Université de Lille

Ecole doctorale Science de la Matière, du Rayonnement et de l'Environnement – SMRE ED 104
UMR Evolution, Ecologie, Paléontologie – CNRS/Université de Lille – UMR 8198

Origin and evolutionary trajectory of microRNA genes in Arabidopsis species

Origine et trajectoire évolutive des gènes de microARN chez les espèces d'Arabidopsis

Par Flavia Pavan

Soutenue le 28 mars 2024 en vue d'obtenir le grade de docteur de l'Université de Lille en biologie de l'environnement, des organismes et des populations, écologie,

devant le jury composé de :

Rapporteurs:
Karine Alix (*Professeure,* AgroParisTech)
Laurent Duret (*Directeur de recherche* CNRS, LBBE - UMR 5558 CNRS/U. Lyon)

Examinateurs:
Clémentine Vitte (*Chargée de recherche* CNRS, GQE Le Moulon - UMR8120 CNRS/U. Paris-Saclay)
Hélène Touzet (*Directrice de recherche* CNRS, CRISTAL - UMR 9189 CNRS/U. Lille)

Directeur de thèse:
Vincent Castric (*Directeur de recherche* CNRS, EEP - UMR 8198 CNRS/U. Lille)

Encadrant:
Sylvain Legrand (*Maître de conférences*, EEP - UMR 8198 CNRS/U. Lille)

Présidente du Jury:
Hélène Touzet

EDSMRE

cnrs

Evo-Eco-Paléo

Table of contents

General introduction

1. Phenotypic evolution and regulation of gene expression

Understanding the molecular basis of observed phenotypic differences between species has long been a central question in evolutionary biology. According to the traditional view, distinct phenotypic characteristics between species should correspond to a high degree of genetic divergence. However, the seminal paper by King and Wilson (1975) revealed that the protein sequences of man and chimpanzee were strikingly similar despite the significant phenotypic differences between these two species. They proposed that the underlying cause of the biological differences between them lay not chiefly in the coding sequences of their genes but rather in variations in the way the expression of these genes was regulated (King and Wilson, 1975). Hence, a true comprehension of how phenotypic evolution proceeds requires an understanding of the evolution of the mechanisms by which gene expression is regulated.

1.1 How is gene expression regulated?

The regulation of gene expression ensures when, where and at which level genes are expressed. This complex process involves transcriptional, post-transcriptional and translational regulations. The first key regulators to be identified were transcription factors by Jacob and Monod in 1961. Transcription factors are DNA-binding proteins targeting specific cis-regulatory elements (short DNA motifs located in the promoter or upstream region of target genes). These elements are either enhancers or silencers, depending on whether they activate or inhibit gene expression (Boeva, 2016). While transcription factors have been the focus of considerable interest (reviewed in Rebeiz et al., 2015; Romani & Moreno 2020), they are only one aspect of a broader regulatory landscape. In particular, under the so-called “central dogma” of molecular biology, the information contained in the DNA sequence of genes is passed on to RNA molecules by transcription, which are then translated into proteins (Crick, 1970), such that RNAs are mostly considered transient carriers of genetic information, with no role beyond that of encoding the primary sequence of proteins. Yet, it is clear that a variety of RNA molecules present in the cell are not encoding protein sequences (non-coding RNAs, ncRNAs) but have important cellular functions (Hangauer et al., 2013).

Those can be divided into housekeeping and regulatory ncRNAs. Housekeeping ncRNAs include essential ncRNA such as ribosomal RNAs (rRNAs), which are key components of ribosomes, the transfer RNAs (tRNAs) delivering the correct amino acids to the ribosome in accordance with the mRNA sequence, the small nuclear RNAs (snRNAs) part of the spliceosome and involved in the process of introns splicing, and the small nucleolar RNAs (snoRNAs), which are mainly involved in the chemical modification of rRNAs and their processing in the nucleus (Matera et al., 2007). These ncRNAs are expressed constitutively and are fundamental to cell function. However, the discovery of the first microRNA (miRNA), *lin-4,* by Ambros and Ruvkun groups in *Caenorhabditis elegans* highlighted the role of ncRNAs in regulating gene expression (Lee et al., 1993; Wightman et al., 1993). Since then, subsequent research revealed a variety of endogenous sRNAs (20-30 nucleotides) with essential roles in development and responses to biotic and/or abiotic stresses. Their role is typically to guide effector proteins to specific loci, leading either to post-transcriptional gene silencing (PTGS) by transcript cleavage and degradation or translation inhibition, or to transcriptional gene silencing (TGS) by regulating DNA or histone modifications (Carthew and Sontheimer, 2009). sRNAs are categorized according to their biogenesis and mode of action into microRNAs, small interfering RNAs (siRNAs) and PIWI-interacting RNAs (piRNAs), the latter found primarily in animals (Carthew and Sontheimer, 2009) (Figure 1a). Both miRNAs and siRNAs precursors are processed by endoribonuclease proteins DROSHA and DICER (in animals) or DICER-LIKE (DCL) (in plants) and then cleavage products are loaded into ARGONAUTE (AGO) proteins to target gene silencing. On the contrary, piRNA biogenesis is DICER-independent and piRNAs interact with PIWI proteins to regulate gene expression (Chen and Rechavi, 2022) (Figure 1a). More precisely, miRNAs are around 20-22 nucleotides (nt) long, originate from short hairpin structures and primarily act at the PTGS level by binding to mRNAs to negatively regulate their expression. In contrast, siRNAs are around 22-24 nt long, originate from double-stranded RNA and generally act at the TGS level (Chen and Rechavi, 2022). In plants, siRNAs are further divided into subclasses such as secondary siRNAs for which the production is triggered by other miRNAs or siRNAs, and heterochromatic siRNAs (hc-siRNA) which derived from heterochromatic regions of the genome such as transposons, retrotransposons or repetitive DNA elements. Secondary siRNAs and hc-siRNA function in transposons and gene silencing (reviewed in Zhan and Meyers, 2023) and they are described in some detail below. Transfer RNA fragments have recently emerged as a new class of sRNAs regulating gene expression in plants and animals. Recent studies demonstrated that mature tRNAs can be cleaved by endonuclease proteins to produce tRNA-derived RNAs (tDRs) approximately 13-30 nt long. Specific tDRs were found associated with AGO proteins suggesting that their implication in

