number of miRNAs was significantly underestimated. Moreover, the genes implicated in the generation and function of the miRNAs and siRNAs have not been described in coffee plants.

In this work, we present a thorough analysis of the identification and characterization of the small RNA-guided silencing complex in the *C. canephora* genome. Eleven AGO proteins; nine DCL-like proteins, including a previously unannotated DCL1; eight RDR proteins; and 48 other proteins implicated in the sRNA pathways, including HYL1, HST, HEN1, SE, and TGH, were identified. Furthermore, we conducted a conserved domain, catalytic site, and phylogenetic analysis to characterize the main proteins of the silencing pathway and validated their expression using RNA-seq libraries. We also identified 235 miRNA precursors producing 317 mature miRNAs belonging to 113 MIR families. We structurally and evolutionarily characterized and identified the putative targets of the MIR families *MIR156*, *MIR172*, and *MIR390.* A total of 2239 putative *C. canephora* miRNA targets were identified, and gene ontology analyses showed that significant pathways were targeted by miRNAs, demonstrating the importance of miRNAs in *C. canephora*.

The identification and analysis of the sRNA silencing pathways in *C. canephora* not only provide insights into the species but also provide a basis for further study of *C. canephora* and *C. arabica* regarding sRNA biogenesis and activity. The comprehension of these pathways in such an important crop provides insights into the species for further use of genetic engineering technologies available for crop breeding.

Materials and Methods

miRNA and protein prediction datasets

The *C. canephora* genome data and genome features were accessed and downloaded from The Coffee Genome Hub [39]. Mature plant miRNA sequences and precursor miRNA sequences were downloaded from miRBase version 21. For protein prediction, Arabidopsis (*Arabidopsis thaliana*) ortholog sequences were retrieved from the nucleotide and protein databases at the NCBI (National Center for Biotechnology Information).

Prediction of genes and proteins involved in the sRNA pathway in *C. canephora*

Putative proteins involved in the sRNA pathways were identified and selected by mining *C. canephora* sequences in the Coffee Genome Hub, an integrated web-based database, using the Basic Local Alignment Search Tool (BLAST) algorithm BLASTp with protein sequences from Arabidopsis as queries to search previously annotated protein-coding genes. The resulting protein sequences were retrieved for further analysis.

Prediction of mature miRNAs and their precursors (pre-miRNAs)

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* [40]. First, the genome and transcriptome sequences of *C. canephora* were searched using

BLASTN to identify putative hairpin-like structures. The retrieved sequences were E-inverted (EMBOSS tool) using the maximum repeat parameters of 336 nucleotides and a threshold value of 25. Then, several lters were applied based on the thermodynamics and structural characteristics of known miRNAs. These lters included a GC content (guanine and cytosine) between 20% and 65%, Minimum Free Energy (MFE), homology with known mature miRNAs, homology to repetitive regions in RepeatMasker 4.0.2 [41], and homology to non-coding RNAs, such as rRNA, snRNA, SL RNA, SRP, tRNA, and RNase P, deposited in the Rfam microRNA Registry version 11.0 [42].

The sequences of pre-miRNAs identi ed in *C. canephora* were characterized according to their structures and thermodynamic parameters. The assessed parameters included the MFE, Adjusted Minimum Free Energy (AMFE), Minimum Free Energy Index (MFEI), size, A content, U content, C content, G content, GC and AU contents, GC ratio, AU ratio, Minimum Free Energy of the thermodynamic ensemble (MFEE), Ensemble Diversity (Diversity), and frequency of the MFE structure in the ensemble (Frequency). The adjusted MFE (AMFE) was determined to be a sequence of 100 nt, and the MFEI was determined using the equation MFEI = [(AMFE) X 100]/(G% + C%)] [43,44]. The secondary structures of pre-miRNA, diversity, MFE, frequency ensemble, and MFE were predicted using RNA-fold software (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). The GC content and other structural properties were defined using Perl scripts.

Analyses of the sRNA pathway proteins and miRNA precursors

The protein families, domains, and active sites were analyzed using PFAM (version 27.0, available at http://pfam.sanger.ac.uk) and the Conserved

Domains Database (CDD; http://www.ncbi.nlm.nih.gov/cdd/). The protein sequences from *C. canephora* and their orthologs from different species were used to perform multiple sequence alignments using ClustalX 2.0 based on the default settings (available at http://www.clustal.org/clustal2/; [45]). The homologs and the *C. canephora* pre-miRNAs were aligned using ClustalX 2.0 based on the following alignment parameters: a gap opening of 22.50 and a gap extension of 0.83. They were also aligned in RNAalifold (http://rna.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi). Phylogenetic trees were inferred using the neighbor-joining method, and sequence divergence was estimated using the Jones–Taylor–Thornton model for proteins [46] and Kimura's (1980) two-parameter model for pre-miRNAs [47]. Statistical reliabilities of the internal branches were assessed using 2000 bootstrap replicates for proteins and 5000 bootstrap replicates for pre-miRNAs with values greater than 30 above the branches. Molecular phylogenetic analyses were conducted using MEGA 5 software [48]. The catalytic domains of ARGONAUTE and DICER-like proteins were aligned using Clustal Omega. Pictures highlighting the catalytic residues were generated from the alignment. Multiple Em for Motif Elicitation (MEME) (Version 4.11.2) [49] was then used to find RDR-like catalytic motifs.

RNA-seq analysis

RNA-seq libraries were downloaded from the SRA (https://www.ncbi.nlm.nih.gov/sra/?term=ERP003741) for the three leaf stages (young, expanded, and old) and stems of the *C. canephora* samples.

For *CcDCL1* prediction, the RNA-seq libraries were assembled using Trinity [50]. BLASTN was run against the assembled data using AtDCL1 as a query. The six retrieved sequences were re-assembled using CAP3 [51], and two novel contigs were formed. The protein sequence of the largest contig was predicted using GenScan (http://genes.mit.edu/GENSCAN.html).

For expression validation, the transcriptome in different tissues was assembled using the alignment of the RNA-seq reads against the *C. canephora* genome with the software TopHat2. The subsequent identification of new genes and alternative splicing analysis were completed with the Cufflinks package. After alignment, possible coding sequences were extracted and identified with the Trans Decoder algorithm and subjected to homology analysis with BLAST. After selecting the proteins involved in the sRNA pathways, differential expression analysis was conducted with the CuffDiff software. The results were visualized and plotted using several packages of the statistical environment R, including the cummeRbund package.

Prediction of *C. canephora* miRNA target genes

To search for putative target genes of the predicted miRNAs in *C. canephora*, transcript (CDS+UTR) sequences were retrieved from the Coffee Genome Hub (http://coffee-genome.org/download) and from RNA-seq libraries (transcript-predicted) of two tissue types: leaves and stem. *C. canephora* miRNA target genes were predicted using the webtool psRNATarget [52]. To avoid false-positive predictions for the miRNA target genes, we used a stringent cutoff threshold for a maximum expectation of 2.0. The other parameters were based on default settings, which included a length for complementarity scoring (hspsize) of 20 bp, top number of target genes for each small RNA of 200, target accessibility, maximum energy to unpair the target site (UPE) of 25, anking length around the target site for target accessibility analysis of 17 bp upstream/13 bp downstream, and a range of the central mismatch leading to translational inhibition of 9–11 nt.

Using the RNA-seq sequences, BLAST2GO was run with the resulting predicted targets for each of the miRNAs *MIR156*, *MIR172*, and *MIR390.* BLAST2GO began with a BLASTP search against SwissProt, followed by mapping and annotation.

GO classes of the miRNA targets were classified and grouped using the web tool SEA (Singular Enrichment Analysis) from agriGO (http://bioinfo.cau.edu.cn/agriGO/index.php) [53]. The input was the target genomic IDs, which were compared against all of the IDs of the Coffee Genome Hub.

Results

sRNAs pathways proteins prediction and validation

The proteins involved in the miRNA pathways were identified by BLASTP in the Coffee Genome using Arabidopsis orthologs as queries. The components of the miRNA pathway, HYL1, SE, DDL and TGH [7,9-11], were identified, and one copy of each of these proteins was identified in the *C. canephora* genome (Table 1). Two core proteins of the sRNA pathways, HEN1 and HST, were also identified. One putative CcHEN1 and one CcHST protein were identified (Table 1). In addition, we also identified at least 48 proteins in the *C. canephora* genome associated with the sRNA pathways described in the literature (S1 Table).

Table 1. HYL1, SE, DDL, TGH, HEN1, and HST orthologs of *C. canephora*.

Protein Name	ID Arabidopsis	Size (aa)	*C. canephora* Locus name	Locus Position	Size (aa)
DDL	NP_188691.1	314	Cc05_g13470	chr5:27034635..27039361	402
TGH	NP_001031926.1	900	Cc04_g07720	chr4:6122482..6132431	852
HYL1	NP_563850.1	419	Cc10_g15960	chr10:26908423..26911736	321
HEN1	NP_001190782.1	942	Cc09_g07800	chr9:10021237..10030396	951
SE	NP_565635.1	720	Cc01_g07580	chr1:25540845..25550602	761
HST	NP_187155.2	1202	Cc02_g32190	chr2:43066609..43081800	1199

Protein name, ID, and size in Arabidopsis, *C. canephora* locus name, position, and protein size

The core proteins of the sRNA pathways- DCL-like, AGO-like, and RDR-like - were identified and characterized as described below. The *C. canephora* protein name, locus position, length, and identity with their respective orthologs from Arabidopsis are presented in Table 2.

Table 2. The *Coffea canephora* DCL-like, AGO-like and RDR-like protein orthologs. Protein name, ID, and length in Arabidopsis, BLASTp e-value and Identity of *C. canephora* vs. Arabidopsis. *C. canephora* ortholog name, locus name, locus position, and protein length.

Protein Name	ID Arabidopsis	Protein length (aa)	BLASTP (e-value) vs *A. thaliana*	Identity	*C. Canephora* ortholog	Locus	Location coordinates	Protein length (aa)
AGO1	NP_171612.1	1060	0.0	84%	CcAGO1	Cc04_g08880	chr4:7327522..7334534	1070
AGO2	NP_174413.2	1014	0.0	48%	CcAGO2.2	Cc09_g06780	chr9:7781473..7787026	1103
	NP_174413.2	1014	0.0	46%	CcAGO2.1	Cc09_g06770	chr9:7773251..7777143	1072
AGO4	NP_001189613.1	924	$4e^{-81}$	43%	CcAGO4.1	Cc04 g10830 Cc04 g10840	chr4:10274296..10280759	
AGO4	NP_001189613.1	924	0.0	74%	CcAGO4.2	Cc01_g06780	chr1:24122477..24129690	869
AGO4	NP_001189613.1	924	0.0	69%	CcAGO4.3	Cc00_g14230	chr0:103099681..103105365	867
AGO5					CcAGO5	Cc01_g10060	chr1:28754803..28760661	960
AGO7	NP_177103.1		0.0	69%	CcAGO7	Cc11_g12560	chr11:29570089..29573706	1014
AGO10	NP_001190464.1	988	0.0	81%	CcAGO10.1	Cc03_g04370	chr3:3329168..3336865	992
	NP_001190464.1	988	0.0	73%	CcAGO10.2	Cc06_g09120	chr6:7288302..7294655	932
AGO16					CcAGO16	Cc05_g02730	chr5:12039961..12045923	909
DCL1	NP_171612.1	1909	0.0	76%	CcDCL1	-	chr0:59461839..59481838	1747
DCL2	NP_566199.4	1388	0.0	55%	CcDCL2.1	Cc09_g03980	chr9:3364371..3376041	1352
	NP_566199.4	1388	0.0	47%	CcDCL2.2	Cc02_g14900 Cc02_g14910	chr2:13049228..13060040	778
	NP_566199.4	1388	$3e^{-112}$	51%	CcDCL2.5	Cc06_g19770	chr6:21807446..21809500	354
	NP_566199.4	1388	0.0	48%	CcDCL2.6	Cc06_g19980	chr6:22425311..22432933	762

	NP_566199.4	1388	0.0	50%	CcDCL2.4	Cc02_g14930	chr2:13070716..13077527	802
	NP_566199.4	1388	0.0	48%	CcDCL2.3	Cc02_g14920	chr2:13060040..13066011	727
DCL3	NP_001154662.2	1580	0.0	48%	CcDCL3	Cc08_g06780	chr8:17408330..17423075	1584
DCL4	NP_001190348.1	1688	0.0	51%	CcDCL4	Cc06_g07320	chr6:5843020..5862408	1656
RDR1	NP_172932.1	1107	0.0	63%	CcRDR1.1	Cc11_g06970	chr11:23552744..23560803	1114
	NP_172932.1	1107	0.0	64%	CcRDR1.2	Cc11_g06940	chr11:23487397..23495045	1113
	NP_172932.1	1107	0.0	60%	CcRDR1.3	Cc11_g06960	chr11:23538795..23545065	1132
	NP_172932.1	1107	0.0	56%	CcRDR1.4	Cc11_g06950	chr11:23504270..23516759	1188
RDR2	NP_192851.1	1133	0.0	57%	CcRDR2	Cc00_g08850	chr0:76051887..76058404	1121
RDR3	NP_179581.2	992	0.0	43%	CcRDR3.1	Cc06_g10360	chr6:8381378..8392034	1020
	NP_179581.2	992	0.0	47%	CcRDR3.2	Cc06_g10350	chr6:8366687..8376181	876
RDR6	NP_190519.1	1196	0.0	67%	CcRDR6	Cc08_g00760	chr8:779886..784083	1050

The number of DCLs may vary among species. For instance, there are five DCLs in poplar, maize (*Zea mays*), and sorghum (*S. bicolor*) [34,54]; seven in tomato (*Solanum lycopersicum*) [55]; eight in rice (*O. sativa*) [56]; and six in common bean (*P. vulgaris*) [33].

Table 3. Conserved domain analysis of the *C. canephora* DCL-like orthologs.

Locus Name	Protein Name	DExD	Helicase-C	DUF283	PAZ	RIBOC	RIBOC	DSRM	DSRM
-	CcDCL1	114-266	503-619	693-784	1029-1164	1201-1387	1423-1579	1582-1643	1674-1742
Cc09_g03980	CcDCL2.1	2-137	318-436	507-592	760-887	935-1087	1119-1272	-	-
Cc02_g14900 Cc02_g14910	CcDCL2.2	-	-	-	162-290	338-478	519-705	709-765	-
Cc06_g19770	CcDCL2.5	-	-	-	-	48-85	126-280	284-340	
Cc06_g19980	CcDCL2.6	-	-	-	174-291	339-490	524-679	685-738	-
Cc02_g14930	CcDCL2.4	-	-	-	177-305	353-497	538-692	-	-
Cc02_g14920	CcDCL2.3	-	-	-	153-273	321-465	506-660	664-723	-
Cc08_g06780	CcDCL3	53-215	406-524	603-690	889-1037	1079 -1243	1289-1439	-	-
Cc06_g07320	CcDCL4	81-232	412-534	606-683	873-993	1041-1204	1242-1386	1395-1459	1572-1645

The annotated protein-coding sequences identified from the BLASTP of the DCL-like search in the Coffee Genome Hub were retrieved, and conserved domain analysis revealed that nine of these sequences contained DCL-like conserved domains (Table 3). Two of the sequences (Cc02_14900 and Cc02_14910) that are sequential in chromosome 2 presented complementary domains of a DCL protein. Then, the genomic region comprising both contigs was retrieved, and the resulting protein was predicted using GenScan (http://genes.mit.edu/GENSCAN.html) and used for further analyses.

Multiple alignments with ortholog DCLs from other angiosperm species and phylogenetic analyses were performed to assign the coffee DCLs and to determine the evolutionary relationship among species. One DCL3, one DCL4, and six DCL2s were assigned. No DCL1 was found using this approach, then we identified one putative CcDCL1 from RNA-seq libraries. Conserved domain analysis (Table 3) of the resulting sequence confirmed a DCL protein, and BLASTP at the NCBI database matched DCL1 proteins with 99% coverage and an E-value of 0. The sequence was then searched by tBLASTN in the Coffee Genome Hub and aligned with a genomic sequence in chromosome 0, an arbitrary pseudochromosome created with all of the unmapped sequences from the 11 chromosomes [39] (S1 Fig). Therefore, although present in the genome assembly, the CcDCL1 was not previously annotated as a protein-coding gene on the Coffee Genome Hub.

The new phylogenetic analysis, including the putative CcDCL1, generated a tree in which the CcDCL clustered similarly to their respective orthologs from other species (Fig 1). In total, nine DCL-like proteins were found in the *C. canephora* genome (Table 2) and were distributed in four distinct clades in the phylogenetic tree (Fig 1); the clades matched the four paralogous DCL-like proteins described in Arabidopsis [57].

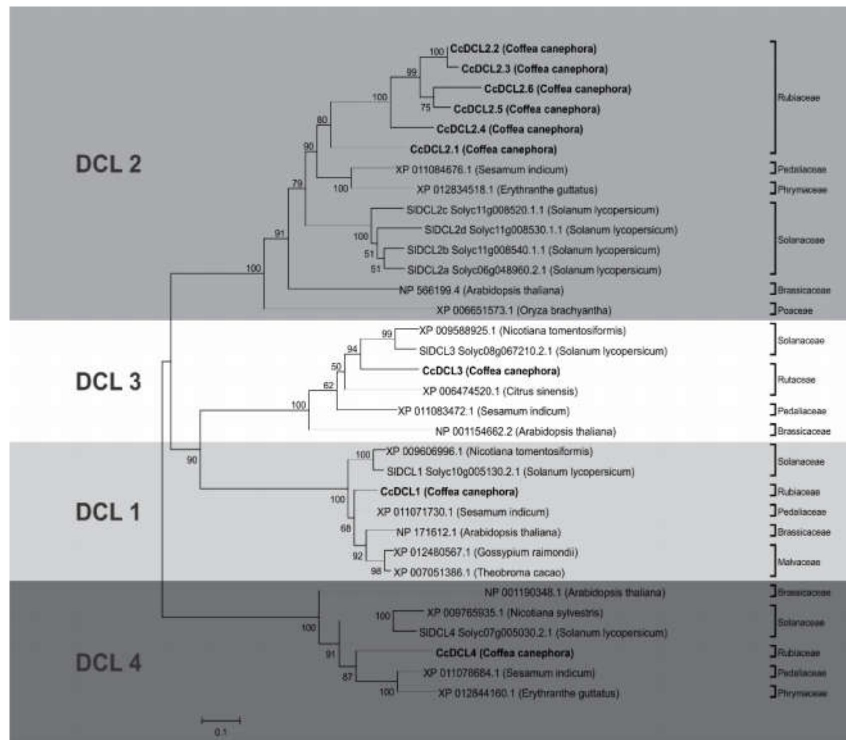


Fig 1. Phylogenetic tree of DCL-like proteins identified in *Coffea canephora.* Phylogenetic tree showing relationships between the paralogous and orthologs proteins of the DCL family. The evolutionary history was inferred using the Neighbor-Joining method [46]. The bootstrap consensus tree inferred from 2000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (2000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site [48]. The analysis involved 33 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 286 positions in the final dataset.

The DCL proteins have six domains types, DExD-helicase (DExDc), Helicase-C (HELICc), Duf283, PAZ, RNAse III (RIBOc), and double-stranded RNA-binding (dsRB), although some of these may not be present [58]. Conserved domain analysis (Table 3) revealed that the CcDCL1-like and CcDCL4-like proteins contain DExD, Helicase-C, Dicer-dimer, PAZ, two

RNAse III (RIBOc), and two dsRB (DSRM) domains. The CcDCL3-like, CcDCL2.1-like, and DCL4-like proteins contain no DSRM domains. The CcDCL2 proteins have five more paralogs, which appear to be partial sequences lacking the N-terminal domains (DExD, Helicase-C, and DUF283). These sequences also lack one (CcDCL2.3, CcDCL2.4, and CcDCL 2.6) or two (CcDCL2.5) DSRM domains. The shortest CcDCL2-like protein, CcDCL2.3, also lacks a PAZ domain.

We also analyzed the conservation of the RNase III catalytic sites of CcDCL-like proteins in the two RNase III domains (RIBOc I and II): glutamate (E), aspartate (D), glutamate (D), aspartate (E) (EDDE) [59]. CcDCL1, CcDCL2.1, CcDCL3, and CcDCL4 contain these conserved catalytic residues (Fig 2).

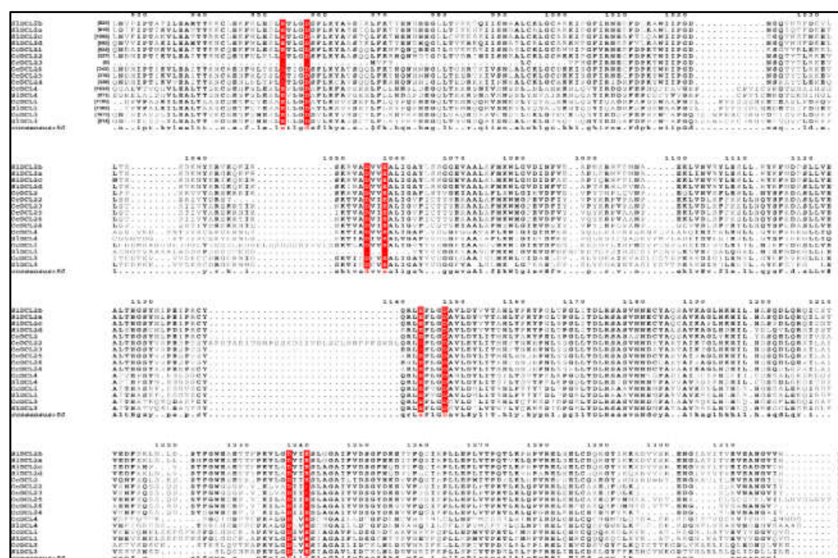


Fig 2. Analysis of the catalytic residues of the CcDCL-like proteins. The two RNase III domains (RIBOc I and II) at the glutamate (E), aspartate (D), glutamate (D), aspartate (E) (EDDE) position. The catalytic sites are highlighted.

ARGONAUTES have been observed in variable numbers in plants. For instance, there are 10 AGOs in Arabidopsis [60], 22 in soybean (*G. max*) [34], 17 in common bean (*P. vulgaris*) [33], 19 in rice (*O. sativa*) [32], and 17 in maize (*Z. mays*) [54]. A BLASTP search using AtAGO as a query in the

Coffee Genome Hub resulted in 12 *C. canephora* protein-coding sequences, which were retrieved and subjected to Conserved Domain analysis to confirm the presence of the conserved domains of ARGONAUTE proteins (N-terminal, PAZ, ArgoMid, and PIWI). Two of the sequences (Cc04_g10830 and Cc04_g10840) that were found sequentially in Chromosome 4 presented as partial sequences, one containing a PIWI domain (Cc04_g10830) and the other containing a PAZ (Cc04_g10840) domain. The genomic sequence comprising both contigs was retrieved, and the protein product was predicted using GenScan (http://genes.mit.edu/GENSCAN.html). BLASTP and Conserved Domain analysis confirmed an AGO protein that was considered for further analyses. Therefore, in total, eleven putative AGO proteins comprising seven homologs were found in *C. canephora* (Table 2).

Conserved domain analysis confirmed the presence of the N-terminal, PAZ, and PIWI domains in all sequences but showed an only variable presence of ArgoMid (Table 4). AGO1 proteins have an additional glycine-rich region at the N-terminus (Gly-rich_Ago1), which was present in one putative AGO sequence. To further determine the evolutionary conservation and assign the AGO-like proteins found in *C. canephora*, we compared the sequences to orthologs from other angiosperm species on a phylogenetic tree. The eleven AGO proteins were assigned and found to cluster with their closest orthologs from other species; the *C. canephora* AGO proteins also similarly grouped into three major phylogenetic clades [17,61]: one AGO1, one AGO5, and two AGO10s in Clade I; two AGO2s and one AGO7 in Clade II; and three AGO4s in Clade III (Fig 3). One AGO16 was also identified, which grouped with the AGO4s in Clade III. . A similar pattern has been found in rice, maize, Arabidopsis, soybean, sorghum, and other species, indicating the conservation of small RNA functions in higher plants [34].

Table 4. Identification of the conserved domains and their start and end positions in the *C. canephora* AGO orthologs.