the regulation of gene expression is similar to the miRNA pathway (Sun et al., 2021; Chen et al., 2021) (Figure 1a).

More recently, long non-coding RNA (lncRNA), ncRNAs longer than 200 nt, have attracted attention due to their high number in plants and animals genomes (Hangauer et al., 2013) (Figure 1b). Similar to mRNAs, most lncRNA are transcribed by the DNA-dependent RNA polymerase II and originate from intergenic regions, intronic regions and sense or antisense transcripts (reviewed in Mattick et al., 2023). Initially, the importance of lncRNAs was questioned due to their low expression levels, leading some to regard them as mere transcriptional noise (Doolittle et al., 2013; Gloss et al., 2016). One study analyzed the sequence diversity levels of a large number of mouse lncRNAs and showed that, although the majority of them evolve under neutral selective constraints, the most conserved are subject to strong selective constraints, suggesting that they play important functions (Wiberg et al., 2015). Among those functions, lncRNAs have roles in the nucleus, in nuclear organization, chromatin regulation and transcription regulation (regulating directly neighboring loci and/or generating a chromatin state influencing the expression of nearby genes). In the cytoplasm, functions of lncRNAs include post-transcriptional regulation via mRNA splicing, mRNA decay, mRNA translation regulation, or via sponging miRNAs (reviewed in Statello et al., 2021). Finally, the circular RNAs (circRNAs) are another type of ncRNAs in plants and animals. Unlike siRNAs and miRNAs, which are formed by linear splicing, circRNAs are formed by covalently connecting the downstream 3′ to the upstream 5′ site from coding or non-coding region (Zhang and Dai, 2022) (Figure 1c). circRNAs can act as miRNA sponge to prevent them from suppressing their target mRNAs which seem to be the most common function in animals. Moreover, they can regulate transcription, mRNA splicing or affect protein function by binding them (Zhang and Dai, 2022).

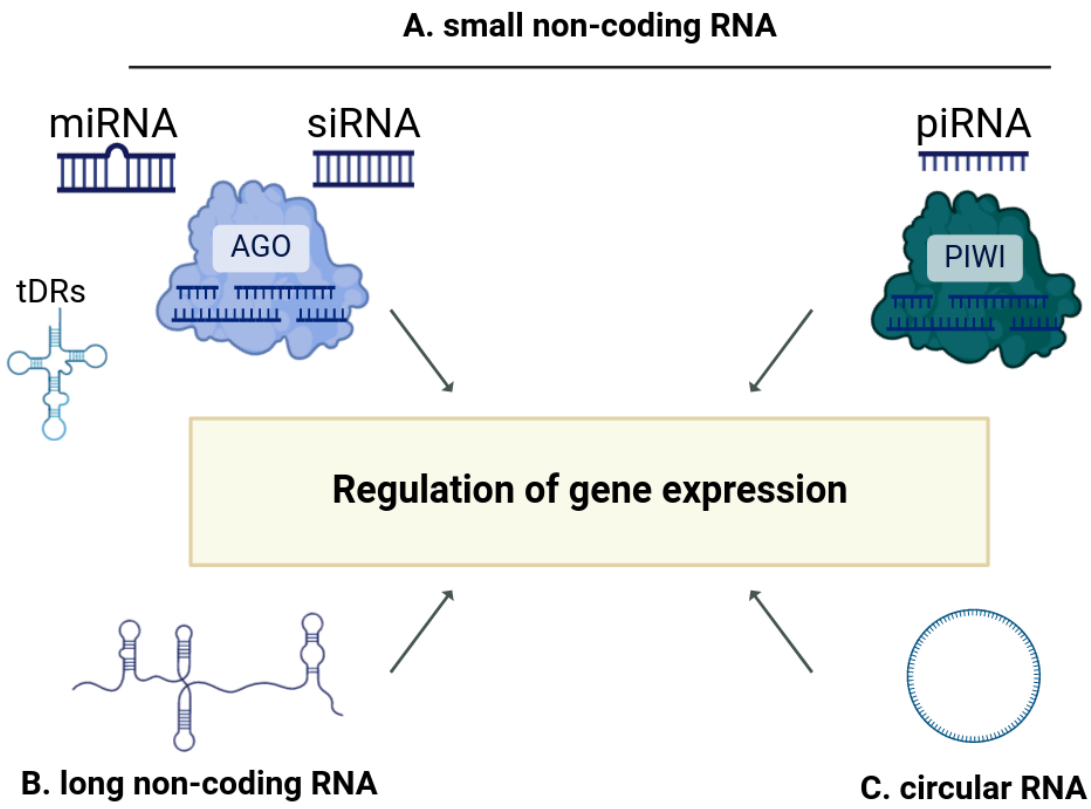


Figure 1: Regulation of gene expression by non-coding RNAs. Non-coding RNAs are major players in the regulation of gene expression. Those can be classified into three main categories: (a) small non-coding RNAs (<200bp) such as miRNAs, siRNAs and piRNAs (only found in animals) are implied in post-transcriptional gene silencing and transcriptional gene silencing. While miRNAs and siRNAs interact with AGO proteins, piRNAs interact with PIWI proteins to regulate gene expression. Recent findings suggest that tRNA-derived RNAs are also loaded into AGO proteins and have regulatory roles; (b) long non-coding RNAs (>200bp) have a wide range of functions, including epigenetic regulation and transcriptional control; (c) circular RNAs regulatory RNAs act mainly as miRNA sponges.