Locus Name	Protein Name	Gly-rich_Ago1	ArgoN	PAZ	ArgoMid	Piwi
Cc04 g08880	CcAGO1	76-186	205-341	407-532	600-674	694-1013
Cc09 g06780	CcAGO2.2	-	253-393	458-581	-	758-1052
Cc09 g06770	CcAGO2.1	-	218-362	426-551	-	728-1022
Cc04 g10830 Cc04 g10840	CcAGO4.1	-	62-184	264-355	-	355-465
Cc01 g06780	CcAGO4.2	-	3-172	238-368	432-495	522-828
Cc00 g14230	CcAGO4.3	-	4-172	238-366	-	520-827
Cc01 g10060	CcAGO5	-	117-257	322-441	508-583	601-919
Cc11 g12560	CcAGO7	-	151-308	380-502	-	666-972
Cc03 g04370	CcAGO10.1	-	136-279	350-471	538-615	630-949
Cc06 g09120	CcAGO10.2	-	88-227	305-419	486-563	578-896
Cc05 g02730	CcAGO16	-	38-202	269-399	-	553-868

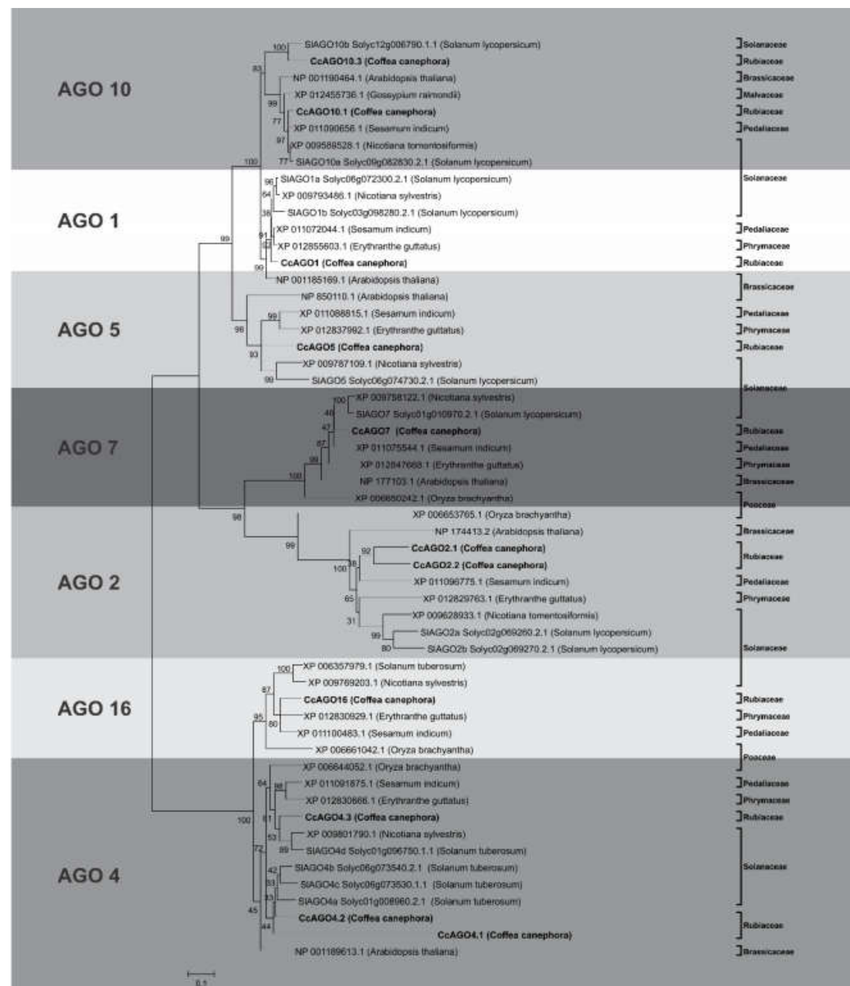


Fig 3. Phylogenetic tree of AGO proteins identified in *Coffea canephora.* Phylogenetic tree showing relationships between the paralogous and orthologs proteins of the AGO family. The evolutionary history was inferred using the Neighbor-Joining method [46]. The bootstrap consensus tree inferred from 2000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (2000 replicates) are shown next to the branches. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site [48] . The analysis involved 55 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 333 positions in the final dataset.

To investigate whether CcAGOs possess conserved catalytic residues and could potentially act as the slicer component of RISC, we aligned the PIWI domains of all of the CcAGOs and searched for the Asp-Asp-His (DDH) catalytic triad in CcAGOs and for a residue corresponding to the conserved H798 residue of AtAGO1 [62]. Four proteins (CcAGO1, CcAGO5, CcAGO7, and CcAGO10.1) possessed the conserved DDH/H798 residues (Table 5). In four CcAGOs, the DDH catalytic motif was conserved, but the H798 was replaced by a serine (CcAGO16), proline (CcAGO4.2 and CcAGO4.3), or glutamine (CcAGO10.2). Two CcAGO proteins contained an aspartate residue in place of the third histidine of the DDH motif (CcAGO2.1 and Cc AGO2.2). CcAGO4.1 contained neither the catalytic DDH nor the H798 residue. The detailed alignment of the PIWI domain is presented in S2 Fig.

Table 5. Analysis of active site amino acids and their respective position in the conserved PIWI domain (PF02171) from the CcAGO proteins.

CcAGO	Motifs*	POSITION
CcAGO1	DDH/H	777-863-1003/815
CcAGO2.1	DDD/H	807-880-1014/845
CcAGO2.2	DDD/H	837-910-1045/875
CcAGO4.1	ENR/R	384-445-489/422
CcAGO4.2	DDH/P	603-686-818/641
CcAGO4.3	DDH/P	601-684-816/639
CcAGO5	DDH/H	683-769-909/721
CcAGO7	DDH/H	750-823-963/788
CcAGO10.1	DDH/H	713-799-939/751
CcAGO10.2	DDH/Q	661-747-887/699
CcAGO16	DDH/S	634-725-857/672

*Motifs show the residues in *C. canephora* AGO proteins that correspond to D760, D845, H986/H798 of AtAGO1

In *C. canephora,* eight putative RDR proteins were found after BLASTP on the Coffee Genome Hub. Conserved domain analysis confirmed the presence of the RNA-dependent RNA polymerase (RdRP) domain, and Multiple Em for Motif Elicitation (MEME) (Version 4.11.2) [49] analysis revealed that six coffee RDR proteins possess a DLDGD motif and two possess a DFDGD motif (Fig 4). Multiple alignments with orthologs

sequences and phylogenetic tree analysis were also performed to assign the coffee RDR proteins and to determine the evolutionary relationship with the other angiosperm species. Four RDRs corresponded to RDR1, one to RDR2, one to RDR6, and two to RDR3 (Fig 5). The name, locus position, length, and identity of the CcRDR proteins with their respective orthologs from Arabidopsis are presented in Table 2.

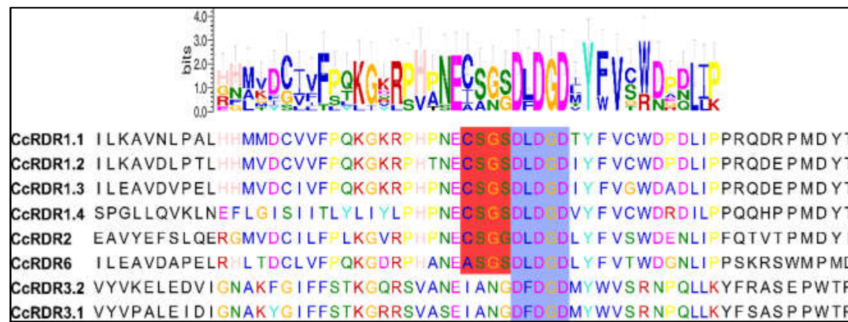


Fig 4. Analysis of the DxDGD catalytic motif of the RNA-dependent RNA polymerase (RdRP) conserved domain. Six coffee RDR possess a DLDGD motif (CcRDR1.1-1.4, CcRDR2 and CcRDR6) and two have the DFDGD motif (CcRDR3.1 and CcRDR3.2), corresponding to the RDRα clade and the RDRγ clade, respectively (Blue box). Additionally, the RDRα displays a conserved subsequences (C/A)SG(S/G) before the DLDGD motif and, all CcRDR1 and the CcRDR2 showed the CSGS sequence, while CcRDR6 showed the ASGS sequence (red box).

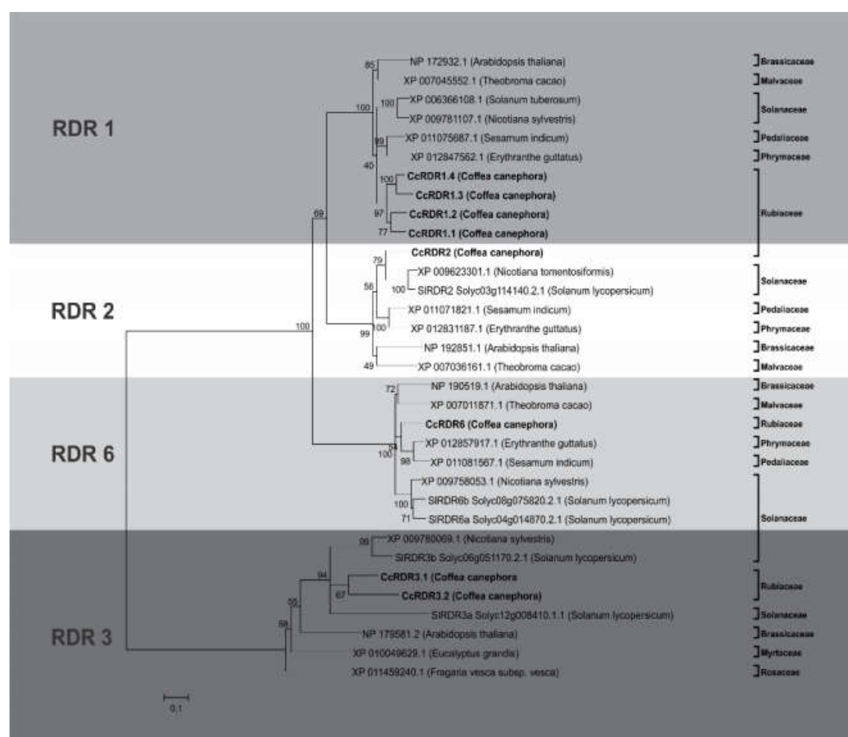


Fig 5. Phylogenetic tree of RDR proteins identified in *C. canephora.* Phylogenetic tree showing relationships between the paralogous and orthologs proteins of the RDR family. The evolutionary history was inferred using the Neighbor-Joining method [46]. The bootstrap consensus tree inferred from 2000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (2000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site [48]. The analysis involved 33 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 312 positions in the final dataset.

In Arabidopsis, the six RDR proteins are divided into four families: RDR1, RDR2, RDR3 (RDR3a and RDR3b), and RDR6 [63]. RDR1, RDR2, and RDR6 function in the formation of dsRNA from ssRNA sequences, which are processed into several types of siRNAs targeting specific endogenous loci

[64]. Among the six Arabidopsis RDR genes, AtRDR1, AtRDR2, and AtRDR6 are involved in processes such as viral resistance, chromatin silencing, and Post-Translational Gene Silencing (PTGS) [65]. The function of the RDR3 genes remains unknown, but the presence of at least one copy of the RDR3 gene in several plant genomes and other organisms suggests that these proteins may have functional significance [66].

In the phylogenetic tree, two main clades are observed, one consisting of RDR1, RDR2, and RDR6 and the other consisting of RDR3. This observation is consistent with the division of the two clades predicted based on their catalytic motifs (Fig 5). Although we found two RDR3 genes in *C. canephora*, similarly to tomato (SlRDR3a and SlRDR3b), the two CcRDR3 genes grouped with SlRDR3a (Fig 5).

To confirm the expression of the main RNA-silencing components, we searched the RNA-seq data of *Coffea canephora* publicly available in the Sequence Read Archive (SRA) of the NCBI (https://www.ncbi.nlm.nih.gov/sra/?term=ERP003741). Sequencing data of leaves collected at different development stages (young, expanded, and old) and stem tissues were analyzed to determine the expression profile of the sRNA silencing components identified in coffee, including *CcAGO, CcDCL, CcRDR, CcHYL1, CcSE, CcDDL, CcTG, CcHEN1,* and *CcHST*. The heatmap showed expression in all the tested tissues (Fig 6). However, Cufflinks analysis assigned three loci annotated as DCL2 in the coffee genome (Cc02_g14900, Cc02_g14910, and Cc02_g14920 – herein referred to as DCL2.2 and DCL2.3) as isoforms of the same genetic locus; therefore, these were not included in the heatmap (S3 Fig). Furthermore, CcAGO4.1 was not expressed in any of the tissues.

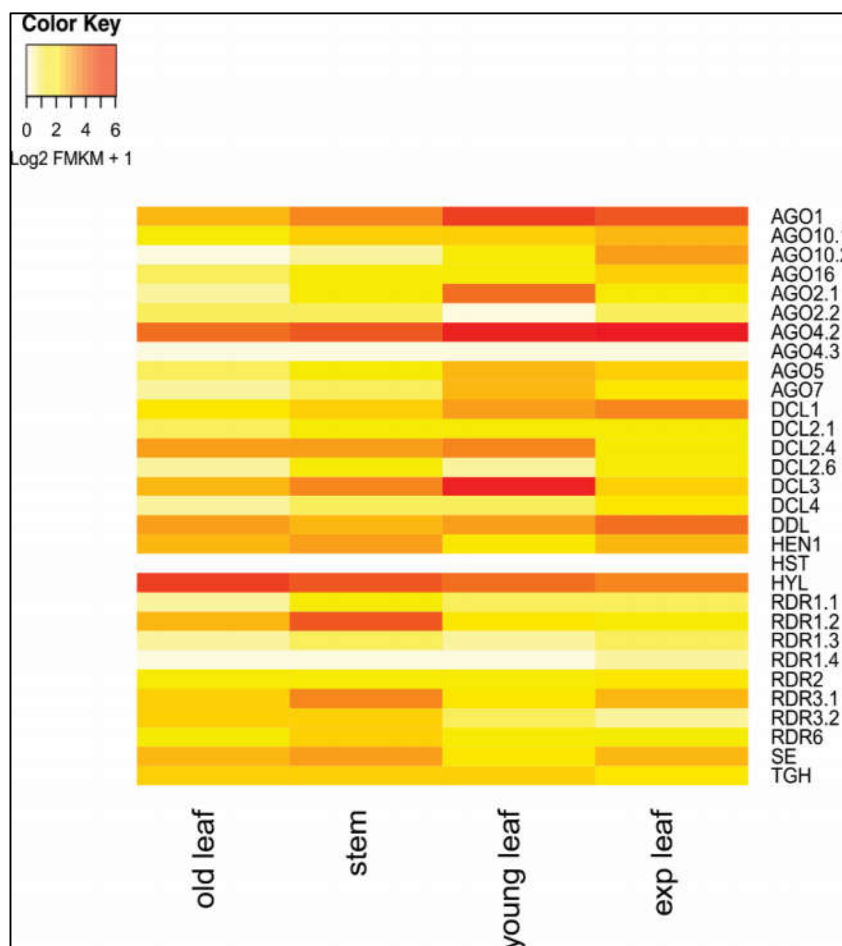


Fig 6. Validation of the main proteins of genes involved in the coffee RNA-guided silencing pathways from RNAseq libraries. Heatmap showing the expression pattern of the *C. canephora* RNA-silencing genes in three leaf developmental stages - Young, Expandend ("exp" in the figure), and Old - and Stem. (Transcriptome available at https://www.ncbi.nlm.nih.gov/sra/?term=ERP003741).

miRNAs and miRNA target prediction

Homology-based miRNA search was conducted by comparing plant miRNAs deposited in the miRBase database version 21 against the coffee genome. After applying filters to retrieve miRNA precursors, a total of 235 precursors and 317 mature miRNAs were identified and characterized,

belonging to 113 MIR families (S2 Table). The mature miRNAs were found in both the 3' and 5’ arms of the precursor, with sizes ranging from 19 to 25 nt, most of which were 21 nt (S2 Table). The preferred first 5’ nucleotide was Uracil (U). The location of the pre-miRNAs in the genome was determined, including the chromosome, start and end point, strand position, and genic/intergenic position (S2 Table). MIR genes were observed in all chromosomes, and chromosome 2 contained the highest number of MIR genes (36 genes). A total of 38 precursors were found either in antiparallel clusters or clustered with a maximum distance of 10 kb between the two miRNAs, but most were widespread throughout the chromosomes. A total of 193 precursors were identified in the intergenic regions, and the other 43 precursors were found within genes (S2 Table).

The precursor sizes varied from 68 to 338 nt, and the AU (Adenine+Uracil) content ranged from 41% to 69% (S3 Table). The thermodynamic aspects of the precursors - Minimal Free Energy (MFE), adjusted MFE (AMFE), MFE index (MFEI), Minimal Free Energy of the thermodynamic ensemble (MFEE), Ensemble Diversity (Diversity) and frequency of the MFE structure in the ensemble (Frequency) - were measured (S3 Table). The MFE ranged from -21.9 to -97.5 kcal mol^{-1}, with a mean of -56.4 kcal mol^{-1}; the AMFE ranged from -21.4 to -59.6 kcal mol^{-1}, with a mean of -36.46 kcal mol^{-1}; and the MFEI varied from 0.7 to 1.7, with a mean of 0.88.

We chose some of the highly conserved MIR families – *MIR156, MIR172,* and *MIR390* – for further characterization. We analyzed the conservation of their sequences and structure as well as their phylogenetic distributions. For each of these MIR families, multiple sequence alignment and secondary structure prediction were performed to verify the primary and secondary conservation relative to other plant species orthologs (Figs 7-9). These MIR families presented high conservation between their primary and secondary structures and their orthologs (Figs 7-9). A phylogenetic tree was created to verify the evolutionary distribution of each MIR family (Figs 7-9).

Fig 7. Alignment of pre-miRNA sequences (a), comparison of secondary structures (b) and phylogenetic tree (c) of ccp-MIR156 miRNAs and their orthologues. ccp- *Coffea canephora*, ath – *Arabidopsis thaliana*, nta – *Nicotiana tabacum*, mtr – *Medicago truncatula*, gma – *Glycine max*, mes – *Manihot esculenta*, ppe – *Prunus persica*, mdm – *Malus domestica*, vvi – *Vitis vinifera*, tcc - *Theobroma cacao*, ptc – *Populus trichocarpa*, aly – *Arabidopsis lyrata*, sly – *Solanum lycopersicum.* The evolutionary history was inferred using the Neighbor-Joining method[46]. The bootstrap consensus tree inferred from 5000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (5000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method [3] and are in the units of the number of base substitutions per site[47]. The analysis involved 23 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 68 positions in the final dataset.

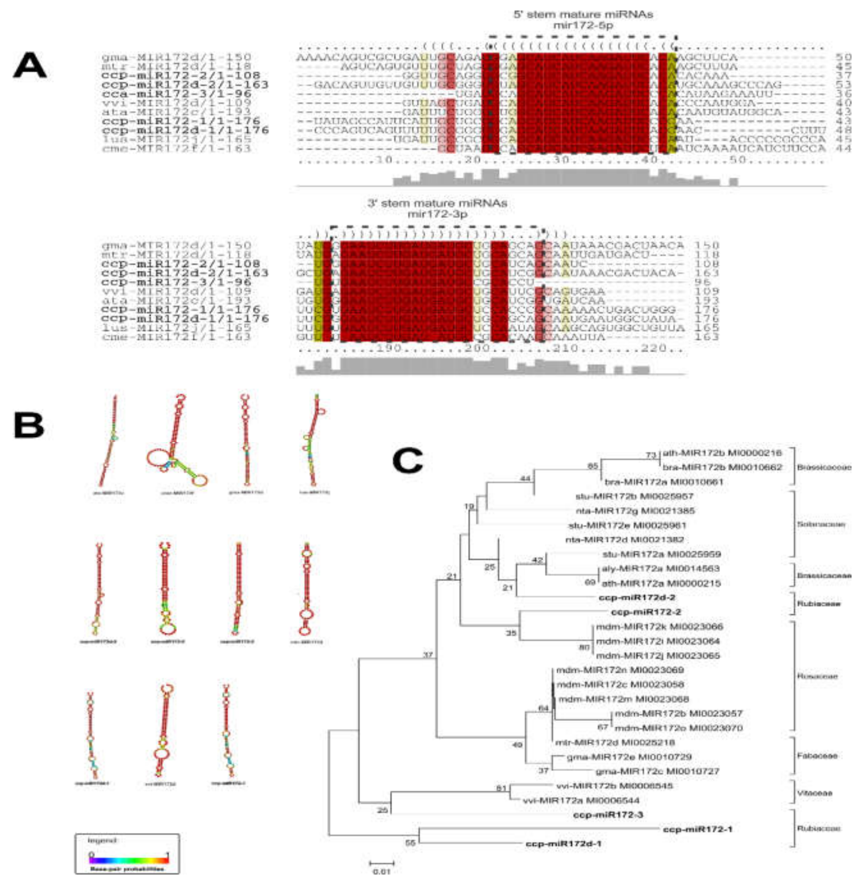


Fig 8. Alignment of pre-miRNA sequences (a), comparison of secondary structures (b) and phylogenetic tree (c) of ccp-MIR172 miRNAs and their orthologues. ccp- *Coffea canephora*, ath – *Arabidopsis thaliana*, cme - *Cucumis melo,* gma – *Glycine max,* lus - *Linum usitatissimum*, mtr – *Medicago truncatula*, vvi – *Vitis vinifera*, bra – *Brassica rapa*, stu – *Solanum tuberosum*, nta – *Nicotiana tabacum*, aly – *Arabidopsis lyrata*, mdm – *Malus domestica.* The evolutionary history was inferred using the Neighbor-Joining method[46]. The bootstrap consensus tree inferred from 5000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (5000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site[47]. The analysis involved 28 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 46 positions in the final dataset.

Fig 9. Alignment of pre-miRNA sequences (a), comparison of secondary structures (b) and phylogenetic tree (c) of ccp-MIR390 miRNAs and their orthologues. ccp- *Coffea canephora,* aly – *Arabidopsis lyrata*, ath – *Arabidopsis thaliana*, bna - Brassica napus, gma – *Glycine max*, ptc – *Populus trichocarpa,* sly – *Solanum lycopersicum,* mdm – *Malus domestica,* tcc - *Theobroma cacao,* lus - *Linum usitatissimum*. The evolutionary history was inferred using the Neighbor-Joining method[46]. The optimal tree with the sum of branch length = 1.87754489 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (5000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site[47]. The analysis involved 22 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 65 positions in the final dataset.

We also identified potential miRNA target genes using psRNATarget [67] based on the *C. canephora* genome. In total, 2239 genes were identified as potential targets of the miRNAs, many of which were targeted by more than one miRNA (S4 Table).

To classify and group the Gene Ontology (GO) classes of the miRNA targets, the web tool SEA (Singular Enrichment Analysis) from agriGO (http://bioinfo.cau.edu.cn/agriGO/index.php) was used [53]. A total of 1356 GO terms were annotated for the target genes in *C. canephora*, and these were summarized in 57 main terms. The genes belonging to the 25 overrepresented terms among the three GO categories, namely the biological process, molecular function, and cellular component categories, are presented (Fig 10).