1.2 Challenges in the identification of gene expression regulatory elements

Despite the development of numerous methods, identifying interactions between transcription factors and DNA remains a challenge due to the proteic nature of transcription factors (Lambert et al., 2018). To date, 1,600 transcription factors have been identified in the human genome with three-quarters of them having a DNA-binding motif (Lambert et al., 2018). Most of these transcription factors have been identified by sequence homology to a previously characterized DNA-binding domain through experimental methods such as one-hybrid assays or DNA affinity purification-mass spectrometry, while *de novo*

identification of transcription factors required the use of methods as protein binding microarrays which are difficult to handle than their cognate DNA arrays because they required that proteins maintain the secondary and tertiary structure (Lambert et al., 2018) (Figure 2a,b). On the other hand, the binding site of transcription factors is typically represented as a position weight matrix (PWM), usually displayed as a sequence logo, providing the probability of occurrence of each nucleotide of a DNA motif. However, this representation does not take into account factors such as DNA structure which is essential in transcription factors binding (Sielemann et al., 2021). The most widely used experimental method to reveal these motifs is chromatin immunoprecipitation sequencing (Chip-seq). In this method, transcription factors are cross-linked to DNA, then the DNA is fragmented, the DNA fragments bound to transcription factors are immunoprecipitated using specific antibodies (ChIP) and subjected to sequencing. Isolated DNA fragments are sequenced to identify transcription factor binding sites in the genome. However, the use of antibodies can make the study more difficult when they are not readily available (Lambert et al., 2018). A final point of difficulty in identifying transcription factor-DNA interactions is that several other factors are involved, such as cooperative binding of other transcription factors or interaction between transcription factors and nucleosomes (Morgunova et al., 2017).

The identification of miRNAs and their target genes also come with their own set of challenges, but they are different. In comparison to transcription factors more miRNA-target interactions have been identified in humans with 2,300 mature miRNAs targeting between 65 to 144 genes each (Alles et al., 2019). Initially, Northern blot was employed to identify miRNAs. This method could detect the accumulation of a specific RNA but could not distinguish between miRNAs, siRNAs and partially degraded RNA fragments. Later, the advent of next-generation sequencing (NGS) and bioinformatics tools has dramatically increased the number of annotated miRNAs (Kozomara et al., 2019). The fundamental characteristic of miRNAs is the precise excision of a miRNA-miRNA* duplex from the stem of a single-stranded loop precursor. However, differentiating miRNAs from other sRNAs like siRNAs has been challenging, leading to false positives in databases such as miRBase (Axtell and Meyers, 2018). Nevertheless, periodic refinement of annotation criteria, has enhanced the reliability of miRNA identification (Ambros et al., 2003; Meyers et al., 2008; Axtell and Meyers, 2018). These criteria included requirements such as hairpin precursor stability, validation of miRNA expression by small RNA sequencing, accuracy of precursor processing, *i.e.* the precision with which the miRNA-miRNA* duplex is cut by DCL proteins, stability of the miRNA-miRNA* duplex (Axtell and Meyers, 2018). One possibility to validate experimentally miRNA predictions is the use of methods such as immunoprecipitation of the

AGO proteins. Briefly, AGO proteins are immunoprecipitated using specific antibodies and small RNA fragments associated are sequenced to identify miRNA.

As compared to the complex DNA-binding motifs of transcription factors, the prediction of miRNA target sites is relatively easier, as it relies on the quantification of simple nucleotide sequence complementarity, a procedure for which many bioinformatic tools have been developed (Figure 2c,d). In plants, experimentally verified miRNA-interactions led to a consensus on the ‘rules’ of base-pairing for a functional plant miRNA-target interaction. In this case, binding between plant miRNAs and targets is based on almost complete sequence complementarity, whereas in animals, target recognition is driven by the seed region, which is at most 2-8 nt long (Wang et al 2015; Bartel et al., 2009). However, the predictions of miRNA targets by simple sequence similarity with perfect or near-perfect complementarity is not optimal because experimental data showed that mismatches, G-U wobbles, and bulges have much stronger effects on targeting efficacy at some positions than others. For example, in *Nicotiana benthamiana*, Liu et al. (2014) assessed the impact of specific mismatches between the miRNA and its target on the latter's expression. They showed that mismatches located at the 5' end of the miRNA-target interaction have more deleterious effects on targeting efficiency than mismatches located at the 3' end. On the other hand, Burghgraeve et al. (2020) showed in *A. halleri* that the interaction between sRNA produced by the S-locus, *i.e.* the locus involved in self-incompatibility in Brassicaceae, and transcripts of the SCR gene, *i.e.* the pollen determinant of self-incompatibility, follows a threshold model, according to which sequence complementarity above a certain level (around three mismatches over the 21 to 24 nucleotides of the sRNA molecule and its target) leads to efficient transcriptional silencing of SCR, whereas complementarity below this threshold does not. The fundamental aspects of miRNA-target interactions in plants are well established, facilitating identification by computational methods. However, predictions are essentially sequence alignments, and knowing that certain elements beyond the base-pairing model play a role in miRNA-targeting efficiency, bioinformatics methods are not sufficient and experimental validation is required (Axtell and Meyers, 2018). Among the experimental methods, the simplest is the profiling of miRNA expression and its target by quantitative polymerase chain reaction (qPCR), however this method does not allow to differentiate interactions of different miRNAs targeting the same gene. Another method consists in incorporating the target site in a reporter gene such as luciferase or green fluorescent protein (GFP) and measuring the expression of the reporter gene relative to miRNA expression, however this method requires transgenic lines that are not easily obtained for some species. A last powerful method, allowing broad-scale validation of miRNAs target is HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation). This technique involves the cross linking of AGO proteins to miRNA