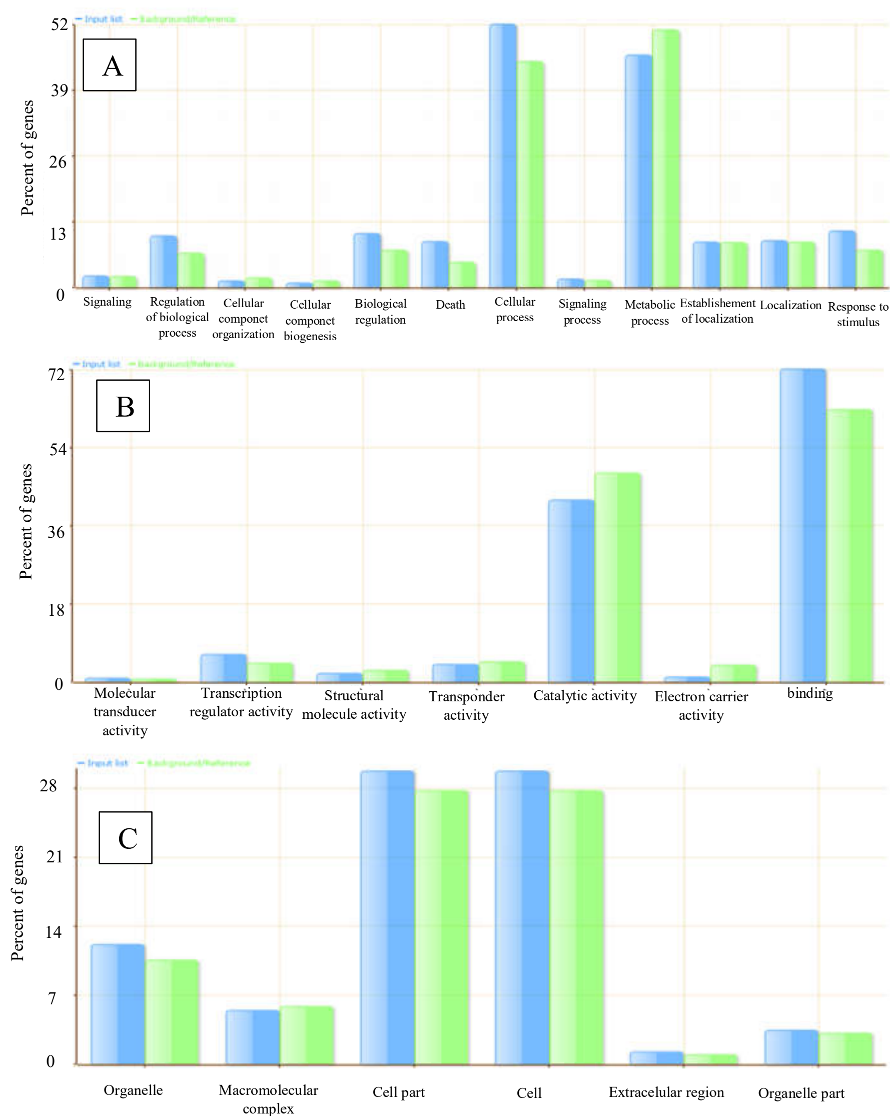


Fig 10. SEA (Singular Enrichment Analysis) of the GO terms of the predicted targets of the ccp-miRNAs. Biological process (A), molecular function (B) and cellular component (C).

We further identified the putative targets of ccp-MIR156, ccp-MIR172, and ccp-MIR390 in the RNA-seq libraries of stem and leaf tissues. The complete list of the targets assigned to these miRNAs is presented in S5 Table.

Discussion

Duplication events and domain and catalytic site configurations reveal insights into the sRNA pathway core members in *C. canephora*

Duplication of DCL2 has been observed in several species [56,68,69]. The largest of the six CcDCL2 members, CcDCL2.1, is located on chromosome 9 and is missing its DsRB (DSRM) domain. DCL2 usually contains only one DsRB (DSRM) domain, but in the four tomato DCL2s, only one member (SlDCL2d) possesses a DsRB (DSRM) domain [55]. The shortest CcDCL2 identified, CcDCL2.5 (354 aa), is located on chromosome 6, along with CcDCL2.6 (762 aa). Both of these proteins are truncated. Similar findings were observed for CcDCL2.2, CcDCL2.3, and CcDCL2.4, which are located sequentially on chromosome 2 and are also incomplete according to the current version of the genome annotation.

Expression analyses demonstrated that at least four DCL2-like genes are active in coffee (Fig 6 – S3 Fig), including the only complete sequence, CcDCL2.1. The other two DCL2 genes that are expressed are DCL2.4 (Cc02_g14930) and DCL2.6 (Cc06_g19980) (Fig 6). In addition to that, a total of seven isoforms were assigned to the same locus (Cc02_g14900) (S3 Fig). This might indicate misannotation of the three DCL2 assigned to the sequential loci at Chromosome 2 (Cc02_g14900, Cc02_g14910 and Cc02_g14920), which are probably exons of a unique gene. Finally, DCL2.5 (Cc06_g19770), which is the most incomplete DCL2 annotated in the genome, is not expressed in either tissue and could not be confirmed. Although it remains unclear how many DCL-like proteins are present and where on the genome their complete sequence can be found, an expansion of the DCL-like proteins appears to have occurred in *C. canephora* through the duplication of the DCL2-like family.

DCL-like proteins might contain the characteristic catalytic residues of RNase III domain-containing proteins [59]. The RNase III domains bind dsRNA and are responsible for cleavage and processing; therefore, they are essential to sRNA generation [58]. Only the incomplete CcDCL2 (CcDCL2.2-CcDCL2.6) proteins did not present the conserved residues (EDDE - Glu-Asp-Asp- Glu) in one or both RNAse III domains, reinforcing the need for further investigation into these short CcDCL2-like proteins.

The presence of CcAGO10, CcAGO2, and CcAGO4 paralogs indicates the occurrence of duplication events in the *C. canephora* genome. Gene duplication is one possible reason for the expansion of AGO proteins. The expansion of the AGO family in flowering plants suggests functional diversi cation of the AGO proteins [61].

PIWI domains contain the three conserved metal-chelating residue motif aspartate, aspartate, histidine (DDH). The DDH motif functions as a catalytic triad. A conserved histidine found at position 798 of AtAGO1 is also important for the catalytic function of the AGO proteins [62]. The four CcAGO proteins that possess the DDH/H motif (CcAGO1, CcAGO5, CcAGO7, and CcAGO10.1) potentially act as the slicer of RISC (Table 5). CcAGO2.1 and CcAGO2.2 showed a third aspartate residue instead of histidine, which was also observed in SlAGO2 [55], AtAGO2 and AtAGO3 [56]; GmAGO3a and SbAGO2 [34]; and OsAGO2 and OsAGO3 [56]. The absence of catalytic amino acids could prevent the processing of target RNA by cleavage; therefore, accessory factors for mediating mRNA turnover could be required [56]. However, the presence of a third aspartate in the triad restores the catalytic activity to function as slicer components of the silencing effector complexes in Arabidopsis and rice AGO2 and AGO3 [56].

In another four CcAGOs (CcAGO4.2, CcAGO4.3, CcAGO10.2, and CcAGO16), the conserved H798 residue has been replaced (Table 5). Previous studies showed variability in the H798 residue in monocots [54,56], while in tomato (*S. lycopersicum*), the H798 sites in the AGO4 group (SlAGO4a, b, c, d and SlAGO6) were replaced by proline [55]. In *C.*

canephora, which is closely related to *Solanaceae*, the H798 residue was also replaced in the AGO4 members, but in CcAGO10.2 and CcAGO16, the H798 residue was replaced by glutamine and serine, respectively.

CcAGO4.1 presented neither of the residues required for catalytic activity, which could represent either functionalization or loss of function. *CcAGO4.1* expression was not found in the RNA-seq libraries, corroborating the hypothesis that this protein is not active due to a lack of effective catalytic residues. However, AGO4 proteins can function either dependent on or independent of their catalytic activity [70]. The expression of *CcAGO4.2* and *CcAGO4.3* indicates that Transcriptional Gene Silencing (TGS) guided by RNA is upregulated in coffee because AGO4 has been implicated in RNA-Directed DNA Methylation (RdDM) [71].

In the RDR-like proteins, the RdRP domain contains a DxDGD catalytic motif [72]. RDR1, RDR2, and RDR6 (RDRα clade) share a DLDGD catalytic motif, whereas RDR3 (RDRγ clade) possesses a DFDGD motif [63,72]. The putative catalytic domains of the CcRDRs presented with the respective expected motifs of the α (CcRDR1.1-1.4, CcRDR2, and CcRDR6) and γ (CcRDR3.1 and CcRDR3.2) clades (Fig 4). Additionally, the RDRα clade displays a conserved subsequence (C/A)SG(S/G) upstream of the DLDGD motif [72], and all CcRDR1s and CcRDR2 present the CSGS sequence, whereas CcRDR6 possessed an ASGS sequence.

Interestingly, four RDR1 genes were found in *C. canephora*, all of which were located sequentially on chromosome 11 (Table 2). RDR1 is involved in plant defenses against biotic and abiotic components [17,73]. Most enriched GO terms in *C. canephora* belong to the defense response class [39]. It was also observed that the *C. canephora* genome includes several species-specific gene family expansions, including the defense-related genes [39]; this could also be the case for the RDR1 genes.

The *C. canephora* genome possesses several conserved and non-conserved *MIR* loci that target major cellular processes

Using a robust pipeline, we were able to significantly enrich the number of predicted miRNAs annotated in *Coffea spp* [35-39]. We identified 235 precursors and 317 mature sequences, whereas previous analyses of the coffee genome identified only 92 precursors [39]. The precursors belonged to 113 MIR families, representing a considerable increase relative to the 33 families originally described in the coffee genome report [39]. Our stringent and robust pipeline predicted sequences that were real miRNA precursors and identified more paralogous loci for several families already described.

The major MIR family was MIR171, with a total of 15 pre-miRNAs. Many highly conserved MIR families among plants were identified, including MIR171, MIR172, MIR156, MIR159, MIR160, MIR164, MIR167, MIR169, MIR390, and several others [74]. In contrast, some of the precursors identified belong to MIR families annotated for one species in miRBase v.21, such as ptc-MIR6476a (*Populus trichocarpa*) and stu-MIR8001b (*Solanum tuberosum*) [75,76].

Some of the most conserved families in plants, MIR156, MIR172, and MIR390 [43], have been identified in several species [33,43,75-77] and play central roles in plant development and stress responses. For instance, miR156 targets SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) transcription factor family members, and miR156-SPL networks define an essential regulatory module that controls phase transitions, leaf trichome development, male fertility, embryonic patterning, and anthocyanin biosynthesis [78-82]. In the *C. canephora* genome, miR156 has 24 putative targets (S4 Table). Based on the transcriptomes of the stem and leaf tissue samples, we found that miR156 potentially targets SPL-6 and SPL-12 in both tissues (S5 Table). In total, 15 putative targets were identified in the stems and

leaves, some of which were identified either in both tissues or in only one (S5 Table).

The MIR172 family consists of five precursors and ten mature miRNAs (S2 Table). This highly conserved family is found in several species and is related to the regulation of flowering time and floral organ identity by targeting APETALA2-like transcription factors in Arabidopsis [83,84]. miR172 acts downstream of miR156 to regulate phase transition [84], as an increase in miR156 levels corresponds to lower expression of miR172 and vice versa in several species [84-87]. In the *C. canephora* genome, 118 putative targets for miR172 were identified (S4 Table). Based on the transcriptome data, a total of 66 putative targets were identified , including AP2 in stem tissue (S5 Table).

miR390 is involved in the regulation of development and the response to several stresses [88-91]. Among its targets, miR390 regulates the Auxin Response Factor (ARF) by mediating non-protein coding Trans-Acting siRNA locus 3 (TAS3) generation in an AGO7-dependent manner [92]. miR390 also targets Leucine-Rich Repeat Receptor-like kinases (LRK) and regulates a LRK protein in *Oryza sativa* in response to cadmium stress [91]. In the *C. canephora* genome, 11 putative targets were identified (S4 Table). Four putative targets were identified in the transcriptomes of stems and leaves (S5 Table), among which a LRK (RKF1) was identified in both tissues (S5 Table).

The ccp-MIR156, ccp-MIR172, and ccp-MIR390 members were highly conserved in their primary and secondary structures relative to their respective orthologs from other species and relative to their distributions within the phylogenetic tree in a clade of Eudicotyledons, consistent with plant phylogeny (Figs 7-9) [93].

The GO terms of the putative *C. canephora* miRNA targets were categorized and compared with the GO terms of the whole genome as background (Fig 10). In total, 1356 GO terms were assigned to the putative targets, including a total of 14975 GO terms annotated to the genome. The

main overrepresented subcategories belonging to the ‘Biological Process’ category were ‘cellular process’ and ‘metabolic process’. In the ‘Cellular Component’ category, the main overrepresented terms were ‘cell part’ and ‘cell’. In the ‘Molecular Function’ category, the main overrepresented terms were ‘catalytic activity’ and ‘binding’. Interestingly, the main categories of the potential targets were also the main categories annotated for the genome (green bars – Fig 10). Therefore, one can infer that miRNAs in *C. canephora* target major cellular processes.

Considering the importance of this pioneering work, we elucidated several aspects of sRNAs in *C. canephora*, which offers a significant step towards a better understanding of the transcriptional and post-transcriptional regulation of this major crop. An understanding of the sRNA pathways in coffee provides insights for plant breeding through genetic engineering technology.

Acknowledgements

The authors thank the members of the Laboratory of Plant Molecular Physiology (LFMP) of the Federal University of Lavras (UFLA) for helping with the data mining and organization. We also thank the Laboratory of Bioinformatics and Molecular Analysis (LBAM) of the Federal University of Uberlândia (UFU) – Campus Patos de Minas, for providing computational structure for analyses.

References

1. Brodersen P, Voinnet O (2006) The diversity of RNA silencing pathways in plants. Trends in Genetics 22: 268-280.

2. Axtell MJ (2013) Classification and comparison of small RNAs from plants. Annu Rev Plant Biol 64: 137-159.

3. Chen X (2009) Small RNAs and Their Roles in Plant Development. Annual Review of Cell and Developmental Biology 25: 21-44.

4. Borges F, Martienssen RA (2015) The expanding world of small RNAs in plants. Nat Rev Mol Cell Biol 16: 727-741.

5. Kim YJ, Zheng B, Yu Y, Won SY, Mo B, et al. (2011) The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. Embo j 30: 814-822.

6. Tang G (2010) Plant microRNAs: an insight into their gene structures and evolution. Semin Cell Dev Biol 21: 782-789.

7. Yu B, Bi L, Zheng B, Ji L, Chevalier D, et al. (2008) The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. Proc Natl Acad Sci U S A 105: 10073-10078.

8. Kurihara Y, Takashi Y, Watanabe Y (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. Rna 12: 206-212.

9. Dong Z, Han MH, Fedoroff N (2008) The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. Proc Natl Acad Sci U S A 105: 9970-9975.

10. Lobbes D, Rallapalli G, Schmidt DD, Martin C, Clarke J (2006) SERRATE: a new player on the plant microRNA scene. EMBO Rep 7: 1052-1058.

11. Ren G, Xie M, Dou Y, Zhang S, Zhang C, et al. (2012) Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis. Proc Natl Acad Sci U S A 109: 12817-12821.

12. Rogers K, Chen X (2013) Biogenesis, Turnover, and Mode of Action of Plant MicroRNAs. Plant Cell 25: 2383-2399.

13. Yamaguchi A, Abe M (2012) Regulation of reproductive development by non-coding RNA in Arabidopsis: to flower or not to flower. J Plant Res 125: 693-704.

14. Hu W, Wang T, Xu J, Li H (2014) MicroRNA mediates DNA methylation of target genes. Biochemical and Biophysical Research Communications 444: 676-681.

15. Li J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol 15: 1501-1507.

16. Zeng Y, Cullen BR (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. Nucleic Acids Res 32: 4776-4785.

17. Bologna NG, Voinnet O (2014) The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in Arabidopsis. Annual Review of Plant Biology 65: 473-503.

18. Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R (2005) MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat Cell Biol 7: 719-723.

19. Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet 15: 394-408.

20. Schiebel W, Haas B, Marinkovic S, Klanner A, Sanger HL (1993) RNA-directed RNA polymerase from tomato leaves. II. Catalytic in vitro properties. J Biol Chem 268: 11858-11867.

21. Moissiard G, Parizotto EA, Himber C, Voinnet O (2007) Transitivity in Arabidopsis can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins. Rna 13: 1268-1278.

22. Cao X, Jacobsen SE (2002) Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing.

23. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11: 204-220.

24. Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O, et al. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell 120: 613-622.

25. Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, et al. (2012) Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. Genes & Development 26: 1825-1836.

26. Zhou M, Law JA (2015) RNA Pol IV and V in gene silencing: Rebel polymerases evolving away from Pol II's rules. Current Opinion in Plant Biology 27: 154-164.

27. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet 41: 630-634.

28. Matzke MA, Kanno T, Matzke AJM (2015) RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. Annual Review of Plant Biology 66: 243-267.

29. Speth C, Willing EM, Rausch S, Schneeberger K, Laubinger S (2013) RACK1 scaffold proteins influence miRNA abundance in Arabidopsis. Plant J 76: 433-445.

30. Jeong IS, Aksoy E, Fukudome A, Akhter S, Hiraguri A, et al. (2013) Arabidopsis C-terminal domain phosphatase-like 1 functions in miRNA accumulation and DNA methylation. PLoS One 8: e74739.

31. Karlsson P, Christie MD, Seymour DK, Wang H, Wang X, et al. (2015) KH domain protein RCF3 is a tissue-biased regulator of the plant miRNA biogenesis cofactor HYL1. Proceedings of the National Academy of Sciences 112: 14096-14101.

32. Kapoor M, Arora R, Lama T, Nijhawan A, Khurana J, et al. (2008) Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. BMC Genomics 9: 451.

33. de Sousa Cardoso TC, Portilho LG, de Oliveira CL, McKeown PC, Maluf WR, et al. (2016) Genome-wide identification and in silico characterisation of microRNAs, their targets and processing pathway genes in Phaseolus vulgaris L. Plant Biol 18: 206-219.

34. Liu X, Lu T, Dou Y, Yu B, Zhang C (2014) Identification of RNA silencing components in soybean and sorghum. BMC Bioinformatics 15: 4.

35. Loss-Morais G, Ferreira DCR, Margis R, Alves-Ferreira M, Corrêa RL (2014) Identification of novel and conserved microRNAs in Coffea canephora and Coffea arabica. Genetics and Molecular Biology 37: 671-682.

36. Rebijith KB, Asokan R, Ranjitha HH, Krishna V, Nirmalbabu K (2013) In silico mining of novel microRNAs from coffee (Coffea arabica) using expressed sequence tags. Journal of Horticultural Science and Biotechnology 88: 325-337.

37. Akter A, Islam MM, Mondal SI, Mahmud Z, Jewel NA, et al. (2014) Computational identification of miRNA and targets from expressed sequence tags of coffee (Coffea arabica). Saudi Journal of Biological Sciences 21: 3-12.

38. Chaves SS, Fernandes-Brum CN, Silva GF, Ferrara-Barbosa BC, Paiva LV, et al. (2015) New Insights on Coffea miRNAs: Features and Evolutionary Conservation. Appl Biochem Biotechnol 177: 879-908.

39. Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, et al. (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 345: 1181-1184.

40. de Souza Gomes M, Muniyappa MK, Carvalho SG, Guerra-Sa R, Spillane C (2011) Genome-wide identification of novel microRNAs and their target genes in the human parasite Schistosoma mansoni. Genomics 98: 96-111.

41. Smit AFA, Hubley R, Green P RepeatMasker at http://repeatmasker.org. Accessed 20 January 2016.

42. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, et al. (2009) Rfam: updates to the RNA families database.

43. Zhang B, Pan X, Cannon CH, Cobb GP, Anderson TA (2006) Conservation and divergence of plant microRNA genes. Plant J 46: 243-259.

44. Zhang BH, Pan XP, Cox SB, Cobb GP, Anderson TA (2006) Evidence that miRNAs are different from other RNAs. Cellular and Molecular Life Sciences CMLS 63: 246-254.

45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

46. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425.

47. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111-120.

48. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739.

49. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2: 28-36.

50. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644-652.

51. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9: 868-877.

52. Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 39.

53. Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38: W64-70.

54. Qian Y, Cheng Y, Cheng X, Jiang H, Zhu S, et al. (2011) Identification and characterization of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in maize. Plant Cell Rep 30: 1347-1363.

55. Bai M, Yang GS, Chen WT, Mao ZC, Kang HX, et al. (2012) Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in Solanum lycopersicum. Gene 501: 52-62.

56. Kapoor M, Arora R, Lama T, Nijhawan A, Khurana JP, et al. (2008) Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. BMC Genomics 9: 1-17.

57. Liu B, Li P, Li X, Liu C, Cao S, et al. (2005) Loss of function of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. Plant Physiol 139.

58. Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, et al. (2006) The evolution and diversification of Dicers in plants. FEBS Letters 580: 2442-2450.

59. Ji X (2008) The Mechanism of RNase III Action: How Dicer Dices. In: Paddison PJ, Vogt PK, editors. RNA Interference. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 99-116.

60. Vaucheret H (2008) Plant ARGONAUTES. Trends in Plant Science 13: 350-358.

61. Zhang H, Xia R, Meyers BC, Walbot V (2015) Evolution, functions, and mysteries of plant ARGONAUTE proteins. Current Opinion in Plant Biology 27: 84-90.

62. Baumberger N, Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci USA 102.

63. Wassenegger M, Krczal G (2006) Nomenclature and functions of RNA-directed RNA polymerases. Trends in Plant Science 11: 142-151.

64. Voinnet O (2008) Use, tolerance and avoidance of amplified RNA silencing by plants. Trends Plant Sci 13: 317-328.

65. Wang XB, Wu Q, Ito T, Cillo F, Li WX, et al. (2010) RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. Proc Natl Acad Sci U S A 107: 484-489.

66. Willmann MR, Endres MW, Cook RT, Gregory BD (2011) The Functions of RNA-Dependent RNA Polymerases in Arabidopsis. The Arabidopsis Book / American Society of Plant Biologists 9: e0146.

67. Dai XB, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Research 39: W155-W159.

68. Liu H, Guo S, Xu Y, Li C, Zhang Z, et al. (2014) OsmiR396d-Regulated OsGRFs Function in Floral Organogenesis in Rice through Binding to Their Targets OsJMJ706 and OsCR4. Plant Physiology 165: 160-174.

69. Tworak A, Urbanowicz A, Podkowinski J, Kurzynska-Kokorniak A, Koralewska N, et al. (2016) Six Medicago truncatula Dicer-like protein genes are expressed in plant cells and upregulated in nodules. Plant Cell Reports 35: 1043-1052.

70. Qi Y, He X, Wang XJ, Kohany O, Jurka J, et al. (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature 443: 1008-1012.

71. Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science 299: 716-719.

72. Zong J, Yao X, Yin J, Zhang D, Ma H (2009) Evolution of the RNA-dependent RNA polymerase (RdRP) genes: Duplications and possible losses before and after the divergence of major eukaryotic groups. Gene 447: 29-39.

73. Zhang C, Wu Z, Li Y, Wu J (2015) Biogenesis, Function, and Applications of Virus-Derived Small RNAs in Plants. Frontiers in Microbiology 6: 1237.

74. Axtell MJ, Bartel DP (2005) Antiquity of MicroRNAs and Their Targets in Land Plants. The Plant Cell 17: 1658-1673.

75. Puzey JR, Karger A, Axtell M, Kramer EM (2012) Deep annotation of Populus trichocarpa microRNAs from diverse tissue sets. PLoS One 7: e33034.

76. Zhang R, Marshall D, Bryan GJ, Hornyik C (2013) Identification and Characterization of miRNA Transcriptome in Potato by High-Throughput Sequencing. PLoS ONE 8: e57233.

77. Liang G, Li Y, He H, Wang F, Yu D (2013) Identification of miRNAs and miRNA-mediated regulatory pathways in Carica papaya. Planta 238: 739-752.

78. Wang JW, Czech B, Weigel D (2009) miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana. Cell 138: 738-749.

79. Wang Y, Wang Z, Amyot L, Tian L, Xu Z, et al. (2015) Ectopic expression of miR156 represses nodulation and causes morphological and developmental changes in Lotus japonicus. Mol Genet Genomics 290: 471-484.

80. Xing S, Salinas M, Hohmann S, Berndtgen R, Huijser P (2010) miR156-targeted and nontargeted SBP-box transcription factors act in concert to secure male fertility in Arabidopsis. Plant Cell 22: 3935-3950.

81. Yu N, Cai WJ, Wang S, Shan CM, Wang LJ, et al. (2010) Temporal control of trichome distribution by microRNA156-targeted SPL genes in Arabidopsis thaliana. Plant Cell 22: 2322-2335.

82. Ostria-Gallardo E, Ranjan A, Chitwood DH, Kumar R, Townsley BT, et al. (2016) Transcriptomic analysis suggests a key role for SQUAMOSA PROMOTER BINDING PROTEIN LIKE, NAC and YUCCA genes in the heteroblastic development of the temperate rainforest tree Gevuina avellana (Proteaceae). New Phytologist 210: 694-708.

83. Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. Plant Cell 15: 2730-2741.

84. Wu G, Park MY, Conway SR, Wang JW, Weigel D, et al. (2009) The sequential action of miR156 and miR172 regulates developmental timing in Arabidopsis. Cell 138.

85. Belli Kullan J, Lopes Paim Pinto D, Bertolini E, Fasoli M, Zenoni S, et al. (2015) miRVine: a microRNA expression atlas of grapevine based on small RNA sequencing. BMC Genomics 16: 1-23.

86. Chuck G, Cigan AM, Saeteurn K, Hake S (2007) The heterochronic maize mutant Corngrass1 results from overexpression of a tandem microRNA. Nat Genet 39.

87. Zhu QH, Helliwell CA (2011) Regulation of flowering time and floral patterning by miR172. J Exp Bot 62: 487-495.

88. Sunkar R, Girke T, Jain PK, Zhu JK (2005) Cloning and characterization of microRNAs from rice. Plant Cell 17: 1397-1411.

89. Sunkar R, Li YF, Jagadeeswaran G (2012) Functions of microRNAs in plant stress responses. Trends in Plant Science 17: 196-203.

90. Chen L, Wang T, Zhao M, Tian Q, Zhang WH (2012) Identification of aluminum-responsive microRNAs in Medicago truncatula by genome-wide high-throughput sequencing. Planta 235: 375-386.

91. Ding Y, Ye Y, Jiang Z, Wang Y, Zhu C (2016) MicroRNA390 Is Involved in Cadmium Tolerance and Accumulation in Rice. Front Plant Sci 7: 235.

92. Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, et al. (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell 133: 128-141.

93. Stevens PF (2001 onwards) Angiosperm Phylogeny Website. Version 12, July 2012 [and more or less continuously updated since]. http://wwwmobotorg/MOBOT/research/APweb/.

Supporting Information

The supporting information is available at:

http://www.dbi.ufla.br/lfmp/Material_Suplementar_Fernandes-Brum.zip

ARTICLE 2 - Characterization and Profiling of the small RNAs transcriptome of two phases of flowering from two cultivars of *Coffea arabica*

Article prepared in the Plant Molecular Biology periodic style for future submission to this journal

Characterization and Profiling of the small RNAs transcriptome of two phases of flowering from two cultivars of *Coffea arabica*

Christiane Noronha Fernandes-Brum, Sandra M. Mathioni, Atul Kakrana, Blake C. Meyers, Matheus de Souza Gomes, Antonio Chalfun-Júnior

Abstract

Coffee is one of the most important crops in the world and the second most globally traded commodity, representing an important source of income in several countries. After reaching the stage of development of approximatelly 3.1-6.0 mm in length (G4 stage), buds enter in a dormancy period, which coincides with the dry and cold season in Brazil. After re-watering, the buds that achieved the G4 stage grow rapidly to reach 6-10 mm (G5 – light green; and G6 – white), and anthesis is observed within 10-12 days. Little is known about the molecular networks involved in the regulation of coffee flowering and bud dormancy release. microRNAs (miRNAs) are 21-24 nt sRNAs that guide mRNA cleavage or translational repression of their target mRNAs. Several miRNAs are involved in the control of flowering time and flower development. To better understand the role miRNAs play in *C. arabica* flowering and bud dormancy release, we assessed the sRNA transcriptome of two contrasting cultivars, 'Red Catuaí IAC 144' and 'Siriema VC4', in two stages of flower development, during dormancy (G4) and after regrowth (G5). We identified a total 155 miRNAs, 17 of which were differentially expressed. Precursors loci, secondary structure of differentially expressed and their potential targets were predicted using *C. canephora*, the

closest species with sequenced genome. This diverse set of miRNAs is an important resource for the investigation of the mechanisms of flower maturation, aiming at the comprehension of the bud dormancy and resume of growth in *C. arabica*.

Introduction

Coffee is one of the most important crops in the world and the second most worldwide traded commodity, representing an important source of income in several countries. Brazil is the largest consumer and the world's largest coffee producer, responsible for 36% of coffee production in 2016 (ICO 2016a). It is also the main exporter, with 32% of total world coffee coming from Brazil, totalizing 34 mi bags exported (ICO 2016b). Only two coffee species are commercially important, *Coffea arabica* and *Coffea canephora*, representing 64% and 36% of the world's coffee production, respectively (ICO 2016b). More than 84% of the Brazilian coffee production is from Arabica coffee. The total Brazilian production in 2016 was of around 52 mi 60kg bags, 41.29 mi of these corresponding to *C. arabica* (CONAB 2016).

Coffee plants show peculiar features such as asynchronous flowering and floral bud dormancy (Camargo and Franco 1985). Asynchronous development is observed in vegetative buds before flower induction and in flower meristems within each bud during flower development (Oliveira et al. 2014), what leads to sequential flowering and uneven fruit ripening stages. After reaching the stage of development of around 3.1-6 mm (named as G4 stage) (Morais et al. 2008), in which all the floral whorls have already been formed (Oliveira et al. 2014), the buds enter a dormancy period (Rena and Maestri 1985). This period coincides with the dry and cold season in Brazil, from June to August. During this period, buds that are in lower stages of development (due to asynchronous growth) can reach the dormancy stage (G4) (Rena and Maestri 1985). After the return of water availability, through rain or irrigation, the buds that achieved the G4 stage grow rapidly to reach 6-

10 mm (G5 – light green buds ; and G6 – white buds) (Morais et al. 2008), and flower opening is observed within 10-12 days (Camargo 1985; Rena and Maestri 1985). A slightly low water potential period followed by the return of watering, through either rainfall or irrigation can induce a main blossoming event (Crisosto et al. 1992), what can mitigate asynchronous flowering and fruit ripening. On the other hand, severe drought events and/or irregular rainfall periods caused by climate changes can cause losses in quality and in productivity (Camargo 2010). Therefore, the comprehension of the regulatory networks of coffee flowering can provide insights into the key regulators and provide tools for the enhancement of this crop.

Some of the molecular aspects of coffee flower development and the quiescent period have been recently investigated. For instance, the MADS-box transcription factors, the main gene family responsible for floral organ identity (Coen and Meyerowitz 1991), were identified and characterized in coffee (de Oliveira et al. 2010; Oliveira et al. 2014). Peculiar expression patterns of MADS-box family members are thought to be responsible for some specificities in coffee floral development, such as epipetalous stamen and mucilage secretion (Oliveira et al. 2014). Furthermore, the expression of flowering repressors, *FLC* and *SVP1*, during flower development, specially *CaFLC* in later stages, suggests involvement in dormancy of floral buds (Oliveira et al. 2014). Another important aspect is the resume of growth towards anthesis after re-watering the plants. An ethylene burst is proposed to be the main regulator of drought-re-watering based promotion of floral buds dormancy break (Lima 2015). Additionally, studies have shown that positive regulation of ethylene biosynthesis and signaling genes induces anthesis in *C. arabica* (Lima 2015). Further investigation of the molecular networks involved in the regulation of coffee flowering will provide a thorough comprehension of this crucial development stage.

Small RNAs (sRNAs) are of non-coding RNAs involved in transcriptional and/or post-transcriptional mediation of gene silencing, in association with an ARGONAUTE (AGO) protein in a RNA-induced

silencing complex (RISC) (Voinnet 2009). microRNAs (miRNAs) are 20-24 nt sRNAs that guide mRNA cleavage or translational repression of their target mRNA by the RISC (Mallory and Vaucheret 2006; Vaucheret 2006). MIR genes are transcribed by RNA polymerase II (Pol II) (Kim et al. 2011) and undergo several modifications from transcription stage to the mature stage. The primary transcripts (pri-miRNAs) are processed by the endonuclease activity of DICER-LIKE1 (DCL1) (Kurihara et al. 2006) into precursors (pre-miRNAs), assisted by other proteins like HYPONASTIC LEAVES 1 (HYL1) (Kurihara et al. 2006), SERRATE (SE) (Dong et al. 2008; Lobbes et al. 2006) and TOUGH (TGH) (Ren et al. 2012). The pre-miRNAs are then processed by the DCL complex to form a 21-24 nt duplex structure, containing two 3'overhanging nucleotides at each end. The duplex is 3' methylated by the methyltransferase HUA ENHANCER1 (HEN1), protecting the duplex from further modification and degradation (Li et al. 2005). The exportin HASTY (HST) is responsible for binding the duplex and transport it from the nucleus to the cytoplasm (Zeng and Cullen 2004). In the cytoplasm, one of the strands is incorporated to an ARGONAUTE (AGO) protein, containing the PAZ and PIWI domains, to form the RISC complex (RNA Induced Silencing Complex). The PIWI domain has an endonuclease activity capable of cleaving the target mRNA, which is recognized by a nearly perfect complementarity with the miRNA (Liu et al. 2005; Rogers and Chen 2013).

miRNAs participate in several important regulatory processes, such as vegetative growth, phase change, reproductive development, response to salinity, drought, and other stresses (Eldem et al. 2012; Gentile et al. 2015; Shuai et al. 2013; Spanudakis and Jackson 2014; Wang et al. 2011; Xie et al. 2014; Yamaguchi and Abe 2012). miRNAs are well studied and characterized in model species such as Arabidopsis and rice (Reinhart et al. 2002; Sun 2012; Sunkar et al. 2005), and with the advancement of high-throughput sequencing technologies, it has become possible to analyze non-model species to discover not only conserved but also less-conserved and species-specific miRNAs

(Belli Kullan et al. 2015; Chen et al. 2016; Liu et al. 2015; Roy et al. 2016; Tian et al. 2014).

Flowering is one of the developmental processes in which miRNAs are involved (Yamaguchi and Abe 2012). Among them, are the conserved miR156, miR159, miR160, miR164, miR166/165, miR167, miR169, miR172, and miR319 (Luo et al. 2013), which are related either in phase transition by determination of flowering time and/or floral organ identity and development (Achard et al. 2004; Aukerman and Sakai 2003; Jung et al. 2007; Lee et al. 2010; Rubio-Somoza and Weigel 2013; Schommer et al. 2012; Spanudakis and Jackson 2014; Wang et al. 2009; Wollmann et al. 2010; Yamaguchi and Abe 2012; Zhang et al. 2015; Zhou and Wang 2013). Novel miRNAs have also been identified in flowers by deep-sequencing crop and ornamental species, and in many cases are suggested to play important roles in specific morphological and/or physiological aspects in these species, such as flower shape, color, development and opening process (Aceto et al. 2014; Belli Kullan et al. 2015; Li et al. 2015; Roy et al. 2016; Wang et al. 2014).

In this study, with the goal of better understanding flower development in *Coffea arabica*, we generated sRNA libraries from two stages of flower development (G4 and G5) from two cultivars – 'Red Catuaí IAC 144' and 'Siriema VC4', which are contrasting in several aspecs. 'Red Catuaí IAC 144' is a widely cultivated cultivar in Brazil, of late maturation cycle and sensitive to deseases and drought (de Carvalho et al. 2008), while 'Siriema VC4' is a recently released cultivar (2015), with early reproductive cycle and tolerance to diseases and drought (de Carvalho et al. 2008; Matiello et al. 2015).

A total of 155 miRNAs were predicted, from which 17 were differentially expressed between developmental stages, and no differential expression between cultivars was observed. The detailed investigation of this diverse set of miRNAs will provide important insights into the flower maturation, more specifically in the bud dormancy and resume of growth mechanisms in *C. arabica.*

Methods

Plant materials

Plants of two cultivars of *C. arabica*, ‘Siriema VC4’ and ‘Red Catuaí IAC 144’, with 3 and 5 years old, respectively, grown at the experimental field of the Federal University of Lavras (UFLA), were used in this experiment. The samples were harvested in the month of September/2014 from three plants of each cultivar at the G4 (3.1-6.0 mm) and G5 (6.0-10.0 mm – light green color) stages, following the classification of Morais et al. (2008) (Figure 1). All samples were immediately frozen in liquid nitrogen and stored at 80 °C until extraction of total RNA was performed.

Figure 1. Coffee flower development according to Morais et al. (2008). G4 - Floral buds ranging from 3.1 to 6.0 mm in length. G5 - Floral buds ranging from 6.1 to 10.0 mm in length (light green color)

Small RNA library construction and Illumina sequencing

Total RNA was isolated with PureLink® Plant RNA Reagent (Invitrogen; Cat#12322-012) according to the manufacturer’s instructions with modifications, and quality and quantity of RNA was verified using agarose gel electrophoresis and NanoVue (GE Healthcare Life Sciences) spectrophotometer. Size selection for 20-30 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) as protocol previously described (Mathioni et al., 2017). All libraries were single-end

sequenced with 51 cycles on an Illumina HiSeq 2500 instrument at the University of Delaware Sequencing and Genotyping Center in the Delaware Biotechnology Institute.

Deep sequencing data analysis

The libraries were pre-processed using the script "prepro.py" version 0.2 (https://github.com/atulkakrana/helper.github) with default settings as described earlier (Patel et al. 2016). Preprocessing included trimming of 3' adapters followed by retention of only reads between 18 and 34 nt for downstream analyses. These reads were then aligned to the *C. canephora* genome (v1.0) using Bowtie (version - 0.12.8) with no allowed mismatches. Mapped reads from all these sRNA libraries were then normalized to empirically derived, 30 million reads base depth (transcript per 30 million, TP30M).

Mapped sRNA reads from all libraries were used as input to two different computational pipelines for 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 modified version of miREAP (https://sourceforge.net/projects/mireap/) and CentroidFold (Sato et al. 2009).

In '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 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 ratio of all reads mapped to sense strand against total reads mapped to both strands. In addition to strand bias, abundance bias was computed as ratio of two most abundant reads against all the reads mapped to same precursor. Candidate precursors with strand bias ≥ 0.9 and abundance bias ≥ 0.7 were selected, and foldback structure for precursor was predicted using CentroidFold. Each precursor was manually

inspected to match 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 matched to any known miRNA were considered as lineage or species-specific. In '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. Stem-loop structure of candidate precursors was visually inspected, 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.

Differential expression analysis of miRNAs between G5/G4 buds

The fold change was calculated with the formula $\log_2$(G5/G4) and the significance (p-value) was calculated with the Student's Two-Tailed Test (T-Test) considering the samples unpaired. Then, False Discovery Rate (FDR) (Benjamini and Hochberg 1995) was calculated. The miRNAs were considered differentially expressed when FDR ≤ 0.05 and $\log_2$(G5/G4) > 1. We tested Siriema (G5/G4), Catuaí (G5/G4) and Siriema+Catuaí (G5/G4).

Prediction of secondary structures and potential targets of differentially expressed genes

The secondary structures of pre-miRNA, diversity, MFE, frequency ensemble, and MFE were predicted using RNA-fold software (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi).

To search the putative target genes of the predicted miRNA, transcripts from the *C. canephora* genome (CDS+UTR) were retrieved from the Coffee Genome Hub (http://coffee-genome.org/download), and miRNAs target genes were predicted using the webtool psRNATarget [17]. To avoid false positive prediction rates for the miRNA target genes, we used a stringent

cut-off threshold for Maximum expectation of 2.0. The other parameters used were the default settings: length for complementarity scoring (hspsize) – 20 bp; top target genes for each small RNA – 200; target accessibility – allowed maximum energy to unpair the target site (UPE) – 25; anking length around target site for target accessibility analysis – 17 bp upstream/13 bp downstream; range of central mismatch leading to translational inhibition: 9–11 nt.

GO classes of the miRNA targets were classified and grouped using the web tool SEA (Singular Enrichment Analysis) from agriGO (http://bioinfo.cau.edu.cn/agriGO/index.php) (Du et al. 2010). The input was the target genomic IDs, which were compared against all the IDs of the Coffee Genome Hub.

Results and discussion

Deep sequencing data analysis

In this study, sRNA libraries from floral buds in two stages of development of two coffee cultivars were constructed and sequenced using an Illumina HiSeq 2500. After trimming adapter sequences, the total reads were mapped to the *C. canephora* genome assembly (v1.0), the closest species to *C. arabica* with the sequenced genome. The total sRNA reads varied from 18,565,555 to 37,037,953 (Table 1). Then, the reads which matched to the genome ranged from 13,750,793 to 28,515,365, and distinct reads which matching to the genome varied from 1,957,856 to 5,405,988 (Table 1).

Table 1. Summary of sRNA reads from libraries prepared from G4 and G5 buds from Siriema VC4 and Red Catuaí IAC 144 coffee cultivars. Genome matched reads reflects the reads mapped to the *C. canephora* genome. Distinct reads reflect the number of unique reads in each library. Total sequences refer to a total number of reads from each library. All data were obtained after removing short reads (<18 bases) and low-quality reads, and after trimming the adapter sequences.

Sample	Total Sequences	Genome Matched Reads	Distinct Genome Matched Reads	Max. Len. Of Reads	Min. Len. Of Reads
G4-CatR1	20,090,789	14,977,669	2,674,231	34	18
G4-CatR2	18,830,975	14,517,660	1,972,426	34	18
G4-CatR3	22,326,506	16,940,523	3,247,921	34	18
G5-CatR1	18,565,555	13,750,793	2,475,142	34	18
G5-CatR2	22,621,119	16,448,487	3,282,311	34	18
G5-CatR3	22,464,546	15,275,208	3,312,314	34	18
G4-Sir-R1	22,441,403	17,219,531	2,592,693	34	18
G4-Sir-R2	31,970,262	24,306,782	1,957,856	34	18
G4-Sir-R3	36,098,033	28,503,985	2,959,089	34	18
G5-Sir-R1	28,498,217	21,480,588	3,726,432	34	18
G5-Sir-R2	37,037,953	28,515,365	5,405,988	34	18
G5-Sir-R3	24,324,433	18,445,833	3,352,651	34	18

miRNA identification

We further analyzed the sRNAs present in the libraries, and 155 mature miRNAs were identified (Supplementary Table 1). From these sequences, 49 were previously known miRNAs, belonging to 29 families. A total of 106 new miRNAs were identified from the libraries. The abundancies of the identified miRNAs (TP30M) were up to 4.4 mi TP30M (miR166_1 – Supplementary Table 1). Many novel miRNAs presented very low abundances (Supplementary Table 1), but in other tissues these miRNAs can be more abundant and further studies can confirm this. These newly discovered miRNAs can be considered genus-specific miRNAs, since the

mature sequences were identified in a *C. arabica* sRNA transcriptome and the precursors were predicted in the *C. canephora* genome.

Homology based search of all the known miRNAs using miRBase (version 21) revealed that the libraries of *C. arabica* exhibited the highest number of homolog sequences with soybean (*Glycine max*), followed by Arabidopsis and *Populus trichocarpa* (Figure 4).

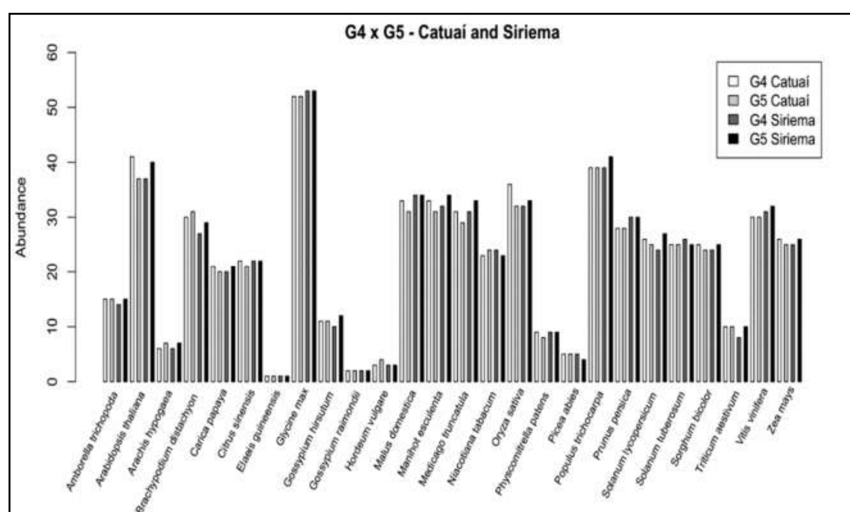


Figure 4. The number of homolog sequences of identified miRNAs with other plant species deposited on miRBase. Values on Y axis indicate the number of miRNA families with homology between *Coffea arabica* and other plant species.

The sizes of the miRNAs identified ranged from 20-24 nt, mostly 21 and 24, with 65 and 56 sequences, respectively (Figure 5). The known miRNAs presented size of 20-22 nt, and the novel miRNAs presented size of 20-24 nt. DICER-like proteins are able to process double-stranded RNA (dsRNA) into small-RNA (sRNA) (Margis et al. 2006). DCL1 is implicated in miRNA processing (21 nt), while DCL2 (22 nt), DCL3 (24nt) and DCL4 (21 nt) are also implicated in short interfering RNA (siRNA) generation (Henderson et al. 2006). DCL3-generated 24 nt miRNA products have been described and are known to be involved in DNA methylation (Hu et al. 2014). The distance between the RNase III and PAZ domains is suggested to be determinant in miRNA length (Rogers and Chen 2013).

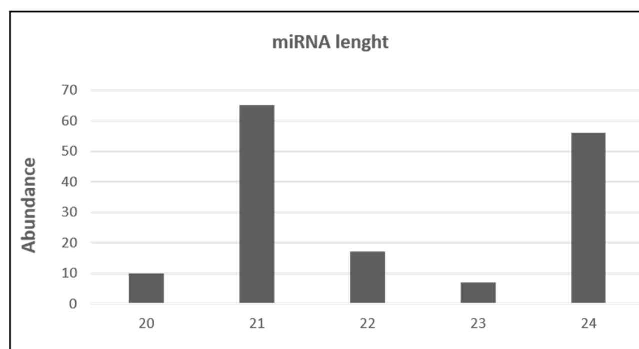


Figure 5. Size distribution of miRNAs identified from all samples of *C. arabica* in G4 and G5 phases.

Analysis of nucleotide bias was performed on known and novel miRNAs. The analysis showed that uracil (U) appeared at the first position with a high frequency in the known miRNAs (Figure 6A), while in the novel miRNAs adenine (A) and U were nearly equally frequent as the first base (Figure 6B). The first 5′ nucleotide theis one of the major determinants for miRNA association to an ARGONAUTE (AGO) protein (Mi et al. 2008). AGO1 harbors microRNAs (miRNAs) that favor a 5' terminal uridine, while AGO4 preferentially recruit small RNAs with a 5' terminal adenosine (Mi et al. 2008).

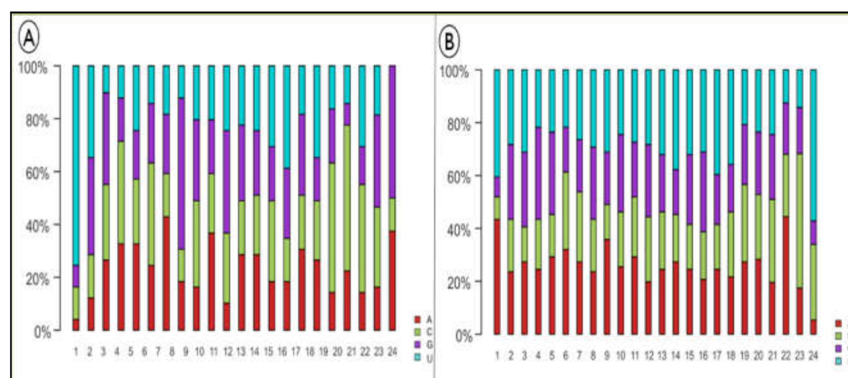


Figure 6. Examination of nucleotide bias within conserved (A) and novel miRNA (B) sequences. Height of bar is proportional to the percentage frequency of the corresponding base at the given position from 1st to 24th.

Furthermore, because the *C. arabica* genome is not public available yet, the precursors were predicted in the *C. canephora* genome, along with their chromosome position and strand, totalizing 211 precursors (Table 2). MIR loci were found in all chromosomes, both in plus and minus strands, and MIR clusters within 10kb were also observed. The MIR genes are evolutionarily conserved, and they can be located in intergenic regions of target genes or on the antisense strand of neighboring genes (Zhou et al. 2007). Arrangements in tandem are rare in plant miRNAs, and they may be clustered in some species (Budak and Akpinar 2015; Voinnet 2009; Zhou et al. 2011).