targets followed by immunoprecipitation of AGO proteins and sequencing of RNAs associated. Thus, broad-scale validation of the targeting effect of miRNAs remains challenging (Devi et al., 2018).

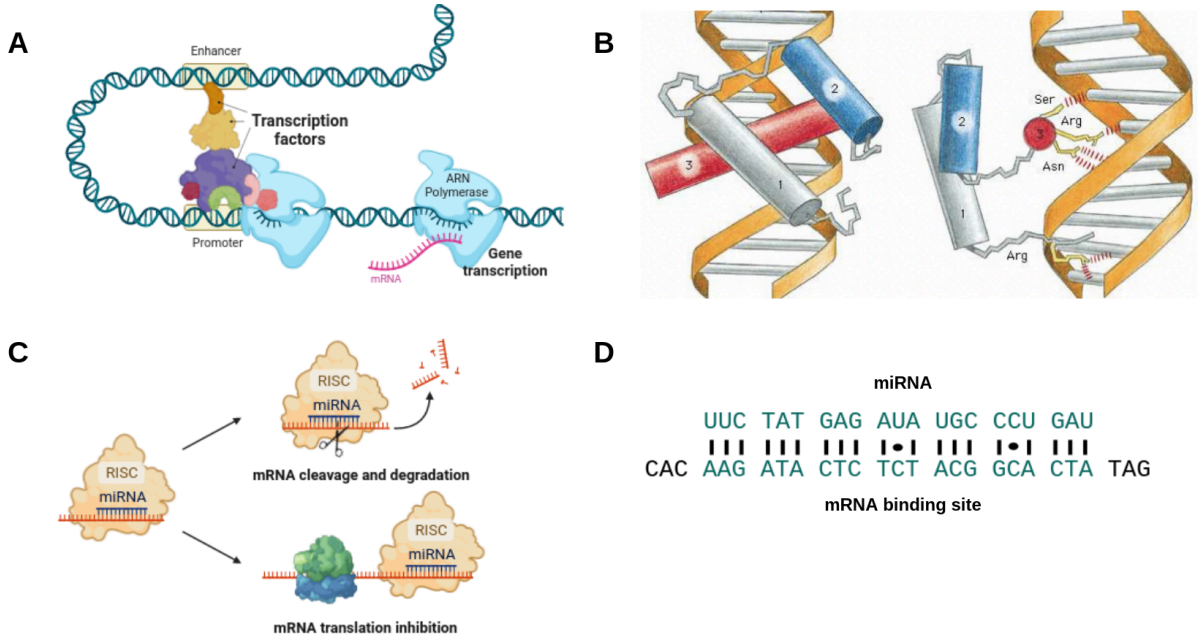


Figure 2: miRNA-target interactions are easier to identify than transcription factors-DNA interactions. (a) Protein nature of transcription factors make the DNA binding site identification difficult. (b) Example of transcription factor binding to DNA motif (Manolis Kellis, book chapter). (c) On the contrary, miRNAs target mRNA by simple sequence complementarity facilitating the identification of miRNA binding sites. (d) Example of hypothetical miRNA binding site.

2. sRNA biosynthesis and mode of action in plants

2.1 Canonical miRNA biosynthesis

In plants, most genes encoding miRNAs are located in intergenic regions of the genome with a minority in intronic regions (Rajagopalan et al., 2006). These miRNA genes are transcribed by the RNA polymerase II into 5' capped and 3' polyadenylated primary transcripts (pri-miRNAs) (Xie et al., 2005) (Figure 3). Concurrently to the transcription of the miRNA gene occurs the processing of the pri-miRNA (Fang et al., 2015; Cambiagno et al., 2021). Pri-miRNAs have a hairpin-like structure that is recognized by the nuclear endoribonuclease DCL which form the D-bodies by associating with other proteins and cleaves the pri-miRNA twice to release the miRNA/miRNA duplex, *i.e.* the mature miRNA and its complementary strand the miRNA*, with a two-nucleotides overhang in 3' and 5' (Figure 3). The *A. thaliana*