The largest family was MIR395 with 12 loci, followed by MIR156 with nine and MIR166 and MIR171 both with seven members. From the novel miRNAs, MIR048 and MIR045 were the largest families, both with four members, followed by four novel MIR with three members each (ccp-MIR005, ccp-MIR022, ccp-MIR40, ccp-MIR109) (Table 2).

Table 2. Identification of pre-miRNAs in *Coffea canephora*. Precursor names, chromosome numbers, start and end positions, strand, status (novel or known), mature sequence and length.

Final Names	Ch	Start	End	Strand	Status	Mature Sequence	Len
ccp-m001	2	43767035	43767217	+	New	AAACCGGTTCGTGTAA TTCTTACT	24
ccp-m002	12	159324653	159324748	+	New	AAAGAGTCTTCTGAAA ATTTGCCT	24
ccp-m004	12	188804663	188804903	+	New	AACACTCTTTCTAGCGT TTGGAAG	24
ccp-m005	8	26164907	26165119	+	New	AACCAATGACTATTCA TGATTC	22
ccp-m005	12	204407875	204408087	+			
ccp-m005	12	130993055	130993267	-			
ccp-m006	4	24763622	24763756	-	New	AACGGACAAGTCATGG TTTGGACT	24
ccp-m007	7	11391711	11391962	+	New	AACTGATGAAAGGACC AAAATGCC	24
ccp-m008	6	30942699	30942818	+	New	AACTGTAGACACCAAA TTTTGAAT	24
ccp-m009	9	15946419	15946531	+	New	AAGACACTATTTACAT TTGAAACT	24
ccp-m010	3	10096476	10096677	-	New	AAGAGACGGAGGCACA GCAGACTC	24
ccp-m012	12	76705375	76705532	-	New	AAGATCCGGCAGAACC ATCTGAGC	24
ccp-m012	12	76723763	76723920	-			
ccp-m013	9	14811515	14811732	+	New	AAGGACCAGCCAGTGA AAGAAATC	24

ccp-m014	2	44292219	44292393	-	New	AAGGGACCTGAGAAGT GTTGAGAG	24
ccp-m015	7	11157580	11157675	-	New	AATATACTGAGAAATG AGCCT	21
ccp-m016	6	5446706	5446904	+	New	AATGAAAGACAAGATT TGAATTCT	24
ccp-m017	4	22254996	22255235	+	New	AATGATCAATTGGCTG TTGCAAAT	24
ccp-m019	2	16776762	16776982	+	New	AATTACCGTAGCACTTT TTATGAT	24
ccp-m020	12	131514841	131514964	+	New	ACATAATATTCGGATT GTGAGTCT	24
ccp-m021	12	59348128	59348304	+	New	ACCAATTCTTGGAGCT GTCGCGCC	24
ccp-m022	3	16238997	16239250	+	New	ACTCGACTCATGTAAG AACTCACA	24
ccp-m022	12	196961858	196962111	+			
ccp-m022	12	134078140	134078393	-			
ccp-m023	10	6454003	6454175	-	New	ACTGGCAAATAGACGG AAAGCACC	24
ccp-m024	11	5202800	5202983	-	New	AGAAGCCGAAGGATTG ACGAGC	22
ccp-m025	10	24256757	24256895	+	New	AGACTGGTTAATTTCGT GATTACT	24
ccp-m027	8	28146078	28146169	+	New	AGAGAATAGTTGTGTT TCGTCATG	24
ccp-m028	10	24247190	24247398	+	New	AGAGATTCTTCCTGAC CATTGGAC	24
ccp-m029	12	92016241	92016349	+	New	AGAGCAATAGATGTCC ATCACTGT	24
ccp-m030	12	34376915	34377004	-	New	AGCCAAGGGAATAGGA TCGTCATC	24
ccp-m031	11	26897425	26897595	-	New	AGCTTAATCTTCAGGA AATTAGT	23
ccp-m032	3	15675380	15675593	-	New	AGGACCATCACTTTGT ACAATTGG	24
ccp-m033	2	22967750	22967862	+	New	AGGAGACCTGTGAATG CTATTACT	24
ccp-m034	9	10867843	10867975	+	New	AGGAGAGTTGAACTGG GTCCGCCT	24
ccp-m034	11	33161302	33161432	-			
ccp-m035	12	164311679	164311782	+	New	AGGAGATGAACTCCCG AATAGCCT	24
ccp-m036	10	27331312	27331397	+	New	AGGCACTTTTCTGGTTT TGAAGGC	24
ccp-m037	12	145774497	145774578	-	New	AGGGTATCGGTGGGAC GCCTGCCT	24
ccp-m038	5	19313366	19313448	+	New	AGGGTCAGAGGCTTTT GTCAACCT	24
ccp-m039	10	9055071	9055156	+	New	AGGTACTTTTCTTGTTT TGAAGGC	24
ccp-m040	11	17552661	17552734	+	New	AGTGGCTGAGATCTTT GTTTTGAT	24
ccp-m040	12	76154022	76154095	-			
ccp-m040	12	152308344	152308417	-			
ccp-m041	3	15932639	15932787	+	New	AGTTGCTAATGTCTTGG ATTTTAG	24
ccp-m043	4	5233120	5233247	-	New	ATACGGCGTAACTTTG TTTGC	21
ccp-m045	2	19706412	19706517	+	New	ATGAATGTAGTTTTCA ACGCCACC	24
ccp-m045	7	20687190	20687313	+			
ccp-m045	12	32161016	32161271	+			

ccp-m045	12	120004115	120004236	-			
ccp-m046	12	105807318	105807539	-	New	ATGAGACAAAAGATTC GTTGAGCT	24
ccp-m048	3	9289702	9289934	+	New	ATGTATCTAATAGGAT GTGGGCCT	24
ccp-m048	12	30833949	30834183	+			
ccp-m048	12	86547585	86547734	+			
ccp-m048	12	98588741	98588886	-			
ccp-m049	10	23547599	23547699	-	New	ATTAGAAACAATGAAT CAAAAGAT	24
ccp-m050	12	3730055	3730304	+	New	ATTCTCGATGTAACTG AATGCACC	24
ccp-m051	6	506785	506975	+	New	ATTCTTAGATGATCCGT GGACATC	24
ccp-m052	9	1225681	1225938	+	New	ATTGAGATTGGATTCG ATTAGACT	24
ccp-m053	1	25770018	25770283	-	New	ATTGGTACACAGCAAA GGACT	21
ccp-m054	12	108641072	108641306	+	New	CAAACCCAACTCTTTAT GCCA	21
ccp-m055	6	513563	513776	-	New	CATGTGCCTGTGTTCTC CATC	21
ccp-m056	3	6452370	6452515	-	New	CCAGGAAAGTGTATAA CCACC	21
ccp-m057	7	4250593	4250699	-	New	CCCATTCTTGTGAATCT CGCC	21
ccp-m058	1	15407952	15408128	-	New	CCGTCGGTGGCATTTG AAACTAGA	24
ccp-m059	4	5529749	5529978	-	New	CCTGTCGCTCACTTTGC CCGCCC	23
ccp-m060	12	145655806	145655917	+	New	CTAGGAGTAGAGTAAT TGATC	21
ccp-m061	8	8543700	8543786	+	New	CTGACCCGAACCTGAT CCACCAAC	24
ccp-m062	3	7988594	7988856	-	New	CTTGGAAACTGTTTAA CTGCC	21
ccp-m064	9	17441063	17441326	+	New	GATTTCAACGTCGGAG CAAAGAGT	24
ccp-m065	5	24196315	24196421	-	New	GCATCAGAGGAGTCAG GCAGG	21
ccp-m065	5	24201495	24201600	-			
ccp-m066	1	23438361	23438492	-	New	GCGCATCTGGTGTAGT GGTATC	22
ccp-m067	11	5265876	5266014	-	New	GGAATCGAAGAATTGA AGAGC	21
ccp-m068	12	167871761	167871833	+	New	GGCACTCGATTTCGAC GTCGGAGC	24
ccp-m069	11	29416867	29417020	+	New	GGGATGGAACCTGAGA ACACCGG	23
ccp-m071	11	26954314	26954538	-	New	GGGGTGGGAGACTGGG GAAGA	21
ccp-m072	5	4689227	4689397	-	New	GTAGTGGATCAAGAAG TATAGACT	24
ccp-m073	1	1240046	1240133	+	New	TAAAACAAACAATGTG TGCGGACT	24
ccp-m074	3	22224806	22224902	-	New	TAAAATTTTGAAGTTG AGCTT	21
ccp-m075	7	18671852	18671932	-	New	TAAGATTGATCCGTCA TATGAACT	24
ccp-m076	6	1721204	1721291	+	New	TAAGGAGGACCCACAT TACTTACT	24
ccp-m077	9	3976895	3976985	+	New	TAAGGAGGACCCACGT TACTTACT	24
ccp-m078	11	22292163	22292362	+	New	TATGATAGAATTAAGT GCACC	21

ccp-m079	3	28291451	28291670	+	New	TCAGTCTTTTTTTCTCTCCT	20
ccp-m080	3	13558233	13558486	-	New	TCATTGTAAACTTTTTTGGCT	21
ccp-m081	8	13837773	13837864	+	New	TCCATTTCCAGCTTGTTTGAC	21
ccp-m083	11	5202808	5202971	-	New	TCGCAGAAGACAGCCGCATC	20
ccp-m084	1	30156217	30156321	-	New	TCGCAGGAGAGATGAAACCGAA	22
ccp-m085	1	9581075	9581190	+	New	TCGGCGTCCTCCTTGCACCAC	21
ccp-m086	1	1500800	1500882	+	New	TCGTGGACCAGATCATGCATC	21
ccp-m087	6	17455	17660	-	New	TCTCCAGTGGATTCTCTCTCC	21
ccp-m088	12	175663025	175663149	+	New	TCTCTCCCTCCCCTGCCACCC	21
ccp-m090	11	5248280	5248456	-	New	TCTGTCGCAGGTGACTTTCGCC	22
ccp-m091	11	5265975	5266100	-	New	TCTTCCCCTCTTCGGCTTACT	21
ccp-m092	10	11790262	11790459	-	New	TGACTTGGTCAACTTTTGCATGCC	24
ccp-m093	3	8602543	8602681	-	New	TGAGAGACATTGAAGGACTTA	21
ccp-m094	12	119733383	119733464	+	New	TGAGGTATGGACCAGCAAGGACT	23
ccp-m095	12	119745549	119745630	+	New	TGAGGTTATGACCTGCAAGGACT	23
ccp-m096	6	5375123	5375305	+	New	TGATTCATTCAGCAAGCTGTG	21
ccp-m097	11	26964592	26964747	-	New	TGATTTCCTGTGGATTCTCC	20
ccp-m098	11	1608789	1608912	-	New	TGCACCCGTTTCCCTCTCTCC	21
ccp-m099	11	5236527	5236648	-	New	TGTCGCAGGAGACTGGCGCCT	21
ccp-m100	7	1961572	1961704	+	New	TGTCGCAGGAGGAATGGCACC	21
ccp-m101	12	119243680	119243760	-	New	TTATAAATGACTGCACTTGGAC	22
ccp-m102	2	9859009	9859222	-	New	TTCAAATCCTCGGTCCCGAAGCCT	24
ccp-m104	10	18193552	18193683	+	New	TTCTGCATTTCCTGGTTTAGGAAA	24
ccp-m105	7	10003754	10003869	+	New	TTGCATACGCGCCTGAATCGG	21
ccp-m106	10	24881594	24881729	+	New	TTGCCAGACCTCCCATCCCATA	22
ccp-m109	5	10389313	10389547	-	New	TTGTAATTGGATGATATTGAGCCT	24
ccp-m109	8	6529593	6529775	+			
ccp-m109	12	73426965	73427147	+			
ccp-m109	12	96067402	96067638	+			
ccp-m109	12	122413865	122413998	-			
ccp-m110	11	16252040	16252171	+	New	TTGTAGATCTTGATAGAGCCT	21
ccp-m111	7	1958205	1958310	+	New	TTGTCGCAGGAGATATGGCACT	22
ccp-m113	9	10820289	10820461	+	New	TTTATACCCGTTTTGACTGCAAGT	24
ccp-m116	11	26367679	26367877	-	New	TTTCCCAGACCTCCCGTGCCGG	22
ccp-m118	8	6416608	6416735	+	New	TTTGATTTCAGGCTGTTCCATTC	23

ccp-m119	8	6419264	6419391	+	New	TTTGATTTTAGGTTGTT CCATTC	23
ccp-m120	12	39680015	39680131	-	New	TTTGGATGAGTAAATG TTGAGATT	24
ccp-m121	1	23110899	23111038	-	New	TTTGGCTGTCTAACTAT TAACTGT	24
ccp-m122	2	13816398	13816652	+	New	TTTGGCTTTTCTCTGTT AGGC	21
ccp-m123	7	23212883	23213029	-	New	TTTTGATGAGACCATAT TATC	21
ccp-m123	12	119005610	119005756	+			
ccp-m124	3	3898384	3898532	+	New	TTTTGCACAAGCCATCC AAC	20
ccp-miR156_1	6	9429865	9430000	+	Known	CGACAGAAGAGAGTGA GCAC	20
ccp-miR156_2	12	173196178	173196323	+	Known	TGACAGAAGAGAGTGA GCAC	20
ccp-miR156_3	4	5161779	5161914	-	Known	TTGACAGAAGATAGAG AGCAC	21
ccp-miR156_3	7	122545	122648	-			
ccp-miR156_3	7	6213181	6213290	-			
ccp-miR156_3	7	122529	122663	-			
ccp-miR156_3	7	6213167	6213305	-			
ccp-miR156_4	5	28048983	28049098	-	Known	TTGACAGAAGAGAGAG AGCAC	21
ccp-miR156_4	5	28049058	28049236	-			
ccp-miR160_1	9	1683410	1683545	+	Known	TGCCTGGCTCCCTGTAT GCCA	21
ccp-miR162_1	4	6151821	6151982	+	Known	TCGATAAACCTCTGCA TCCAG	21
ccp-miR164_1	2	47655457	47655557	-	Known	CACGTGCTCCCCTTCTC CAAC	21
ccp-miR164_1	2	47655442	47655569	-			
ccp-miR166_1	1	35697983	35698134	+	Known	TCGGACCAGGCTTCAT TCCCC	21
ccp-miR166_1	2	19855672	19855856	+			
ccp-miR166_1	2	54264929	54265121	+			
ccp-miR166_1	2	19855672	19855856	+			
ccp-miR166_1	3	3908863	3909003	-			
ccp-miR166_1	8	31212091	31212302	+			

ccp-miR166_2	6	5768088	5768294	-	Known	TCTCGGACCAGGCTTCATTCC	21
ccp-miR167_1	1	35954214	35954340	-	Known	TGAAGCTGCCAGCATGATCTA	21
ccp-miR167_1	3	3324003	3324123	+			
ccp-miR167_2	2	36665745	36665864	-	Known	TGAAGCTGCCAGCATGATCTGA	22
ccp-miR168_1	8	6480112	6480265	+	Known	TCGCTTGGTGCAGGTCGGGAA	21
ccp-miR169_1	7	2455343	2455442	+	Known	CAGCCAAGGATGACTTGCCGG	21
ccp-miR169_2	6	32478049	32478184	+	Known	CAGCCAGGGATGACTTGCCGA	21
ccp-miR171_1	7	8354897	8355053	+	Known	TATTGGCCTGGTTCACTCAGA	21
ccp-miR171_2	3	18565898	18565996	+	Known	TGATTGAGCCGTGCCAATATC	21
ccp-miR171_2	9	1285034	1285132	+			
ccp-miR171_2	10	844801	844951	+			
ccp-miR171_3	6	6553647	6553762	-	Known	TTGAGCCGCGCCAATATCACT	21
ccp-miR171_4	5	18036725	18036834	+	Known	TTGAGCCGCGTCAATATCTCT	21
ccp-miR171_5	5	18097383	18097513	+	Known	TTGAGCCGTGCCAATATCACG	21
ccp-miR172_1	2	8974707	8974818	+	Known	GGAATCTTGATGATGCTGCAT	21
ccp-miR172_2	2	6092592	6092779	-	Known	GGAATCTTGATGATGCTGCAG	21
ccp-miR2111_1	6	1734967	1735088	+	Known	GTCCTCAGGATACAGATTACC	21
ccp-miR319_1	4	5038385	5038613	-	Known	TTGGACTGAAGGGTTTCCTTC	21
ccp-miR319_2	8	30318324	30318512	-	Known	TTGGACTGAAGGGAGCTCCCT	21
ccp-miR3627_1	4	25888272	25888378	-	Known	TCGCAGAAGAGATGGCACCTA	21
ccp-miR390_1	1	32829593	32829735	+	Known	AAGCTCAGGAGGGATAGCGCC	21
ccp-miR390_1	1	32831307	32831449	+			
ccp-miR390_1	1	32829593	32829735	+			

ccp-miR390_1	1	32831307	32831449	+			
ccp-miR390_1	6	718646	718803	-			
ccp-miR393_1	2	5587557	5587711	-	Known	TCCAAAGGGATCGCATTGATCC	22
ccp-miR393_1	2	5587541	5587726	-			
ccp-miR394_1	10	765120	765282	+	Known	TTGGCATTCTGTCCACCTCC	20
ccp-miR395_1	9	20573212	20573328	+	Known	CTGAAGTGTTTGGGGGAACTC	21
ccp-miR395_1	9	20596829	20596973	+			
ccp-miR395_1	9	20610657	20610776	+			
ccp-miR395_1	9	16746623	16746767	-			
ccp-miR395_1	9	16746623	16746767	-			
ccp-miR395_1	11	30224783	30224893	-			
ccp-miR395_1	11	30226755	30226874	-			
ccp-miR395_1	12	135561900	135562016	+			
ccp-miR395_1	12	141412197	141412399	-			
ccp-miR395_1	12	187220911	187221055	-			
ccp-miR395_1	12	202360574	202360722	-			
ccp-miR395_2	6	21798044	21798153	+	Known	TGAAGTGTTTGGGGGAACTC	20
ccp-miR396_3	2	756983	757168	+	Known	TCCACAGGCTTTCTTGAACG	20
ccp-miR396_3	6	9752060	9752196	+			
ccp-miR396_4	6	6505953	6506112	+	Known	TTCCACAGCTTTCTTGAACTG	21
ccp-miR396_5	4	3379788	3379886	+	Known	TTCCACAGCTTTCTTGAACTT	21
ccp-miR396_5	6	6513941	6514111	-			
ccp-miR397_1	5	26533024	26533156	-	Known	TCATTGAGTGCAGCGTTGATG	21

ccp-miR398_1	2	34102356	34102467	+	Known	TGTGTTCTCAGGTCGCCCCTG	21
ccp-miR399_1	2	13915837	13915939	+	Known	TGCCAAAGGAGAATTGCCCTG	21
ccp-miR399_1	2	13865570	13865703	+			
ccp-miR399_2	2	13918938	13919090	-	Known	TGCCAAAGGAGATTTGCCCCG	21
ccp-miR399_3	2	13862020	13862173	-	Known	TGCCAAAGGAGATTTGCCCGG	21
ccp-miR399_4	8	28925034	28925175	-	Known	TGCCAAAGGAGAGTTGCCCT	20
ccp-miR403_1	1	24017127	24017268	-	Known	TTAGATTCACGCACAAACTCG	21
ccp-miR408_1	2	49819183	49819308	-	Known	TGCACTGCCTCTTCCCTGGCTG	22
ccp-miR477_1	8	30572650	30572741	+	Known	CACTCTCCCTCAAGGGCTTCCG	22
ccp-miR479_1	2	943834	943937	+	Known	TGAGCCGAACCAATATCACTC	21
ccp-miR479_1	2	943822	943952	+			
ccp-miR482_1	11	26816056	26816170	-	Known	GGAATGGGCTGCTAGGGATGG	21
ccp-mir482_2	11	26937326	26937564	-	Known	TTTCCCAGTCCTCCCATTCCTA	22
ccp-mir482_3	11	26937255	26937370	-	Known	TTCCCTAGTCCTCCCATCCCGG	22
ccp-mir482_3	11	26946797	26946922	-			
ccp-mir482_3	11	26964595	26964739	-			
ccp-miR482_4	11	26969627	26969803	-	Known	TTTCCCAATCCACCCATACCGA	22
ccp-miR530_1	6	3344220	3344478	+	Known	AGGTGCAGATGCTGATGCAGA	21
ccp-miR7122_1	3	3950526	3950691	-	Known	TTATACAGAGAAACCGCGGTTG	22

Differential expression analysis

The abundances of normalized reads (TP30M) of G5 and G4 stages were used to calculate the fold change $\log_2(G5/G4)$ and the p-value and FDR were calculated to confirm differential expression. There were no differentially expressed miRNAs between G5/G4 within Siriema or within Catuaí, but between G5/G4 in Siriema+Catuaí, 17 miRNAs showed FDR<=0.05 and $\log_2(G5/G4)>1$ and were, therefore, differentially expressed

(Supplementary Table 1). When the p-value was calculated within each cultivar, the three biological replicates of each stage in the cultivar were considered, while when we compared using the two cultivars, the three replicates of each cultivar were used as replicates, increasing the N (N=6) for calculation of p-value. With the lower number of replicates, the p-value was not significant and differential expression could not be observed (data not shown). Therefore, the replicates from both cultivars were combined for the next analyses and only the development stages were investigated.

From the 17 differentially expressed miRNAs, 16 were up and one was downregulated (Figure 7), from which 12 were novel miRNAs (ccp-m065, ccp-m096, ccp-m055, ccp-m087, ccp-m064, ccp-m045, ccp-m105, ccp-m005, ccp-m015, ccp-m083, ccp-m069, ccp-miR091) and five were conserved (ccp-miR156_2, ccp-miR172_2, ccp-miR167_1, ccp-miR171_1, ccp-miR396_4). The only downregulated miRNA was ccp-miR091.

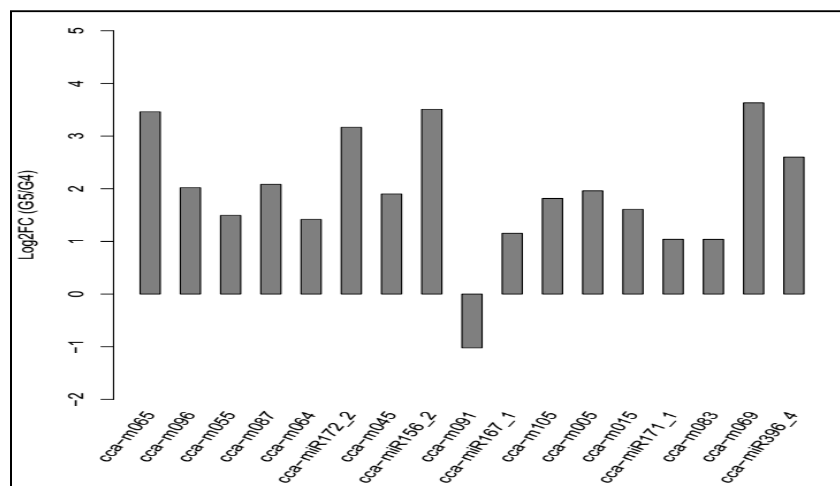


Figure 7. Fold change analysis of the differentially expressed miRNAs between G5 and G4 in Coffea arabica floral buds.

The differentially expressed miRNAs are potentially playing crucial roles in resume of growth from G4 to G5 stage. Interestingly, more novel than known miRNAs are found differentially expressed, and further investigation needs to be performed to addressed on whether they participate in this phase

change. All the known miRNAs have already been described in flowering regulation. For instance, miR156 regulate genes involved in the control of the transition from the vegetative phase to the reproductive phase, such *SQUAMOSA PROMOTER BINDING PROTEIN LIKE (SPL)* (Zhang et al. 2015; Zhou and Wang 2013). miR156 gradually decreases as plants age, and an increase in SPL promotes flowering through activating *FT*, MADS-box and *LFY* genes (Wang et al. 2009). Overexpression of miR156 decreases SPL expression, which results in delayed flowering (Chuck et al. 2007). On the other hand, overexpression of miR172 in Arabidopsis accelerates flowering (Aukerman and Sakai 2003). miR172 increases as plants approach flowering, and also accumulates in leaves and floral buds (Aukerman and Sakai 2003). miR172 acts in the regulatory process of flowering time and also in the determination of floral organ identity, modulating the expression of the *APETALA2-like (AP2)* genes, which are repressors of flowering and A-function floral identity genes (Aukerman and Sakai 2003; Wollmann et al. 2010). miR172 is also activated in the photoperiod and temperature pathways (Jung et al. 2007; Lee et al. 2010; Yamaguchi and Abe 2012). Both miR156 and miR172 have already been identified in late stages of flower development (Li et al. 2015). The upregulation of both miRNAs between the two stages of development in *C. arabica* indicates a role in flower resume of development towards anthesis, but is controversial with their antagonistic expression pattern and requires further investigation.

miR167 targets *AUXIN RESPONSIVE FACTOR* family members (*ARF6/ARF8*), and is implicated in sepal, petal and stamen development (Rubio-Somoza and Weigel 2013). ARF6 and ARF8 regulate both stamen and gynoecium maturation (Nagpal et al. 2005), and miR167 is essential for correct patterning of gene expression (Wu et al. 2006). miR167 cleavage of ARF8 in *C. arabica* and *C. canephora* has already been confirmed by 5'RACE in our previous report (Chaves et al. 2015). miR167_1 is low expressed in the tested tissues, with mean values of 63,5 TP30M (G4) and 141,17 TP30M (G5), but differential expression between the two phases indicates that flower

maturation may occur in G4 and ARF6/ARF8 are probably active in this phase, while in G5 miR167_1 is upregulated to control the expression on ARF/6/ARF8.

MIR171 is a well conserved miRNA family known to regulate members of the GRAS/SCARECROW-LIKE (SCL) transcription factor family. In Arabidopsis, miR171a is most highly expressed in the inflorescence where it regulates SCL6-III and SCL6-IV expression through mRNA cleavage (Llave et al. 2002; Reinhart et al. 2002). In barley, miR171 accumulates in reproductive organs, and overexpression of miR171 causes delayed flowering, along with altered shoot architecture (Curaba et al. 2013). Interestingly, in *Vitis vinifera L.* (grapevine) and *Prunus mume* (peach), miR171 decreases during inflorescence development (Belli Kullan et al. 2015; Wang et al. 2014), diverging from the upregulation observed in *C. arabica*. In *Cymbidium ensifolium*, an important ornamental flower in China, miR171 was identified in two phases of floral development (buds < =0.5 cm and with 2-3cm), but not differentially expressed (Li et al. 2015).

In Arabidopsis, seven out of the nine Growth-Regulating Factors (GRFs) are proved to be regulated by miR396, which plays critical roles in cotyledon and leaf growth and development (Casadevall et al. 2013; Debernardi et al. 2014). Recently, an important role has been described for miR396 in controlling carpel number and pistil development via regulation of the GRF/GRF-INTERACTING FACTORS complex in Arabidopsis (Liang et al. 2014). They demonstrated that overexpression of MIR396a resulted in pistils with a single carpel, while the miR396-resistant version of mGRF7 or mGRF9 could rescue miR396-overexpressing plants (Liang et al. 2014). In rice, both knockdown of OsGRF6 and overexpression of OsmiR396d showed similar defects in floral organ development, and overexpression of OsGRF6 was able to almost completely rescued these defects (Liu et al. 2014). Additionally, overexpression of ptc-MIR396, with identic mature sequence to ath-MIR396, resulted in downregulation of NtGRFs in tobacco, and abnormal floral organs were also observed (Baucher et al. 2013). In tomato, short

tandem target mimic (STTM), a method to destruct small RNAs by small RNA degrading nucleases (Yan et al. 2012), has been used to study the MIR396 function in flower and fruit development (Cao et al. 2016). As a result, the flowers, sepals, and fruits of transgenic plants were all significantly larger than those of non-tansgenic plants, although the cotyledons were shorter (Cao et al. 2016).

Interestingly, miR396 downregulates the MADS-box flowering repressor SHORT VEGETATIVE PHASE (SVP), a flowering repressor, via mRNA decay rather than cleavage based on degradome analysis (Yang et al. 2015). It has been already identified expression of *FLC* and *SVP1* in coffee buds in later stages of development, as an indication of their involvement in bud dormancy (Oliveira et al. 2014). Further investigation of whether ccp-MIR396 down-regulates SVP in coffee plants can contribute to the elucidation of the dormancy release that occurs between G4 and G5 phases in *C. arabica.*

Although not differentially expressed, miR166_1 and miR166_2 were highly expressed in both stages, with mean values over 3mi TP30M (Supplementary Table 1). The Arabidopsis MIR166/165 group targets five members of the HD-ZIP III transcription factor genes, such as REVOLUTA (REV), PHABULOSA (PHB), PHAVOLUTA (PHV), CORONA (CNA)/ATHB15, and ATHB8 (Byrne 2006; Juarez et al. 2004; Kidner and Martienssen 2004; Kim et al. 2005; Mallory et al. 2004; Williams et al. 2005). MIR166 are involved in stamen and carpels development (Jung and Park 2007), and were also found significantly high expressed in these tissues in *Vitis vinifera* L. (Belli Kullan et al. 2015).

Secondary structure and Target prediction

The secondary structures of the novel miRNAs that presented differential expression were predicted using RNA-fold software (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). As mentioned above, the structure prediction was performed in the *C. canephora* genome, the closest

species with sequenced genome available. Precursors formed characteristic stem-loop structures (Figure 8).

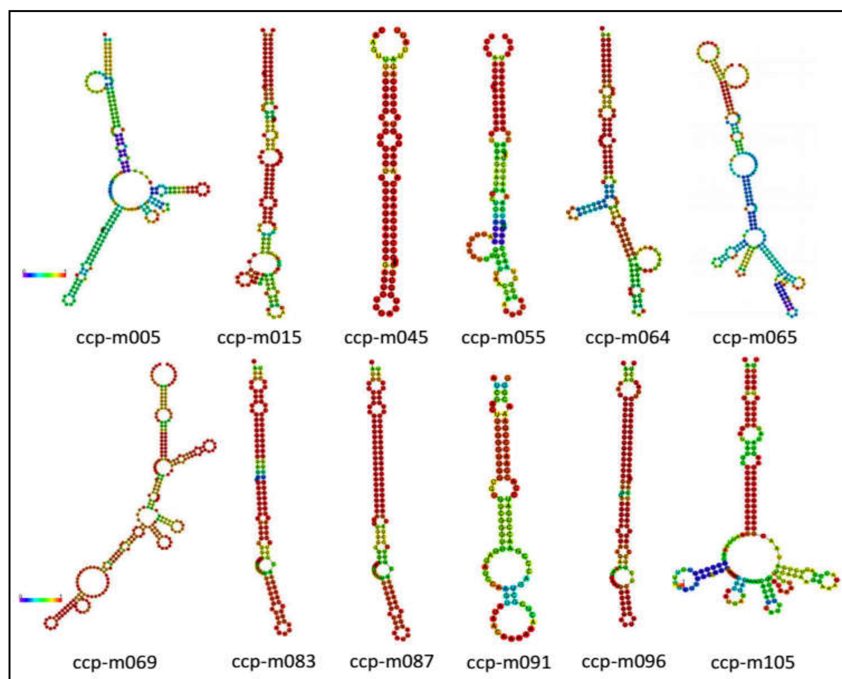


Figure 8. Predicted fold-back structures for the novel miRNAs in *C. canephora.*

The putative target genes of the differentially expressed miRNAs were predicted in the *C. canephora* genome (Supplementary Table 2). Most of the predicted targets are hypothetical proteins or non-characterized proteins, but some annotated genes were also identified. miR156 showed eight putative targets among them three SPB genes, and miR172 presented five putative targets among them two AP2-like genes (RAP2-7). miR167 had 10 predicted targets, from which five are Gibberellin 20 oxidase 1-D, that catalyzes consecutive steps of oxidation in the late part of the GA biosynthetic pathway (Qin et al. 2013). For miR396, 22 putative targets were predicted in the *C. canephora* genome, among them a MADS-box gene (AGL62). Several targets were also predicted for the novel miRNAs.

To investigate the general function of the putative targets, Gene Ontology (GO) was used to classify and group the classes of the miRNA targets. The web tool SEA (Singular Enrichment Analysis) from agriGO (http://bioinfo.cau.edu.cn/agriGO/index.php) was used (Du et al. 2010). A total of 84 GO terms were annotated for the target genes in *C. canephora*. The genes belonging to the 13 overrepresented terms among the three GO categories, namely the biological process, molecular function, and cellular component categories, are presented (Figure 9). The main overrepresented subcategories belonging to the 'Biological Process' categories were 'cellular process', 'regulation of biological process', 'biological regulation', 'metabolic process', 'establishment of localization', and 'localization'. In the 'Cellular Component' category, the main overrepresented terms were 'macromolecular complex', 'cell part', 'organelle', and 'cell'. In the 'Molecular Function' category, the main overrepresented terms were 'transcription regulator activity', 'catalytic activity', and 'binding'. These overrepresented terms imply that the predicted targets are involved in regulatory processes, and one can infer that the miRNAs differentially expressed between G4 and G5 stages in *C. arabica* are key regulators of the resume of growth of the buds.

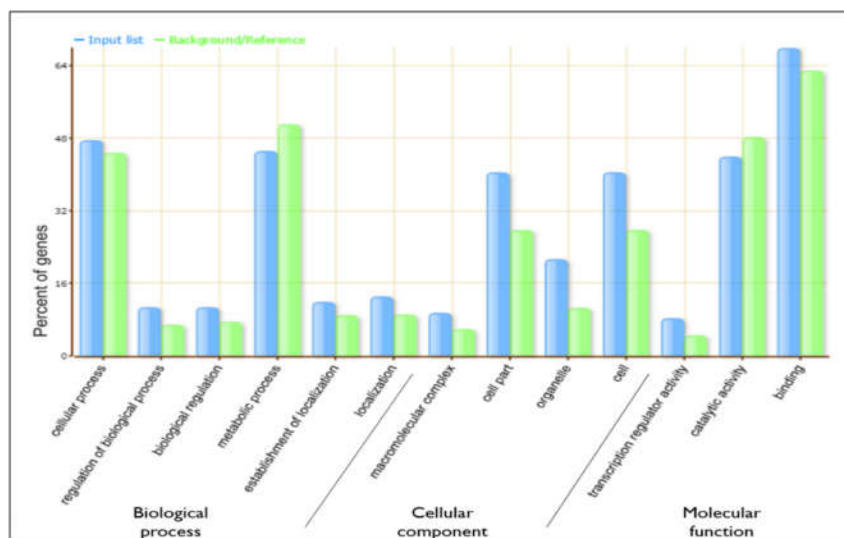


Figure 9. Gene ontology enrichment of *C. canephora* targeted by differently expressed miRNAs. The percentage of significantly enriched GO terms were categorized per biological process, cellular component, and molecular function. Blue bars represent the input list (putative targets list) and Green Bars represent the Backgroung/Reference List (*C. canephora* genome).

Conclusion

Understanding the mechanism of processes associated with flower development in different flowering plants is an important aspect of miRNA regulated processes. This study established a miRNA database for two contrasting cultivars – 'Red Catuaí IAC 144' and 'Siriema VC4' – of *C. arabica* in two stages of flower bud development of coffee, G4 and G5, which includes the identification of 155 mature miRNAs, 49 previously known and 106 novel miRNAs. A total of 211 MIR loci were identified in the *C. canephora* genome. No differential expression between the cultivars was observed, but we reported differential expression of 17 miRNAs between G5 and G4 stages, which might play a crucial role in flower development and resume of growth processes. The putative novel miRNAs might provide further clue in gene regulation of flower development processes in *Coffea* species. Furthermore, the secondary structure of the precursors of the novel

miRNAs and the putative targets of the differentially expressed genes were predicted in the *C. canephora* genome, and Gene Ontology enrichment analysis revealed that the predicted targets are involved mainly in regulatory processes. This diverse set of miRNAs will provide useful resource for the investigation of *C. arabica* flower maturation, aiming at the comprehension of the bud dormancy and resume of growth mechanisms.

References

Aceto S, Sica M, De Paolo S, D'Argenio V, Cantiello P, Salvatore F, Gaudio L (2014) The Analysis of the Inflorescence miRNome of the Orchid *Orchis italica* Reveals a *DEF*-Like MADS-Box Gene as a New miRNA Target PLoS ONE 9:e97839 doi:10.1371/journal.pone.0097839.

Achard P, Herr A, Baulcombe DC, Harberd NP (2004) Modulation of floral development by a gibberellin-regulated microRNA Development (Cambridge, England) 131:3357-3365 doi:10.1242/dev.01206.

Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes Plant Cell 15:2730-2741 doi:10.1105/tpc.016238.

Baucher M, Moussawi J, Vandeputte OM, Monteyne D, Mol A, Pérez-Morga D, El Jaziri M (2013) A role for the miR396/GRF network in specification of organ type during flower development, as supported by ectopic expression of Populus trichocarpa miR396c in transgenic tobacco Plant Biology 15:892-898 doi:10.1111/j.1438-8677.2012.00696.x.

Belli Kullan J et al. (2015) miRVine: a microRNA expression atlas of grapevine based on small RNA sequencing BMC Genomics 16:1-23 doi:10.1186/s12864-015-1610-5.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing Journal of the Royal Statistical Society Series B (Methodological) 57:289-300.

Budak H, Akpinar BA (2015) Plant miRNAs: biogenesis, organization and origins Functional & Integrative Genomics 15:523-531 doi:10.1007/s10142-015-0451-2.

Byrne ME (2006) Shoot Meristem Function and Leaf Polarity: The Role of Class III HD–ZIP Genes PLoS Genet 2:e89 doi:10.1371/journal.pgen.0020089.

Camargo A (1985) Florescimento e frutificação de Café arabica nas diferentes regiões cafeeiras do Brasil Pesquisa Agropecuária Brasileira 20:831-839.

Camargo A, Franco C (1985) Clima e fenologia do cafeeiro. Ministério da Agricultura e Comércio, Rio de Janeiro.

Camargo M (2010) The impact of climatic variability and climate change on arabic coffee crop in Brazil Bragantia 69:239-247 doi:10.1590/s0006-87052010000100030.

Cao D et al. (2016) Regulations on growth and development in tomato cotyledon, flower and fruit via destruction of miR396 with short tandem target mimic Plant Science 247:1-12 doi:http://dx.doi.org/10.1016/j.plantsci.2016.02.012.

Casadevall R, Rodriguez RE, Debernardi JM, Palatnik JF, Casati P (2013) Repression of growth regulating factors by the microRNA396 inhibits cell proliferation by UV-B radiation in Arabidopsis leaves Plant Cell 25:3570-3583 doi:10.1105/tpc.113.117473.

Chaves SS et al. (2015) New Insights on Coffea miRNAs: Features and Evolutionary Conservation Appl Biochem Biotechnol 177:879-908 doi:10.1007/s12010-015-1785-x.

Chen JL et al. (2016) Identification of miRNAs and their targets through high-throughput sequencing and degradome analysis in male and female Asparagus officinalis BMC Plant Biol 16:19 doi:10.1186/s12870-016-0770-z.

Chuck G, Meeley R, Irish E, Sakai H, Hake S (2007) The maize tasselseed4 microRNA controls sex determination and meristem cell fate by targeting Tasselseed6/indeterminate spikelet1 Nature genetics 39:1517-1521 doi:10.1038/ng.2007.20.

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development Nature 353:31-37.

CONAB (2016) Acompanhamento da Safra Brasileira, Café vol 3.

Crisosto CH, Grantz DA, Meinzer FC (1992) EFFECTS OF WATER DEFICIT ON FLOWER OPENING IN COFFEE (COFFEA-ARABICA L) Tree Physiology 10:127-139.

Curaba J, Talbot M, Li Z, Helliwell C (2013) Over-expression of microRNA171 affects phase transitions and floral meristem determinancy in barley BMC Plant Biol 13:1-10 doi:10.1186/1471-2229-13-6.

de Carvalho CHS et al. (2008) Cultivares de Café Arábica de Porte baixo. In: de Carvalho CHS (ed) Cultivares de Café: Origem, características e recomendações. Embrapa, Brasília, DF.

de Oliveira RR, Chalfun-Junior A, Paiva LV, Andrade AC (2010) In Silico and Quantitative Analyses of MADS-Box Genes in Coffea arabica Plant Molecular Biology Reporter 28:460-472 doi:10.1007/s11105-009-0173-5.

Debernardi JM et al. (2014) Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity The Plant Journal 79:413-426 doi:10.1111/tpj.12567.

Dong Z, Han MH, Fedoroff N (2008) The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1 Proc Natl Acad Sci U S A 105:9970-9975 doi:10.1073/pnas.0803356105.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community Nucleic acids research 38:W64-70 doi:10.1093/nar/gkq310.

Eldem V, Çelikkol Akçay U, Ozhuner E, Bakır Y, Uranbey S, Unver T (2012) Genome-Wide Identification of miRNAs Responsive to Drought in Peach (*Prunus persica*) by High-Throughput Deep Sequencing PLoS ONE 7:e50298 doi:10.1371/journal.pone.0050298.

Gentile A, Dias LI, Mattos RS, Ferreira TH, Menossi M (2015) MicroRNAs and drought responses in sugarcane.

Henderson IR, Zhang X, Lu C, Johnson L, Meyers BC, Green PJ, Jacobsen SE (2006) Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning Nature genetics 38:721-725 doi:10.1038/ng1804.

Hu W, Wang T, Xu J, Li H (2014) MicroRNA mediates DNA methylation of target genes Biochemical and Biophysical Research Communications 444:676-681 doi:http://dx.doi.org/10.1016/j.bbrc.2014.01.171.

ICO (2016a) Historical Data on the Global Coffee Trade - Total production by all exporting countries. http://www.ico.org/new_historical.asp. Accessed 02/05/2017 2016.

ICO (2016b) Historical Data on the Global Coffee Trade - Total exports by all exporting countries. http://www.ico.org/new_historical.asp. Accessed 02/05/2017 2016.

Jeong D-H, Park S, Zhai J, Gurazada SGR, Paoli E, Meyers BC (2011) Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage Plant Cell 23 doi:10.1105/tpc.111.089045.

Jeong DH et al. (2013) Comprehensive investigation of microRNAs enhanced by analysis of sequence variants, expression patterns, ARGONAUTE loading, and target cleavage Plant Physiol 162:1225-1245 doi:10.1104/pp.113.219873.

Juarez MT, Kui JS, Thomas J, Heller BA, Timmermans MC (2004) microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity Nature 428:84-88 doi:10.1038/nature02363.

Jung J-H, Park C-M (2007) MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and floral development in Arabidopsis Planta 225:1327-1338 doi:10.1007/s00425-006-0439-1.

Jung JH, Seo YH, Seo PJ, Reyes JL, Yun J, Chua NH, Park CM (2007) The GIGANTEA-regulated microRNA172 mediates photoperiodic flowering independent of CONSTANS in Arabidopsis Plant Cell 19:2736-2748 doi:10.1105/tpc.107.054528.

Kidner CA, Martienssen RA (2004) Spatially restricted microRNA directs leaf polarity through ARGONAUTE1 Nature 428:81-84 doi:10.1038/nature02366.

Kim J et al. (2005) microRNA-directed cleavage of ATHB15 mRNA regulates vascular development in Arabidopsis inflorescence stems The Plant journal : for cell and molecular biology 42:84-94 doi:10.1111/j.1365-313X.2005.02354.x.

Kim YJ, Zheng B, Yu Y, Won SY, Mo B, Chen X (2011) The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana The EMBO journal 30:814-822 doi:10.1038/emboj.2011.3.

Kurihara Y, Takashi Y, Watanabe Y (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis RNA (New York, NY) 12:206-212 doi:10.1261/rna.2146906.

Lee H et al. (2010) Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in Arabidopsis Nucleic acids research 38:3081-3093 doi:10.1093/nar/gkp1240.

Li J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis Current biology : CB 15:1501-1507 doi:10.1016/j.cub.2005.07.029.

Li X et al. (2015) Characterization and comparative profiling of the small RNA transcriptomes in two phases of flowering in Cymbidium ensifolium BMC Genomics 16:622 doi:10.1186/s12864-015-1764-1.

Liang G, He H, Li Y, Wang F, Yu D (2014) Molecular mechanism of microRNA396 mediating pistil development in Arabidopsis Plant Physiol 164:249-258 doi:10.1104/pp.113.225144.

Lima AA (2015) Ethylene Regulation Under Different Watering Conditions and Its Possible Involvement in Coffee (*Coffea arabica* L.) Flowering (PhD Thesis). Universidade Federal de Lavras.

Liu H et al. (2014) OsmiR396d-Regulated OsGRFs Function in Floral Organogenesis in Rice through Binding to Their Targets OsJMJ706 and OsCR4 Plant Physiology 165:160-174.

Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R (2005) MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies Nature cell biology 7:719-723 doi:10.1038/ncb1274.

Liu J, Zhang X, Zhang F, Hong N, Wang G, Wang A, Wang L (2015) Identification and characterization of microRNAs from in vitro-grown pear shoots infected with Apple stem grooving virus in response to high temperature using small RNA sequencing BMC Genomics 16:945 doi:10.1186/s12864-015-2126-8.

Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA Science 297:2053-2056 doi:10.1126/science.1076311.

Lobbes D, Rallapalli G, Schmidt DD, Martin C, Clarke J (2006) SERRATE: a new player on the plant microRNA scene EMBO reports 7:1052-1058 doi:10.1038/sj.embor.7400806.

Luo Y, Guo Z, Li L (2013) Evolutionary conservation of microRNA regulatory programs in plant flower development Dev Biol 380:133-144 doi:10.1016/j.ydbio.2013.05.009.

Mallory AC, Reinhart BJ, Jones-Rhoades MW, Tang G, Zamore PD, Barton MK, Bartel DP (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region The EMBO journal 23:3356-3364 doi:10.1038/sj.emboj.7600340.

Mallory AC, Vaucheret H (2006) Functions of microRNAs and related small RNAs in plants Nature genetics 38 Suppl:S31-36 doi:10.1038/ng1791.

Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM (2006) The evolution and diversification of Dicers in plants FEBS Letters 580:2442-2450 doi: http://dx.doi.org/10.1016/j.febslet.2006.03.072.

Matiello JB, de Almeida SR, da Silva MB, Ferreira IB (2015) SIRIEMA VC 4, CULTIVAR CLONAL DE CAFEEIROS COM RESISTÊNCIA MÚLTIPLA, À FERRUGEM E AO BICHO MINEIRO. Paper presented at the 41o Congresso de Pesquisas Cafeeiras, Poços de Caldas.

Mathioni, S.M., Kakrana, A., and Meyers, B.C. 2017. Characterization of plant small RNAs by next generation sequencing. *Curr. Protoc. Plant Biol.* 2:39-63. doi: 10.1002/cppb.20043.

Mi S et al. (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide Cell 133:116-127 doi:10.1016/j.cell.2008.02.034.

Morais H, Caramori PH, Koguishi MS, Ribeiro AMdA (2008) Escala fenológica detalhada da fase reprodutiva de Cooffea arabica Bragantia 67:257-260 doi:10.1590/s0006-87052008000100031.

Nagpal P et al. (2005) Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation Development (Cambridge, England) 132:4107-4118 doi:10.1242/dev.01955.

Oliveira RR, Cesarino I, Mazzafera P, Dornelas MC (2014) Flower development in Coffea arabica L.: new insights into MADS-box genes Plant Reproduction 27:79-94 doi:10.1007/s00497-014-0242-2.

Patel P, Ramachandruni SD, Kakrana A, Nakano M, Meyers BC (2016) miTRATA: a web-based tool for microRNA Truncation and Tailing

Analysis Bioinformatics (Oxford, England) 32:450-452 doi:10.1093/bioinformatics/btv583.

Qin X, Liu JH, Zhao WS, Chen XJ, Guo ZJ, Peng YL (2013) Gibberellin 20-oxidase gene OsGA20ox3 regulates plant stature and disease development in rice Molecular plant-microbe interactions : MPMI 26:227-239 doi:10.1094/mpmi-05-12-0138-r.

Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants Genes Dev 16:1616-1626 doi:10.1101/gad.1004402.

Ren G, Xie M, Dou Y, Zhang S, Zhang C, Yu B (2012) Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis Proc Natl Acad Sci U S A 109:12817-12821 doi:10.1073/pnas.1204915109.

Rena AB, Maestri M (1985) Fisiologia do Cafeeiro Informe Agropecuário 11:26-40.

Rogers K, Chen X (2013) Biogenesis, Turnover, and Mode of Action of Plant MicroRNAs Plant Cell 25:2383-2399 doi:10.1105/tpc.113.113159.