genome contains four DCL paralogs (Mukherjee et al., 2013): DCL1 (the canonical DCL involved in miRNA biogenesis) and DCL4 generate duplexes of 21 nt, DCL2 of 22 nt and DCL3 of 24 nt (Roger and Chen 2013). The secondary structure of pri-miRNAs is crucial in determining the specific sites where DCL1 cuts. Some pri-miRNAs are cleaved from “base-to-loop” due to an imperfect lower stem of 15 to 17 nt located between specific bulges and the miRNA-miRNA∗ duplex (Figure 4). On the contrary, some pri-miRNAs have a long upper stem located between the loop and the miRNA-miRNA* duplex, leading to “loop-to-base” processing (Bologna et al., 2013) (Figure 4). After the miRNA-miRNA* duplex is released, the 3′ ends of miRNA and miRNA* are methylated by the nuclear 2′-O-methyltransferase HUA ENHANCER 1 (HEN1) to stabilize the duplex (Figure 3). The miRNA* strand is usually degraded, while the miRNA strand is loaded in AGO1 to form an RNA-induced silencing complex (RISC) in association with other proteins (Figure 3). The *A. thaliana* genome encodes ten AGO genes (AGO1-10) (Vaucheret et al., 2008). AGO proteins sort miRNAs according to their size: AGO1/5/10 and AGO2/3/7 clades load mainly 21 and 22 nt sRNAs, while AGO4/6/9 clades load 24 nt sRNAs (Mi et al., 2008). The 5’ nucleotide is also determinant for sRNA sorting. AGO1 preferentially binds sRNAs with a 5′ uridine, while AGOs2-4-6-7-9 prefer a 5′ adenosine and AGO5 prefers a 5′ cytosine (Mi et al., 2008). However, other factors may influence AGO sorting such as base pairing at position 15 of the miRNA duplex, with AGO1 favoring duplexes with a central mismatch, while AGO2 favors duplexes without a mismatch (Zhang et al., 2014). miRNA loading onto AGO1 occurs in the nucleus, as AGO1 is capable of shuttle between the nucleus and the cytoplasm due to the presence of nuclear localization signal and a nuclear export signal in the AGO1 protein sequence (Bologna et al., 2013) (Figure 3). Once, the RISC complex is transported into the cytoplasm, the mRNA target is recognised through near-complete sequence complementarity with the miRNA, leading to negative regulation through mRNA cleavage and degradation or translation inhibition (Mallory et al., 2010) (Figure 3). Cleavage of the targeted transcript can be performed by AGO1, AGO2, AGO4 or AGO7 thanks to the slicing activity of the PIWI (P-element induced wimpy testis) domain present in the proteins (Carbonell et al., 2012; Qi et al., 2006). In translational repression when AGO1-RISC complex binds the 5′ untranslated region (UTR) of a transcript, the translational initiation is sterically blocked, whereas targeting the open reading frame (ORF) sterically block the translational elongation (Iwakawa and Tomari, 2013). Finally, in rare cases miRNAs direct DNA methylation of their target. For example, in rice, 24-nt miRNAs generated by DCL3 are loaded into AGO4 to direct DNA methylation (Wu et al., 2010).