Roy S, Tripathi AM, Yadav A, Mishra P, Nautiyal CS (2016) Identification and Expression Analyses of miRNAs from Two Contrasting Flower Color Cultivars of Canna by Deep Sequencing PLoS ONE 11:e0147499 doi:10.1371/journal.pone.0147499.

Rubio-Somoza I, Weigel D (2013) Coordination of Flower Maturation by a Regulatory Circuit of Three MicroRNAs PLoS Genet 9:e1003374 doi:10.1371/journal.pgen.1003374.

Sato K, Hamada M, Asai K, Mituyama T (2009) CENTROIDFOLD: a web server for RNA secondary structure prediction Nucleic acids research 37:W277-280 doi:10.1093/nar/gkp367.

Schommer C, Bresso E, Spinelli S, Palatnik J (2012) Role of microRNA miR319 in plant development Signaling and Communication in Plants 15:29-47.

Shuai P, Liang D, Zhang Z, Yin W, Xia X (2013) Identification of drought-responsive and novel Populus trichocarpa microRNAs by high-throughput sequencing and their targets using degradome analysis BMC Genomics 14:233.

Spanudakis E, Jackson S (2014) The role of microRNAs in the control of flowering time Journal of Experimental Botany 65:365-380 doi:10.1093/jxb/ert453.

Sun GL (2012) MicroRNAs and their diverse functions in plants Plant Molecular Biology 80:17-36 doi:10.1007/s11103-011-9817-6.

Sunkar R, Girke T, Jain PK, Zhu JK (2005) Cloning and characterization of microRNAs from rice Plant Cell 17:1397-1411 doi:10.1105/tpc.105.031682.

Tian Y et al. (2014) Identification and characterization of microRNAs related to salt stress in broccoli, using high-throughput sequencing and bioinformatics analysis BMC Plant Biol 14:226 doi:10.1186/s12870-014-0226-2.

Vaucheret H (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations Genes & Development 20:759-771 doi:10.1101/gad.1410506.

Voinnet O (2009) Origin, biogenesis, and activity of plant microRNAs Cell 136:669-687 doi:10.1016/j.cell.2009.01.046.

Wang JW, Czech B, Weigel D (2009) miR156-regulated SPL transcription factors define an endogenous flowering pathway in Arabidopsis thaliana Cell 138:738-749 doi:10.1016/j.cell.2009.06.014.

Wang T, Chen L, Zhao M, Tian Q, Zhang W-H (2011) Identification of drought-responsive microRNAs in Medicago truncatula by genome-wide high-throughput sequencing BMC Genomics 12:367.

Wang T, Pan H, Wang J, Yang W, Cheng T, Zhang Q (2014) Identification and profiling of novel and conserved microRNAs during the flower opening process in Prunus mume via deep sequencing Molecular Genetics and Genomics 289:169-183 doi:10.1007/s00438-013-0800-6.

Williams L, Grigg SP, Xie M, Christensen S, Fletcher JC (2005) Regulation of Arabidopsis shoot apical meristem and lateral organ formation by microRNA miR166g and its AtHD-ZIP target genes Development (Cambridge, England) 132:3657-3668 doi:10.1242/dev.01942.

Wollmann H, Mica E, Todesco M, Long JA, Weigel D (2010) On reconciling the interactions between APETALA2, miR172 and AGAMOUS with the ABC model of flower development Development (Cambridge, England) 137:3633-3642 doi:10.1242/dev.036673.

Wu MF, Tian Q, Reed JW (2006) Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction Development (Cambridge, England) 133:4211-4218 doi:10.1242/dev.02602.

Xie F, Stewart CN, Taki FA, He Q, Liu H, Zhang B (2014) High-throughput deep sequencing shows that microRNAs play important roles in switchgrass responses to drought and salinity stress Plant Biotechnology Journal 12:354-366 doi:10.1111/pbi.12142.

Yamaguchi A, Abe M (2012) Regulation of reproductive development by non-coding RNA in Arabidopsis: to flower or not to flower J Plant Res 125:693-704 doi:10.1007/s10265-012-0513-7.

Yan J et al. (2012) Effective Small RNA Destruction by the Expression of a Short Tandem Target Mimic in Arabidopsis The Plant Cell 24:415-427.

Yang C-Y et al. (2015) MicroRNA396-Targeted SHORT VEGETATIVE PHASE Is Required to Repress Flowering and Is Related to the Development of Abnormal Flower Symptoms by the Phyllody Symptoms1 Effector Plant Physiology 168:1702-1716 doi:10.1104/pp.15.00307.

Zeng Y, Cullen BR (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5 Nucleic acids research 32:4776-4785 doi:10.1093/nar/gkh824.

Zhang T, Wang J, Zhou C (2015) The role of miR156 in developmental transitions in Nicotiana tabacum Science China Life sciences 58:253-260 doi:10.1007/s11427-015-4808-5.

Zhou C-M, Wang J-W (2013) Regulation of Flowering Time by MicroRNAs Journal of Genetics and Genomics 40:211-215 doi:http://dx.doi.org/10.1016/j.jgg.2012.12.003.

Zhou M et al. (2011) Genome-wide analysis of clustering patterns and flanking characteristics for plant microRNA genes The FEBS journal 278:929-940 doi:10.1111/j.1742-4658.2011.08008.x.

Zhou X, Ruan J, Wang G, Zhang W (2007) Characterization and Identification of MicroRNA Core Promoters in Four Model Species PLoS Computational Biology 3:e37 doi:10.1371/journal.pcbi.0030037.

Supplementary Table 1. Normalized abundance (TP30M) of all miRNAs identified in G4 and G5 phases in Siriema and Catuai, average of G4 and G5 of 6 replicates, p-value and False Discovery Rate (FDR).

Supplementary Table 2. Target prediction of the mature miRNAs in the C. canephora genome with psRNATarget. miRNA names, Target ID (Locus Name) in C. canephora, Expectation scoring, unpaired energy (UPE) required to open the secondary structure around the miRNA target site, the start and end position on the miRNA and the Target, the sequence alignment of the miRNA and Target sequences, and the type of inhibition method.

miRNA	G4 Cat R1	G4 Cat R2	G4 Cat R3	G4 Sir R1	G4 Sir R2	G4 Sir R3	G5 Sir R1	G5 Sir R2	G5 Sir R3	G5 Cat R1	G5 Cat R2	G5 Cat R3	g4 ave.	g5 ave.	LOG2	p.value	FDR
ccp-m009	1	0	4	0	0	0	1	1	2	0	0	0	0,83	0,67	-0,322	0,8265	0,935
ccp-m016	1	1	7	2	0	0	1	1	5	0	1	1	1,83	1,50	-0,290	0,8029	0,932
ccp-m025	0	0	0	0	0	1	0	0	1	0	0	1	0,17	0,33	1,000	0,5497	0,775
ccp-m052	1	0	0	0	0	0	0	1	0	0	0	0	0,17	0,17	0,000	1,0000	1,000
ccp-m054	5	7	11	15	2	13	17	6	16	0	2	12	8,83	8,83	0,000	1,0000	1,000
ccp-m061	0	0	2	2	0	1	0	2	0	0	1	0	0,83	0,50	-0,737	0,5416	0,774
ccp-m075	0	0	1	19	1	8	23	0	38	0	1	0	4,83	10,33	1,096	0,4784	0,739
ccp-m079	0	0	0	0	0	1	0	0	1	0	0	0	0,17	0,17	0,000	1,0000	1,000
ccp-m081	0	1	0	2	1	1	2	1	1	0	2	0	0,83	1,00	0,263	0,7344	0,889
ccp-m098	0	0	0	1	0	0	0	1	2	0	0	0	0,17	0,50	1,585	0,4086	0,660
ccp-m123	0	0	0	0	0	0	1	0	0	0	1	0	0,00	0,33	-	0,1747	0,381
ccp-miR396_3	8	0	0	0	0	0	0	0	0	0	0	0	1,33	0,00	-	0,3632	0,632

ccp-miR530_1	0	3	0	0	0	0	0	0	0	0	0	0	0,50	0,00	-	0,3632	0,632
ccp-m019	7	0	11	12	2	5	18	5	10	1	4	21	6,17	9,83	0,673	0,3668	0,632
ccp-m038	0	0	0	0	0	0	0	1	0	1	0	0	0,00	0,33	-	0,1747	0,381
ccp-m072	1	0	12	6	0	1	5	1	4	1	11	13	3,33	5,83	0,807	0,4015	0,659
ccp-m030	27	6	30	14	5	6	10	17	8	3	11	29	14,67	13,00	-0,174	0,7833	0,932
ccp-m041	7	1	20	8	0	3	5	1	5	3	2	4	6,50	3,33	-0,963	0,3456	0,616
ccp-m065	0	1	0	0	1	0	2	5	2	3	5	5	0,33	3,67	3,459	0,0020	0,026
ccp-m101	9	3	25	13	0	4	15	4	17	3	10	26	9,00	12,50	0,474	0,5111	0,755
ccp-m120	3	0	14	3	1	2	12	11	11	3	5	25	3,83	11,17	1,543	0,0852	0,264
ccp-miR477_1	0	1	4	9	5	12	8	15	2	3	7	5	5,17	6,67	0,368	0,5886	0,815
ccp-m121	8	7	32	20	2	8	7	15	25	4	15	16	12,83	13,67	0,091	0,8820	0,963
ccp-m043	12	4	9	7	2	1	6	5	10	6	4	7	5,83	6,33	0,119	0,8032	0,932
ccp-m093	8	1	6	9	8	4	4	4	7	6	5	1	6,00	4,50	-0,415	0,3439	0,616

ccp-m104	15	14	22	2	6	7	8	26	9	6	26	3	11,00	13,00	0,241	0,7068	0,870
ccp-m119	1	0	4	1	0	0	2	2	4	6	4	13	1,00	5,17	2,369	0,0569	0,215
ccp-miR2111_1	13	19	8	10	21	8	5	17	8	6	11	5	13,17	8,67	-0,603	0,1642	0,374
ccp-m012	12	12	45	20	2	4	17	20	7	7	28	42	15,83	20,17	0,349	0,6180	0,837
ccp-m088	15	4	7	67	77	108	126	13	106	7	9	20	46,33	46,83	0,015	0,9863	1,000
ccp-m021	11	6	6	5	1	3	7	4	4	9	10	7	5,33	6,83	0,358	0,4039	0,659
ccp-m109	8	3	28	3	2	1	11	5	21	9	12	16	7,50	12,33	0,718	0,3438	0,616
ccp-m051	16	11	44	16	0	4	9	11	8	12	13	33	15,17	14,33	-0,082	0,9130	0,971
ccp-miR397_1	25	4	22	12	1	4	5	4	5	13	6	10	11,33	7,17	-0,661	0,3781	0,644
ccp-m049	25	0	30	14	0	8	15	9	16	15	6	14	12,83	12,50	-0,038	0,9529	0,990
ccp-miR319_1	13	19	18	7	0	1	7	4	11	16	11	3	9,67	8,67	-0,158	0,8057	0,932
ccp-m096	5	10	12	5	0	1	28	18	27	17	26	18	5,50	22,33	2,022	0,0002	0,009
ccp-miR169_1	12	18	18	8	4	4	19	18	15	17	45	29	10,67	23,83	1,160	0,0401	0,183

ccp-miR171_3	23	32	43	28	3	26	28	29	56	19	34	5	25,83	28,50	0,142	0,7671	0,922
ccp-m029	40	10	38	9	2	11	17	10	16	22	16	21	18,33	17,00	-0,109	0,8533	0,939
ccp-m067	44	29	67	30	10	17	8	13	23	22	17	25	32,83	18,00	-0,867	0,1412	0,359
ccp-miR164_1	15	8	17	19	7	15	32	42	35	22	15	65	13,50	35,17	1,381	0,0277	0,157
ccp-m020	12	8	18	21	0	10	34	15	26	23	15	54	11,50	27,83	1,275	0,0436	0,188
ccp-m055	17	8	28	17	6	11	49	48	35	23	38	52	14,50	40,83	1,494	0,0010	0,025
ccp-miR171_5	23	29	21	3	2	4	29	14	26	23	29	18	13,67	23,17	0,761	0,1250	0,336
ccp-m027	39	6	40	23	2	6	32	13	36	25	15	30	19,33	25,17	0,380	0,4877	0,739
ccp-m032	100	10	169	21	6	12	56	41	47	25	35	48	53,00	42,00	-0,336	0,7069	0,870
ccp-m035	32	17	32	42	16	14	30	28	28	25	39	43	25,50	32,17	0,335	0,2582	0,507
ccp-m024	48	36	52	33	6	25	24	19	17	26	24	55	33,33	27,50	-0,278	0,5254	0,768
ccp-m074	36	8	65	50	0	18	62	14	68	26	35	103	29,50	51,33	0,799	0,2268	0,469

ccp-m102	15	11	18	7	3	6	8	17	8	26	24	14	10,00	16,17	0,693	0,1492	0,367
ccp-miR395_2	25	7	43	0	1	1	4	10	13	26	7	12	12,83	12,00	-0,097	0,9180	0,971
ccp-m037	24	19	25	12	7	9	23	34	25	28	35	29	16,00	29,00	0,858	0,0076	0,054
ccp-m053	12	7	26	21	4	9	21	22	20	28	21	60	13,17	28,67	1,122	0,0667	0,230
ccp-m059	39	19	41	98	53	71	38	29	27	28	22	37	53,50	30,17	-0,827	0,0952	0,287
ccp-m040	35	1	38	20	3	7	25	28	26	29	21	79	17,33	34,67	1,000	0,1533	0,370
ccp-m057	4	7	5	5	2	4	20	15	11	29	11	51	4,50	22,83	2,343	0,0327	0,163
ccp-m046	52	4	70	41	1	10	24	20	39	31	19	31	29,67	27,33	-0,118	0,8542	0,939
ccp-miR171_4	33	18	48	21	7	9	7	15	11	31	15	7	22,67	14,33	-0,661	0,2873	0,543
ccp-m002	43	8	77	33	2	15	41	40	43	33	28	38	29,67	37,17	0,325	0,5444	0,774
ccp-miR156_1	11	47	26	171	47	72	26	73	30	33	30	90	62,33	47,00	-0,407	0,5715	0,798
ccp-miR169_2	41	33	27	42	63	61	24	62	28	35	71	81	44,50	50,17	0,173	0,6365	0,843
ccp-m017	31	14	58	21	5	30	35	35	48	36	23	54	26,50	38,50	0,539	0,2047	0,435

ccp-miR172_1	24	11	31	28	28	22	58	52	47	36	34	55	24,00	47,00	0,970	0,0013	0,025
ccp-m013	56	50	86	39	32	55	32	22	60	41	24	41	53,00	36,67	-0,532	0,1199	0,336
ccp-m058	29	1	8	34	9	30	51	1	72	42	0	5	18,50	28,50	0,623	0,4913	0,739
ccp-m007	80	17	86	57	2	23	80	44	60	44	61	102	44,17	65,17	0,561	0,2500	0,500
ccp-m001	45	17	58	22	1	13	38	20	50	45	61	58	26,00	45,33	0,802	0,1028	0,301
ccp-miR399_2	27	99	73	526	125	105	29	36	53	45	19	62	159,17	40,67	-1,969	0,1738	0,381
ccp-m087	25	14	13	9	24	22	61	82	94	47	56	113	17,83	75,50	2,082	0,0020	0,026
ccp-m033	83	39	84	77	28	101	43	67	38	49	58	71	68,67	54,33	-0,338	0,3017	0,564
ccp-miR160_1	63	48	87	29	30	34	30	21	41	51	56	38	48,50	39,50	-0,296	0,4271	0,682
ccp-m010	84	34	83	31	4	13	31	43	53	52	28	82	41,50	48,17	0,215	0,6907	0,866
ccp-m031	32	1	78	23	1	1	42	38	76	58	61	24	22,67	49,83	1,137	0,0962	0,287
ccp-m064	31	25	70	31	0	15	72	72	87	60	100	68	28,67	76,50	1,416	0,0025	0,026
ccp-m073	27	28	76	71	4	35	224	90	158	60	92	200	40,17	137,33	1,774	0,0140	0,087

ccp-m028	160	58	165	82	9	58	75	106	105	63	62	162	88,67	95,50	0,107	0,8234	0,935
ccp-m004	72	50	99	49	4	34	52	39	35	65	61	38	51,33	48,33	-0,087	0,8394	0,939
ccp-m039	126	10	201	74	2	16	92	64	67	67	71	174	71,50	89,17	0,319	0,6444	0,846
ccp-m006	151	79	122	94	11	72	50	69	75	68	67	147	88,17	79,33	-0,152	0,7212	0,880
ccp-m008	97	17	153	95	7	44	173	98	142	70	89	198	68,83	128,33	0,899	0,0831	0,263
ccp-m050	37	12	61	55	6	22	72	73	92	71	46	73	32,17	71,17	1,146	0,0069	0,051
ccp-miR172_2	12	4	8	8	4	5	59	53	82	76	43	55	6,83	61,33	3,166	0,0002	0,009
ccp-miR399_3	32	150	103	417	210	96	29	23	88	77	7	26	168,00	41,67	-2,011	0,0718	0,240
ccp-m122	63	22	76	35	1	13	57	53	67	80	57	110	35,00	70,67	1,014	0,0391	0,183
ccp-m078	17	4	38	19	1	15	238	61	272	81	86	267	15,67	167,50	3,418	0,0140	0,087
ccp-mir482_2	136	40	81	45	10	67	83	70	92	93	68	118	63,17	87,33	0,467	0,2491	0,500
ccp-miR3627_1	196	142	242	242	98	144	75	64	83	95	90	194	177,33	100,17	-0,824	0,0324	0,163
ccp-m045	45	17	106	41	4	29	167	134	172	96	161	173	40,33	150,50	1,900	0,0002	0,009

ccp-miR156_2	9	4	31	23	7	13	179	245	217	99	116	135	14,50	165,17	3,510	0,0012	0,025
ccp-m060	88	47	119	82	7	39	118	88	87	105	94	148	63,67	106,67	0,744	0,0533	0,207
ccp-m091	183	194	235	262	174	306	101	134	133	105	81	113	225,67	111,17	-1,021	0,0018	0,026
ccp-m084	77	62	86	207	36	109	101	76	93	106	79	115	96,17	95,00	-0,018	0,9645	0,990
ccp-m068	111	90	93	65	36	60	73	104	96	115	101	46	75,83	89,17	0,234	0,3996	0,659
ccp-m111	80	58	76	82	27	77	48	65	74	115	75	13	66,67	65,00	-0,037	0,9207	0,971
ccp-m048	143	10	175	110	6	43	155	86	145	122	102	279	81,17	148,17	0,868	0,1301	0,336
ccp-m022	124	28	157	75	18	65	100	133	132	127	133	182	77,83	134,50	0,789	0,0533	0,207
ccp-miR398_1	1477	121	563	376	86	269	212	69	215	128	69	102	482,00	132,50	-1,863	0,1601	0,370
ccp-m056	83	69	117	66	27	39	110	117	151	131	102	189	66,83	133,33	0,996	0,0050	0,046
ccp-miR479_1	132	101	151	216	20	121	268	145	283	138	118	96	123,50	174,67	0,500	0,2514	0,500
ccp-m062	123	76	170	196	9	61	202	159	160	140	114	546	105,83	220,17	1,057	0,1584	0,370
ccp-m077	182	95	211	127	11	72	138	123	160	157	167	79	116,33	137,33	0,239	0,5427	0,774

ccp-m014	223	101	267	179	61	206	167	113	232	161	133	295	172,83	183,50	0,086	0,8054	0,932
ccp-m023	262	180	268	355	119	225	239	196	203	164	164	263	234,83	204,83	-0,197	0,4415	0,691
ccp-m092	108	74	164	194	19	120	335	300	229	164	150	351	113,17	254,83	1,171	0,0100	0,067
ccp-miR167_1	77	59	84	94	18	49	165	87	169	166	146	114	63,50	141,17	1,153	0,0015	0,026
ccp-m105	64	26	52	41	11	16	108	105	145	173	120	88	35,00	123,17	1,815	0,0003	0,009
ccp-m116	135	105	158	34	41	72	71	127	73	195	167	97	90,83	121,67	0,422	0,3172	0,585
ccp-m113	339	96	627	258	13	107	410	254	416	201	320	568	240,00	361,50	0,591	0,2843	0,543
ccp-m097	222	328	215	128	111	99	260	411	210	208	434	166	183,83	281,50	0,615	0,1287	0,336
ccp-miR394_1	167	185	200	221	109	229	226	219	248	223	229	13	185,17	193,00	0,060	0,8515	0,939
ccp-m034	382	132	325	142	63	107	139	156	87	234	192	393	191,83	200,17	0,061	0,9056	0,971
ccp-m124	268	198	456	382	56	340	395	233	318	234	247	89	283,33	252,67	-0,165	0,6787	0,862
ccp-m036	349	28	588	128	4	60	218	174	201	262	278	263	192,83	232,67	0,271	0,6930	0,866
ccp-m086	282	163	248	194	39	139	180	139	244	272	151	185	177,50	195,17	0,137	0,6786	0,862

ccp-miR395_1	470	169	724	36	39	69	134	107	255	272	123	111	251,17	167,00	-0,589	0,5103	0,755
ccp-miR399_1	162	434	575	359	159	192	92	130	311	281	125	173	313,50	185,33	-0,758	0,1454	0,363
ccp-miR156_4	13	15	17	36	11	37	1543	948	968	285	400	1053	21,50	866,17	5,332	0,0065	0,050
ccp-m005	52	74	102	48	7	25	131	213	146	300	270	139	51,33	199,83	1,961	0,0026	0,026
ccp-m015	223	37	208	265	18	94	434	353	475	301	332	681	140,83	429,33	1,608	0,0027	0,026
ccp-miR408_1	302	96	288	89	53	143	268	196	296	319	287	282	161,83	274,67	0,763	0,0500	0,204
ccp-miR171_2	236	236	313	165	26	109	247	213	314	394	458	253	180,83	313,17	0,792	0,0437	0,188
ccp-m100	147	402	184	242	45	123	184	240	305	461	417	80	190,50	281,17	0,562	0,2671	0,517
ccp-m094	471	252	882	410	80	305	377	380	348	490	383	833	400,00	468,50	0,228	0,6229	0,837
ccp-m080	381	128	736	468	12	145	850	291	773	512	456	820	311,67	617,00	0,985	0,0609	0,225
ccp-m076	626	309	821	446	34	303	539	375	668	545	452	494	423,17	512,17	0,275	0,4830	0,739
ccp-miR393_1	431	441	512	296	142	234	711	251	921	569	593	394	342,67	573,17	0,742	0,0728	0,240

ccp-miR171_1	494	420	692	707	44	265	1033	936	971	633	705	1110	437,00	898,00	1,039	0,0059	0,048
ccp-miR167_2	467	344	697	354	53	265	775	612	993	650	606	292	363,33	654,67	0,849	0,0465	0,195
ccp-m083	494	343	407	332	172	347	738	677	588	736	530	1033	349,17	717,00	1,038	0,0022	0,026
ccp-miR399_4	486	727	808	718	257	453	500	793	673	836	733	1025	574,83	760,00	0,403	0,1296	0,336
ccp-miR7122_1	1557	431	1923	5465	585	3828	2231	1715	2260	988	778	3425	2298,17	1899,50	-0,275	0,6698	0,862
ccp-m066	558	1277	387	940	2252	552	1111	1306	747	1050	1117	412	994,33	957,17	-0,055	0,9088	0,971
ccp-m071	869	424	342	430	604	239	348	727	500	1143	1516	1131	484,67	894,17	0,884	0,0813	0,263
ccp-miR390_1	750	694	981	636	392	850	882	1190	842	1498	1507	613	717,17	1088,67	0,602	0,0636	0,229
ccp-m110	2247	873	2681	1463	60	629	1797	992	1778	1779	1432	2395	1325,50	1695,50	0,355	0,4378	0,691
ccp-m106	1777	1072	1822	1038	418	1393	1227	1165	1474	1833	1756	1408	1253,33	1477,17	0,237	0,3847	0,648
ccp-miR319_2	3596	2262	5737	3887	685	2261	4758	4082	4417	2718	4256	4529	3071,33	4126,67	0,426	0,2136	0,447
ccp-m118	3046	872	5477	2063	178	1673	3952	2716	5179	3006	2604	6029	2218,17	3914,33	0,819	0,1100	0,316

ccp-miR168_1	2165	2385	2564	1718	1056	1923	2241	2120	2968	3316	2688	2780	1968,50	2685,50	0,448	0,0322	0,163
ccp-miR156_3	7201	4261	12656	6936	722	4319	2459	2346	3379	3540	2649	3737	6015,83	3018,33	-0,995	0,1273	0,336
ccp-m095	5128	2826	5995	10663	1080	5610	5692	3876	5107	3774	2994	7683	5217	4854,33	-0,104	0,8156	0,935
ccp-miR482_4	5345	2532	5711	8063	1759	9368	6145	4227	6288	4419	3569	4682	5463	4888,33	-0,16	0,6722	0,862
ccp-miR162_1	7244	4669	11078	6257	1604	4222	6102	4558	7359	4768	4814	6870	5845	5745,17	-0,025	0,9449	0,990
ccp-m069	736	792	896	340	304	459	4937	10331	6818	6981	8322	6307	587	7282,67	3,631	0,0003	0,009
ccp-miR403_1	8473	4852	16152	6874	618	3914	16857	8186	17590	9895	11349	25343	6813	14870,00	1,126	0,0393	0,183
ccp-m090	12782	8040	14386	9792	1686	7599	11116	10076	11403	12321	11943	17165	9047	12337,33	0,447	0,1553	0,370
ccp-miR396_4	4276	1449	3772	4999	597	2222	19170	9325	22592	13583	10570	29734	2885	17495,67	2,600	0,0054	0,046
ccp-m085	5887	7154	4774	1732	2581	3792	7351	17191	6193	14095	26147	9974	4320	13491,83	1,643	0,0284	0,157
ccp-miR482_1	9849	3775	4428	3628	7978	3725	4714	9333	5009	14472	19761	17670	5563	11826,50	1,088	0,0656	0,230

ccp-miR396_5	21051	11633	31563	52722	2923	28421	137665	68839	114784	38885	37149	79727	24718	79508,17	1,685	0,0194	0,115
ccp-m099	38220	42871	41490	16845	16590	19233	33489	46778	35303	55501	88603	15625	29208	45883,17	0,652	0,1857	0,400
ccp-mir482_3	57548	27043	64819	16222	14691	36556	24018	23563	36770	58302	41362	29969	36146	35664,00	-0,019	0,9631	0,990
ccp-miR166_2	1558352	798405	1308965	1725198	505858	1471374	1019153	1091347	1312119	1161468	887218	1268931	122802	1123372,67	-0,129	0,6267	0,837
ccp-miR166_1	3303893	2562230	4408574	2228905	977691	2960813	2744550	2514530	3168685	3908375	3335749	2474227	2740351	3024352,67	0,142	0,6009	0,824