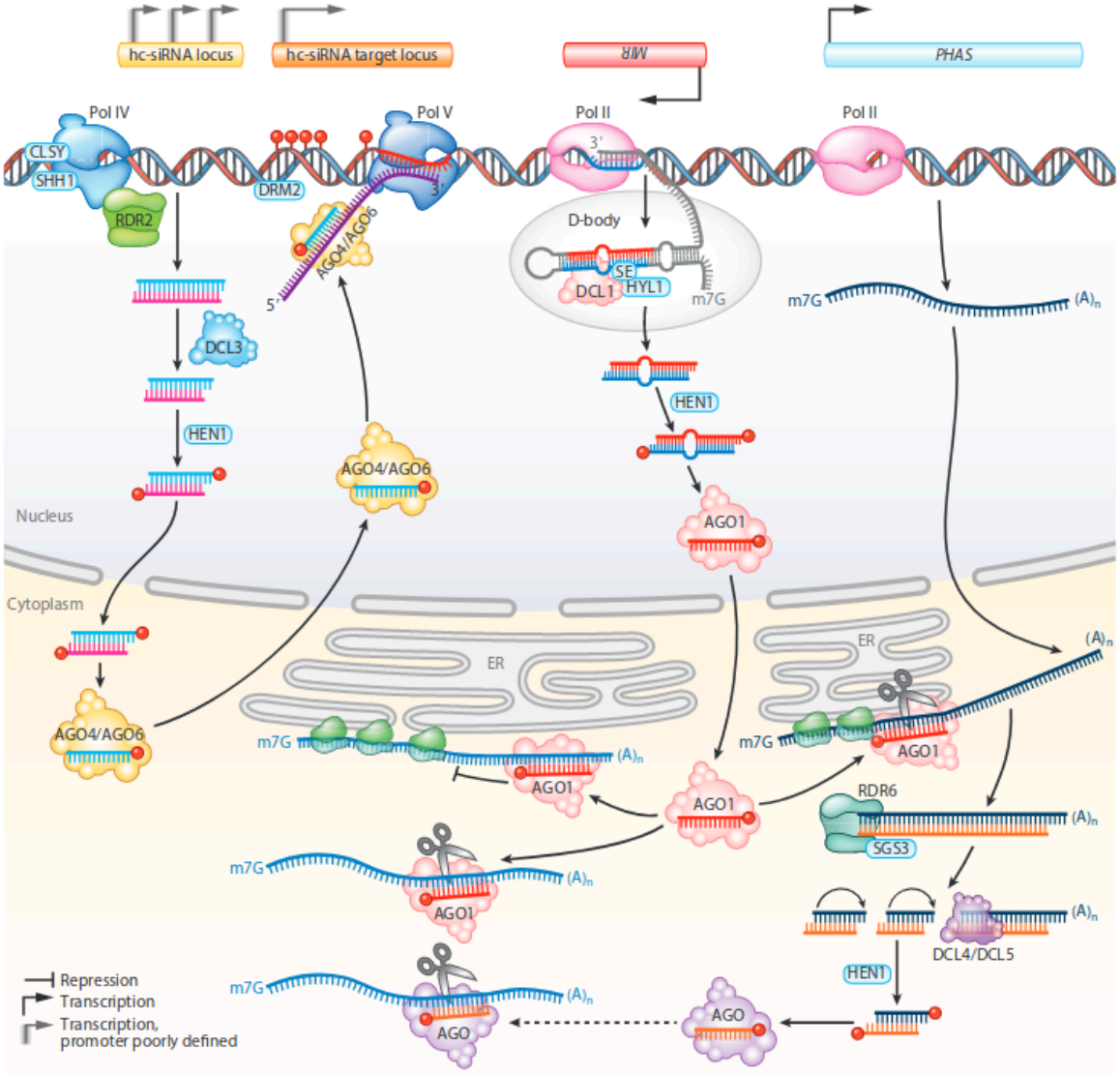


Figure 3: The biosynthetic pathways of miRNAs and siRNAs in plants (Zhan and Meyers, 2023). miRNA genes are transcribed by DNA-dependent RNA polymerase II (Pol II) into pri-miRNAs with a hairpin-like structure. The pri-miRNA is recognized by the nuclear endoribonuclease DCL1 which cleaves it twice to release the miRNA/miRNA duplex. The miRNA is then loaded onto AGO1 in the nucleus. The mRNA target is recognised in the cytoplasm, leading to negative regulation through mRNA cleavage and degradation or translation inhibition. Some miRNAs can cleave the transcript from PHAS loci of which trigger specific fragments are converted to dsRNA by RNA-dependent RNA polymerase 6 (RDR6). Those dsRNA are then diced by DCL4 or DCL5 proteins to generate phasiRNA. hc-siRNAs derived from heterochromatic regions of the genome are transcribed by the plant-specific Pol IV directly followed by the synthesis of a complementary hc-siRNA precursor strand by RDR2. Double strand hc-siRNA precursors are diced by DCL3 into hc-siRNA which are loaded onto AGO4, AGO6 or AGO9. The complex then initiates de novo DNA methylation at the loci targeted by hc-siRNA. Those loci are transcribed by the plant-specific Pol V.

In animals, miRNA biosynthesis is very similar except for a few differences: about half of miRNA genes are located in clusters, often composed of different mature miRNAs; the

nuclear proteins Drosha and DGCR8 cut the pri-miRNA a first time, then the cytoplasmic protein Dicer cuts a second time to release the miR/miR* duplex; the interaction between miRNA and target occurs over a sequence of 2 to 8 nt, whereas in plants complementarity occurs over almost the entire sequence of the mature miRNA (Axtell et al., 2013; Roger and Chen, 2013). Finally, unlike plants, the majority of animal AGOs do not induce cleavage of the miRNA target, but rather induce translational repression of targets by blocking translational initiation or elongation (Bartel et al., 2009).

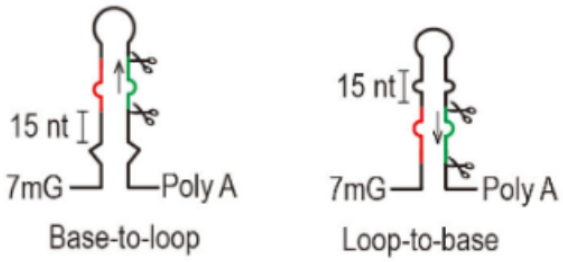


Figure 4: Two main types of pri-miRNA processing by DCL. The localization of an imperfect region in the stem of some pri-miRNAs influenced their processing by DCL proteins from "base-to-loop" (region located between specific bulges and the miRNA-miRNA∗ duplex) or from "loop-to-base" (region located in upper stem, between the loop and the miRNA-miRNA∗ duplex). (Adapted from Jodder, 2021).

2.2 siRNA biosynthesis

siRNAs are also small RNA molecules and have features that resemble miRNAs, but they differ from them in several ways. They can be subdivided into two classes: heterochromatic small interfering RNAs (hc-siRNAs) and secondary siRNAs.

hc-siRNAs derived from heterochromatic regions of the genome such as transposons, retrotransposons or repetitive DNA elements and function in RNA-directed DNA methylation (RdDM). They are transcribed by the plant-specific DNA-dependent RNA polymerase IV (Pol IV) directly followed by the synthesis of a complementary hc-siRNA precursor strand by the RNA-dependent RNA polymerase 2 **(**RDR2). Double strand hc-siRNA precursors are diced by DCL3 into hc-siRNA duplexes which are 24 nt long (Herr et al., 2015), and are methylated by HEN1 before their export into cytoplasm where they are loaded onto AGO4, AGO6 or AGO9 to form RISCs (Havecker et al., 2010). The complex then initiates de novo DNA methylation at the loci targeted by hc-siRNA. Those loci are transcribed by the plant-specific DNA-dependent RNA polymerase V (Pol V) and are mainly transposons (Matzke and Mosher, 2014).

Secondary siRNAs can be further subdivided into phased siRNA (phasiRNA), trans-acting siRNA (tasiRNA) and epigenetically activated siRNA (easiRNA) according to their loci of origin (PHAS loci, TAS genes and active retrotransposons). PhasiRNAs are products of processive cleavage of dsRNAs in a regular way from a well-defined region which is typically defined by AGO cleavage of a single-stranded precursor directed by miRNA or siRNA. After cleavage, the precursor is converted to dsRNA by RNA-DEPENDENT RNA POLYMERASE 6 (RDR6). For example, an asymmetric bulge present in the miRNA-miRNA* duplex will generate a 22/21nt duplex after cleavage by DCL1 (Chen et al., 2010). The 22-nt miRNA loaded in AGO1 can cleave the transcript from PHAS loci of which trigger specific fragments are converted to dsRNA by RDR6. Those dsRNA are then diced by DCL4 or DCL5 proteins to generate a 21 or 24 nt phasiRNA (Liu et al., 2020). Those siRNAs are called phasiRNA due to DCL proteins that cut in a sequential way with a specific phased pattern when reads are mapped.

2.3 Biological functions of miRNAs

In plants, the critical role of sRNAs in gene regulation has become evident. Indeed, experimental studies that analyzed mutants of the main component of miRNA pathway showed that the individuals were strongly affected. For example, mutations that completely abolish the expression of DCL1 are embryo-lethal, and mutants with a weak expression of DCL1 show severe developmental defects such as later flowering, reduced fertility and pleiotropic developmental effects (Schauer et al., 2002). Similarly, null mutant *hen1, i.e* complete lack of the gene product, exhibit strong developmental effects as dwarfism, late flowering, short siliques and impaired photomorphogenesis (Chen et al., 2002; Tsai et al., 2014). Mutations in AGO1 have also pleiotropic effects leading to a complete different plant architecture, *e.g.* rosette leaves lack leaf blade, cauline leaves were filamentous and flowers were infertile and filamentous (Bohmert et al., 1998). This phenotype was very similar to the *Argonauta argo* octopus species which inspired the name of AGO proteins.

By targeting key transcription factors, they occupy a central position in governing plant development. sRNAs play a crucial role in controlling key regulators of meristem identity, leaf polarity and flowering (Ario et al., 2017). For example, in *A. thaliana* the interaction between miR156 and the TF SQUAMOSA PROMOTER BINDING PROTEIN LIKE (SPL) is involved in the transition from juvenile to adult plant, the interaction between miR396 and GROWTH REGULATING FACTOR (GRF) controls leaf morphogenesis, and that between miR169 and NY-FA regulates root development (Samad et al., 2017). Not only do miRNAs act as master

regulators of plant development, they are also involved in regulating the abiotic stress response triggered by various environmental stimuli, such as heat and cold, drought and nutrient deficiency (Song et al., 2019).

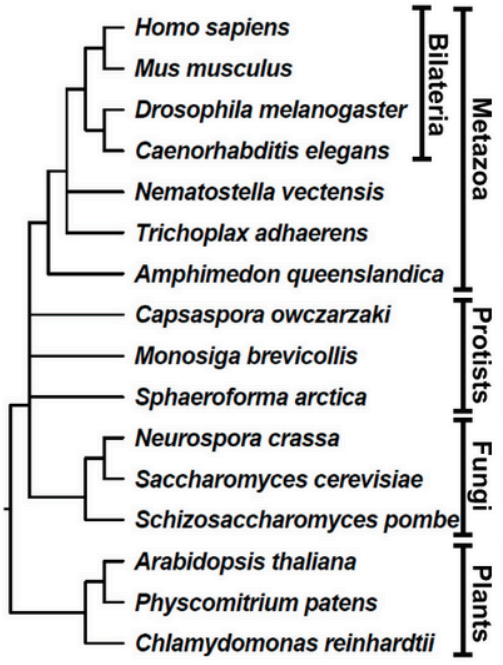
3. Evolution of the miRNA pathway

3.1 Origin of the pathway

Although the sRNAs biosynthesis and mode of action share some similarity between plants and animals, several notable differences are observed. This initially led to the hypothesis that miRNA pathways in plants and animals evolved independently, and that the modern miRNA repertoires would result from convergent evolution (Axtell et al., 2011; Tarver et al., 2012). One argument in support of this hypothesis was that, although DICER proteins are widespread in eukaryotes, plants do not encode a Drosha homolog, but process primary miRNA transcripts only via DCL proteins (Figure 5). However, recent evidence raises the alternative possibility that the miRNA pathway might have already existed in the last common ancestor of eukaryotes (Moran et al., 2017). For example, in the alga *Chlamydomonas reinhardtii*, DCL3 has Drosha-like features (possesses a proline-rich domain and lacks a PAZ domain) that are absent in higher plants, suggesting parallel evolution in different lineages, but a single origin of DCL genes in animals and plants (Valli et al., 2016). On the other hand, the hypothesis of an independent origin of miRNA pathway relied on the fact that the closest unicellular relatives to animals, the choanoflagellates lack Drosha and Pasha genes (Grimson et al., 2008). However, only one unicellular has been studied, and the absence of Drosha and Pasha in choanoflagellates could reflect the loss of an ancient pathway prior to the animal-choanoflagellate divergence (Bartel et al., 2018). This is supported by Brate et al., (2018) who investigated various unicellular sister lineages of animals and showed that Drosha and Pasha originated before the last common ancestor of metazoans.

The second key players of the miRNA machinery, the AGO proteins, are conserved from archaea and bacteria to eukaryotes (Swarts et al., 2014) (Figure 5). However, the differences in the mode of actions observed between plants and animals indicates that AGO proteins evolved independently in the different lineages. For example, in plants, AGO proteins can cleave mRNA targets through a nearly-complete sequence complementarity

between the miRNA and its target (Mallory et al., 2010). This mode of action is similar to the one in prokaryotes where AGO proteins usually cleave the targeted foreign DNA, suggesting that the target cleavage is an ancient mode of action (Moran et al., 2017). On the contrary, the majority of animal AGOs do not induce miRNA target cleavage but induce the destabilization and translational inhibition of targets through seed matching, indicating that this mode is a derived state (Moran et al., 2017). During plant evolution the AGO family expanded with various duplications and losses, suggesting a functional diversification of AGO proteins (Zhang et al., 2015). Based on phylogenetic analyses in land plants, AGO genes seem to form four major clades: AGO1/5/10, AGO2/3/7, AGO4/6/9 and AGO-like clades (Vaucheret et al., 2008; Wang et al., 2021) (Figure 5). The AGO like clade is present in lycophytes, bryophytes and ferns but absent in angiosperms indicating that the functional diversification of AGO proteins occurred early in land plants followed by parallel expansion of the AGO family in angiosperms (You et al., 2017).