RNA Acc.	Target Acc.	miRNA start	miRNA end	Target start	Target end	miRNA aligned fragment	Target aligned fragment	Inhibition	Target Description
ccp-m091	Cc11_g04230	1	20	1488	1507	UCUUCCCCUCUUCGGCUUAC	GAAAGCUGAAGAGGGGAAGA	Cleavage	Polyol transporter 5
ccp-m091	Cc00_g08750	1	21	22	42	UCUUCCCCUCUUCGGCUUACU	AGAGAGCCGAGGUGGGGAAGA	Translation	Hypothetical protein
ccp-m091	Cc10_g08580	1	20	1044	1063	UCUUCCCCUCUUCGGCUUAC	GGAAGCCGAGGAGGAGAAGA	Cleavage	Putative unknown protein
ccp-m091	Cc03_g00650	1	21	24	44	UCUUCCCCUCUUCGGCUUACU	AGGAAGAGGAAGAGGGGAAGA	Cleavage	Putative uncharacterized protein
ccp-m065	Cc03_g09900	1	21	321	341	GCAUCAGAGGAGUCAGGCAGG	CCUGCUUGAUACCUCUGAUGC	Translation	Probable aldo-keto reductase 2
ccp-m065	Cc01_g01000	1	21	1865	1885	GCAUCAGAGGAGUCAGGCAGG	CCUGCUUGAUUCCUUUGCUGC	Cleavage	Pentatricopeptide repeat-containing protein At2g21090
ccp-m065	Cc01_g19440	1	21	1872	1892	GCAUCAGAGGAGUCAGGCAGG	CUUGCUUGCUUCUUCUGAUGC	Cleavage	ARM REPEAT PROTEIN INTERACTING WITH ABF2
ccp-m065	Cc07_g13300	1	20	1297	1316	GCAUCAGAGGAGUCAGGCAG	AUGUUUGAUUUCUCUGAUGC	Cleavage	ARM repeat superfamily protein
ccp-m065	Cc02_g20590	1	21	2559	2579	GCAUCAGAGGAGUCAGGCAGG	UCUUCUUGACUCCUCUGGUGA	Cleavage	DNA polymerase V family
ccp-m065	Cc03_g03740	1	21	4385	4405	GCAUCAGAGGAGUCAGGCAGG	UCUUUCUGAUUCUUCUGGUGC	Cleavage	Putative Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein
ccp-m065	Cc06_g07870	1	20	2103	2122	GCAUCAGAGGAGUCAGGCAG	AUGCCUAAUUUCUCUGAUGC	Cleavage	Putative chloride channel-like protein CLC-g
ccp-m065	Cc07_g08040	1	20	1391	1410	GCAUCAGAGGAGUCAGGCAG	CUGCGUGACUUAUCUGAUGC	Translation	GDP-L-galactose phosphorylase 1
ccp-m065	Cc05_g15020	1	21	704	724	GCAUCAGAGGAGUCAGGCAGG	UCUGCUUGCUUCCACUGAUGC	Cleavage	E3 ubiquitin-protein ligase UPL1
ccp-m065	Cc05_g15750	1	20	4	23	GCAUCAGAGGAGUCAGGCAG	CUGGCUGGUUCAUCUGAUGC	Translation	Putative uncharacterized protein
ccp-m096	Cc02_g02940	1	21	16	36	UGAUUCAUUCAGCAAGCUGUG	CAAAGCAUGCUGAAUGGAUCA	Cleavage	Arabidopsis thaliana protein of unknown function (DUF821)
ccp-m096	Cc02_g29190	1	20	1165	1184	UGAUUCAUUCAGCAAGCUGU	ACGGCUGGCUGGAUGAAUCC	Cleavage	Putative F-box/FBD/LRR-repeat protein At1g78750
ccp-m096	Cc11_g06500	1	20	167	186	UGAUUCAUUCAGCAAGCUGU	ACAGCUUGAUAGAUGAAUCA	Translation	Putative Probable LRR receptor-like serine/threonine-protein kinase At4g08850
ccp-m096	Cc06_g17490	1	20	209	228	UGAUUCAUUCAGCAAGCUGU	ACAGCUAUUUGGAUGAAUCA	Cleavage	Transcription factor bHLH35
ccp-m055	Cc09_g02540	1	20	922	941	CAUGUGCCUGUGUUCUCCAU	AUGGAGAAAACAGGCACAUU	Cleavage	Protein of unknown function (DUF155)

ccp-m055	Cc02_g15690	1	21	1392	1412	CAUGUGCCUGUGUUCUCCAUC	GGUGAAGUACACAGGCAUAUG	Cleavage	Putative AAA-type ATPase family protein
ccp-m055	Cc03_g00710	1	20	2730	2749	CAUGUGCCUGUGUUCUCCAU	AUGGAGAACAUGUGUGCAUG	Cleavage	unknown protein%3B FUNCTIONS IN
ccp-m055	Cc01_g09290	1	21	820	840	CAUGUGCCUGUGUUCUCCAUC	GAUGAUGAUCACAGGCACAUG	Cleavage	Transcription factor TGA2
ccp-m055	Cc07_g05880	1	20	203	221	CAUGUGCCUGUGUUCUCCAU	AUGGAGAACACAGGC-CGUG	Cleavage	Putative Patatin group A-3
ccp-m087	Cc07_g08120	1	21	48	68	UCUCCAGUGGAUUCUCUCUCC	GGAGAGAGAGUCCGUGGGAGA	Cleavage	Putative Acyl-CoA--sterol O-acyltransferase 1
ccp-m087	Cc11_g09620	1	21	220	240	UCUCCAGUGGAUUCUCUCUCC	GGAGAGAGUAUUCAUUGGAGC	Cleavage	Elongator complex protein 2
ccp-m087	Cc04_g10790	1	20	817	836	UCUCCAGUGGAUUCUCUCUC	GAGGGAGUAUUUACUGGAGA	Cleavage	Vignain
ccp-m087	Cc02_g08600	1	20	438	457	UCUCCAGUGGAUUCUCUCUC	GAGGGAGGAUUCAGUGGAGA	Cleavage	Pre-mRNA-splicing factor 3
ccp-m087	Cc10_g13920	1	20	2339	2358	UCUCCAGUGGAUUCUCUCUC	GAGGAAGAACCCAUUGGAGA	Translation	Putative Probable LRR receptor-like serine/threonine-protein kinase At2g16250
ccp-m087	Cc00_g02560	1	20	1573	1592	UCUCCAGUGGAUUCUCUCUC	GUGAGAGAACUUACUGGAGA	Translation	Putative Activating signal cointegrator 1 complex subunit 3
ccp-m064	Cc00_g27990	1	22	575	596	GAUUUCAACGUCGGAGCAAAGA	UUUUUGCUCCGAUGCUGGAGUU	Cleavage	Glucan endo-1%2C3-beta-glucosidase
ccp-m064	Cc02_g36730	1	22	1791	1812	GAUUUCAACGUCGGAGCAAAGA	UAUUUGUUCUGCUGUUGAAAUC	Translation	Putative Dihydroflavonol-4-reductase
ccp-m064	Cc05_g10130	1	20	609	628	GAUUUCAACGUCGGAGCAAA	UUUGCUCAGAUGUUGAAAGC	Cleavage	Probable sugar phosphate/phosphate translocator At3g14410
ccp-m064	Cc02_g37200	1	24	353	376	GAUUUCAACGUCGGAGCAAAGAGU	ACCGUUUGCUUCCAUGGUGAAAUC	Cleavage	Putative uncharacterized protein
ccp-miR172_2	Cc07_g06200	1	21	1826	1846	GGAAUCUUGAUGAUGCUGCAG	CUGCAGCAUCAUCAGGAUUCC	Cleavage	Putative uncharacterized protein
ccp-miR172_2	Cc09_g02450	1	21	1526	1546	GGAAUCUUGAUGAUGCUGCAG	CUGCAGCAUCAUCAGGAUUCU	Cleavage	Putative Ethylene-responsive transcription factor RAP2-7
ccp-miR172_2	Cc01_g11840	1	21	1876	1896	GGAAUCUUGAUGAUGCUGCAG	CUGCAGCAUCAUCAGGAUUCU	Cleavage	Putative Ethylene-responsive transcription factor RAP2-7
ccp-miR172_2	Cc06_g21970	1	21	554	574	GGAAUCUUGAUGAUGCUGCAG	UUGCAGCAUCCUUAAGAUUCC	Translation	Hypothetical protein
ccp-miR172_2	Cc01_g16800	1	21	177	197	GGAAUCUUGAUGAUGCUGCAG	CUUCAGCAUUGUCAAGAUUCA	Cleavage	B-cell receptor-associated protein 31-like
ccp-miR172_2	Cc02_g22390	1	21	351	371	GGAAUCUUGAUGAUGCUGCAG	CAGCAUCAUCAUCAAGGUUCC	Cleavage	Plant protein of unknown function (DUF639)

ccp-miR172_2	Cc08_g09580	1	20	3233	3252	GGAAUCUUGAUGAUGCUGCA	UGAAGCAGCAUCAAGAUUCC	Cleavage	Protein of unknown function (DUF810)
ccp-m045	Cc02_g08390	1	20	985	1004	AUGAAUGUAGUUUUCAACGC	AGGUUGGAAACUAUAUUCAU	Cleavage	ABC transporter B family member 6
ccp-m045	Cc05_g06850	1	20	1807	1826	AUGAAUGUAGUUUUCAACGC	UCUUUGAGAACUAUAUUCAU	Cleavage	Protochlorophyllide reductase%2C chloroplastic
ccp-m045	Cc04_g08400	1	20	377	396	AUGAAUGUAGUUUUCAACGC	GCGUUGAUAAUUAUGUUUAU	Cleavage	Miraculin
ccp-m045	Cc04_g08390	1	20	377	396	AUGAAUGUAGUUUUCAACGC	GCGUUGAUAAUUAUGUUUAU	Cleavage	Miraculin
ccp-m045	Cc01_g16970	1	22	1853	1874	AUGAAUGUAGUUUUCAACGCCA	UAGAGCUGAAAAUUAUAUUCAU	Cleavage	Putative Sec14 cytosolic factor
ccp-m045	Cc07_g21270	1	20	1541	1560	AUGAAUGUAGUUUUCAACGC	GCGAUGAAGAUGACAUUCAU	Translation	Putative disease resistance protein RGA3
ccp-miR156_2	Cc05_g11850	1	20	998	1017	UGACAGAAGAGAGUGAGCAC	UUGCUUACUCUCUUCUGUCA	Cleavage	Putative uncharacterized protein
ccp-miR156_2	Cc11_g17130	1	20	1053	1072	UGACAGAAGAGAGUGAGCAC	GUGCUCUCUCUCUUCUGUCA	Cleavage	Putative Squamosa promoter-binding-like protein 16
ccp-miR156_2	Cc02_g24550	1	20	959	978	UGACAGAAGAGAGUGAGCAC	GUGCUCUCUCUCUUCUGUCA	Cleavage	Putative uncharacterized protein
ccp-miR156_2	Cc06_g23710	1	20	1355	1374	UGACAGAAGAGAGUGAGCAC	GUGCUCUCUCUCUUCUGUCA	Cleavage	Putative uncharacterized protein
ccp-miR156_2	Cc11_g16990	1	20	1349	1368	UGACAGAAGAGAGUGAGCAC	GUGCUCUCUCUCUUCUGUCA	Cleavage	Putative squamosa promoter binding protein-like 2
ccp-miR156_2	Cc11_g11740	1	20	1293	1312	UGACAGAAGAGAGUGAGCAC	GUGCUCUCUCUCUUCUGUCA	Cleavage	Putative uncharacterized protein
ccp-miR156_2	Cc02_g13600	1	20	738	757	UGACAGAAGAGAGUGAGCAC	AUGCUCUCUCUCUUCUGUCA	Cleavage	Squamosa promoter-binding protein 1
ccp-miR156_2	Cc05_g07500	1	20	1028	1047	UGACAGAAGAGAGUGAGCAC	UUGCUCUCUCUCUUCUGUCA	Cleavage	Putative uncharacterized protein
ccp-miR156_2	Cc06_g21540	1	20	16	35	UGACAGAAGAGAGUGAGCAC	CUGCUCAUGCUUUUCUGUCA	Cleavage	Hypothetical protein
ccp-miR156_2	Cc07_g01140	1	20	597	616	UGACAGAAGAGAGUGAGCAC	GCGCUCUCUCUUUUCUGUCG	Cleavage	Putative CASP-like protein POPTRDRAFT_752786
ccp-miR156_2	Cc11_g15480	1	20	1078	1097	UGACAGAAGAGAGUGAGCAC	GUGUUCUCUAUUUUCUGUCA	Translation	High affinity nitrate transporter 2.5
ccp-miR167_1	Cc01_g16410	1	20	2327	2346	UGAAGCUGCCAGCAUGAUCU	AGGUUAUGCUGGCUGUUUCA	Cleavage	Peptidyl-prolyl cis-trans isomerase FKBP62
ccp-miR167_1	Cc09_g03270	1	20	1920	1939	UGAAGCUGCCAGCAUGAUCU	AGAUUAUGCUGGUGGCUACA	Cleavage	Putative uncharacterized protein
ccp-miR167_1	Cc03_g11920	1	20	327	346	UGAAGCUGCCAGCAUGAUCU	UGGUUAUGCUAGCAGCUUCA	Translation	Putative Gibberellin 20 oxidase 1-D

ccp-miR167_1	Cc03_g11570	1	20	294	313	UGAAGCUGCCAGCAUGAUCU	UGGUUAUGCUAGCAGCUUCA	Translation	Gibberellin 20 oxidase 1-D
ccp-miR167_1	Cc03_g12140	1	20	399	418	UGAAGCUGCCAGCAUGAUCU	UGGUUAUGCUAGCAGCUUCA	Translation	Gibberellin 20 oxidase 1-D
ccp-miR167_1	Cc00_g03550	1	20	408	427	UGAAGCUGCCAGCAUGAUCU	UGGUUAUGCUAGCAGCUUCA	Translation	Gibberellin 20 oxidase 1
ccp-miR167_1	Cc00_g03560	1	20	495	514	UGAAGCUGCCAGCAUGAUCU	UGGUUAUGCUAGCAGCUUCA	Translation	Gibberellin 20 oxidase 1-D
ccp-miR167_1	Cc08_g10510	1	20	677	696	UGAAGCUGCCAGCAUGAUCU	AAAUGAUGCUGGUAGCUUCA	Cleavage	K( ) efflux antiporter 4
ccp-miR167_1	Cc10_g06820	1	20	630	649	UGAAGCUGCCAGCAUGAUCU	AGAUUGUGCUGAGAGCUUCA	Translation	Phospho-2-dehydro-3-deoxyheptonate aldolase
ccp-miR167_1	Cc08_g14790	1	20	1970	1989	UGAAGCUGCCAGCAUGAUCU	AGUUCAUGAUGGUAGCUUUA	Cleavage	Putative Transcription factor GLABRA 3
ccp-m105	Cc05_g08680	1	20	226	245	UUGCAUACGCGCCUGAAUCG	AGAUUCAGGUGCGUAUCCAA	Cleavage	Putative F-box/kelch-repeat protein At3g06240
ccp-m105	Cc00_g02930	1	20	166	185	UUGCAUACGCGCCUGAAUCG	CGAUUCAAGUGCGUUUGCAA	Cleavage	Putative F-box/kelch-repeat protein At3g06240
ccp-m005	Cc05_g05710	1	21	529	550	AACCAAUGACUAU-UCAUGAUU	AAUCAUGAUAUAGUCAUUGGUG	Cleavage	DNA repair metallo-beta-lactamase family protein
ccp-m005	Cc10_g11830	1	21	436	456	AACCAAUGACUAUUCAUGAUU	AAUCAUGACUAGUUAAUGGUU	Cleavage	Hypothetical protein
ccp-m015	Cc03_g15970	1	20	1441	1460	AAUAUACUGAGAAAUGAGCC	GGCUCAUUUUUUGGUGUGUU	Cleavage	Pyruvate dehydrogenase E1 component subunit beta
ccp-m083	Cc07_g06820	1	20	80	99	UCGCAGAAGACAGCCGCAUC	GCUGCGUCUGUCUUCUGCGA	Cleavage	Uncharacterized protein At2g23090
ccp-m069	Cc04_g08930	1	20	23	42	GGGAUGGAACCUGAGAACAC	CUGGUUUUAGGUUCCAUCCC	Cleavage	NAC domain containing protein 57
ccp-m069	Cc00_g17350	1	20	1696	1715	GGGAUGGAACCUGAGAACAC	GAGUUCACAGGAUCCAUCCC	Translation	Putative LRR receptor-like serine/threonine-protein kinase GSO1
ccp-miR396_4	Cc06_g08100	1	20	404	423	UUCCACAGCUUUCUUGAACU	AGUUCAAGGAAGCUGUGGGA	Cleavage	Hypothetical protein
ccp-miR396_4	Cc04_g00510	1	21	276	296	UUCCACAGCUUUCUUGAACUG	UAGUUCAAGGAAACUGUGGAA	Translation	Vacuolar cation/proton exchanger 2
ccp-miR396_4	Cc09_g09000	1	20	320	339	UUCCACAGCUUUCUUGAACU	UGUUCAAGAAAGCUGAGGAA	Cleavage	Putative uncharacterized protein
ccp-miR396_4	Cc03_g04610	1	20	263	282	UUCCACAGCUUUCUUGAACU	AGUUUAAGAAGGCUGAGGAA	Cleavage	Pentatricopeptide repeat-containing protein At3g23020
ccp-miR396_4	Cc00_g15760	1	21	141	161	UUCCACAGCUUUCUUGAACUG	CAGUUGAAGAAAGUUCUGGAA	Cleavage	CONTAINS InterPro DOMAIN/s
ccp-miR396_4	Cc09_g08870	1	20	131	150	UUCCACAGCUUUCUUGAACU	UGUUCAAGAAGGCUGCGGAA	Cleavage	Mads box protein putative~ AGL62~ modules

ccp-miR396_4	Cc08_g05740	1	20	1727	1746	UUCCACAGCUUUCUUGAACU	AGUUUGAGAAAGUUGUGAAA	Cleavage	Chitinase-like protein 2
ccp-miR396_4	Cc11_g11830	1	21	1289	1309	UUCCACAGCUUUCUUGAACUG	CAGUUCAAUAAUGCUGUGGAC	Translation	Probable 6-phosphogluconolactonase 1
ccp-miR396_4	Cc06_g05390	1	21	783	804	UUCCACA-GCUUUCUUGAACUG	CCGUUCAAGAAAGCCUGUGGAA	Cleavage	Putative growth-regulating factor 4
ccp-miR396_4	Cc06_g12040	1	21	665	686	UUCCACA-GCUUUCUUGAACUG	CCGUUCAAGAAAGCCUGUGGAA	Cleavage	Putative uncharacterized protein
ccp-miR396_4	Cc02_g00930	1	21	825	846	UUCCACA-GCUUUCUUGAACUG	CCGUUCAAGAAAGCCUGUGGAA	Cleavage	Putative uncharacterized protein
ccp-miR396_4	Cc05_g03020	1	21	1983	2004	UUCCACAGCUUUCUUGA-ACUG	CAGUGUGAAGAAAGCUGUGGAA	Cleavage	Calcium-transporting ATPase 12%2C plasma membrane-type
ccp-miR396_4	Cc11_g10740	1	20	221	240	UUCCACAGCUUUCUUGAACU	AGUUCAAUGAGGUUGUGGAG	Cleavage	Putative Clavaminate synthase-like protein At3g21360
ccp-miR396_4	Cc07_g07350	1	19	1021	1040	UUCCACA-GCUUUCUUGAAC	GUUCAAGAAAGCCUGUGGAA	Cleavage	Putative growth-regulating factor 2
ccp-miR396_4	Cc07_g20540	1	20	2081	2100	UUCCACAGCUUUCUUGAACU	AGUUUAAUAAAGCUGUUGAA	Cleavage	Putative unknown protein%3B FUNCTIONS IN
ccp-miR396_4	Cc02_g36500	1	19	1027	1046	UUCCACA-GCUUUCUUGAAC	GUUCAAGAAAGCAUGUGGAA	Cleavage	Putative uncharacterized protein
ccp-miR396_4	Cc11_g03060	1	21	657	678	UUCCACAGCUUUCUUGA-ACUG	CGGUGUGAAGAAAGCUGUGGAA	Cleavage	Calcium-transporting ATPase 12%2C plasma membrane-type
ccp-miR396_4	Cc03_g04410	1	21	1596	1617	UUCCACAGCUUUCUUGA-ACUG	CGGUGUGAAGAAAGCUGUGGAA	Cleavage	Putative calcium-transporting ATPase 13%2C plasma membrane-type
ccp-miR396_4	Cc07_g09590	1	20	764	783	UUCCACAGCUUUCUUGAACU	AGAUCAAGAGAUCUGUGGGA	Translation	Sugar transporter ERD6-like 6
ccp-miR396_4	Cc02_g13260	1	20	1196	1215	UUCCACAGCUUUCUUGAACU	AGAUGGAGAAGGCUGUGGAA	Cleavage	Pentatricopeptide repeat-containing protein At4g21705%2C mitochondrial
ccp-miR396_4	Cc09_g10660	1	20	1109	1128	UUCCACAGCUUUCUUGAACU	AGUACAAGAAGGAUGUGGAG	Cleavage	SEC1 family transport protein SLY1
ccp-miR396_4	Cc05_g16590	1	20	1386	1405	UUCCACAGCUUUCUUGAACU	AGUGGGAGGAAGCUGUGGAA	Cleavage	Putative uncharacterized protein