AGO/ Piwi	DGCR8/ Pasha	DICER	DROSHA	miRbase V 22.1
8	1	1	1	1917
7	1	1	1	1234
5	1	2	1	258
8	1	1	1	253
6	1	2	1	141
1	0	5	1	/
5	1	4	1	8
0	0	0	0	/
0	0	0	0	/
3	0	0	0	/
2	0	2	0	/
0	0	0	0	/
1	0	1	0	/
10	0	4	0	326
6	0	5	0	/
2	0	3	0	50

Figure 5: miRNA machinery is widespread (figure adapted from Wang et al., 2023). Number of the major proteins involved in miRNA pathway and number of miRNA annotated in miRbase V22.1 in representative eukaryotic species.

3.2 Evolution of miRNA genes

miRNAs are present in plants (Chávez Montes et al., 2014), animals (Guerra-Assunção et al., 2012), fungi (Johnson et al., 2022) and some viruses (Nanbo et al., 2021). Among

animals, only the ctenophore *Mnemiopsis leidyi* lacks miRNA (Maxwell et al., 2012). To date, there is little evidence for the role of miRNAs in organisms other than bilaterian animals and land plants (Moran et al., 2017). For example, only 8 miRNA genes have been found in the demosponge *Amphimedon queenslandicais* (Grimson et al., 2008) and 141 have been identified in the cnidarian *Nematostella vectensis* (Grimson et al., 2008; Moran et al., 2013) (Figure 5). The miRNA families are not conserved between plants and animals (Axtell et al., 2011). The only exception is the miR854 family, which is expressed in *A. thaliana*, but also in *Caenorhabditis elegans*, *Mus musculus* and *Homo sapiens*, and targets a similar region of different homologues of the UBP1b gene encoding an RNA-binding protein (Arteaga-Vazquez et al., 2007). However, it is still debated whether the observed conservation results from shared ancestry or from functional convergence from an independent origin (Arteaga-Vazquez et al., 2007). The lack of conservation of miRNA sequences between plants and animals suggests an independent emergence and diversification of miRNA genes (Axtell et al., 2011). However, an alternative explanation for the lack of homology observed in miRNAs is that sequence turnover is so rapid that homology can no longer be observed between contemporary lineages (Moran et al., 2017). This is in line with several studies which have shown a high rate of death and birth in Arabidopsis and in Drosophila species (Fahlgren et al., 2010; Lu et al., 2008).

In particular, a comparison between *A. thaliana* and *A. lyrata*, a closely related species, showed that 33% of miRNA families were not conserved between the two species, and had therefore been gained or lost over the 10 million years since the two species diverged (Fahlgren et al., 2010). However, determining the evolutionary history of miRNAs presents difficulties, and the high rate of miRNA loss observed could be the result of missannotations (Fromm et al., 2015). Firstly, incomplete genomes and/or small RNA sequencing data lead to incomplete miRNAomes for some organisms. Secondly, annotation errors (random sequence or other classes of small RNAs) are widespread. For example, up to 84% of metazoan miRNA annotations in the miRBase database (Kozomara and Griffiths-Jones, 2014) have been suggested to correspond to false positives (Fromm et al., 2015). Tarver et al., (2018) showed in metazoans that apparent losses of miRNA families are mainly an artifact of poorly sampled and annotated microRNAomes. In plants, efforts have been made to compensate for these problems by sampling a large number of species. Chávez montes et al., (2014) examined miRNA data from 34 phylogenetically representative plant species, ranging from green algae to eudicots (Figure 6). The study revealed that some miRNA families are deeply conserved, while others appear to be species-specific. In particular, they distinguish the following groups of conserved miRNA families: 1) miRNA families conserved in all species, such as miR156, miR159, miR167-169, miR319 2) miRNA families conserved in a lineage but absent in some species, such as miR158, miR162, miR395, miR397 3)

families specific to a group of a species such as miR1919, miR4376, miR5300, which are specific to Solanaceae (Chávez montes et al., 2014) (Figure 6). A recent study conducted a comprehensive analysis of miRNA conservation over 81 phylogenetic plant species ranging from chlorophytes to angiosperms. Their analysis was based on miRNAs annotated in the PmiREN database, in which miRNA have been identified with a standardized workflow using a variety of accessible sRNAseq datasets (Guo et al., 2020). In particular, they showed that across all species studied, around 61.2% of miRNA families were species-specific (Guo et al., 2022).

The rapid evolution of these genes means that their detailed study requires the comparison of species that have diverged recently. In Arabidopsis, the closest comparison is between *A. thaliana* and *A. lyrata*, which diverged about 5 million years ago (Fahlgren et al., 2010; Ma et al., 2010). These studies are therefore not optimally suited to study the evolutionary processes of rapidly evolving genetic elements such as miRNAs, and the large divergence between these species does not allow for fine characterisation of the origin and evolutionary processes of miRNAs. On the other hand, the use of high quality genomes and the use of large set sRNA sequencing data is crucial for the reliable annotation of miRNAs in order to comprehensively determine the miRNAs repertory of a species.

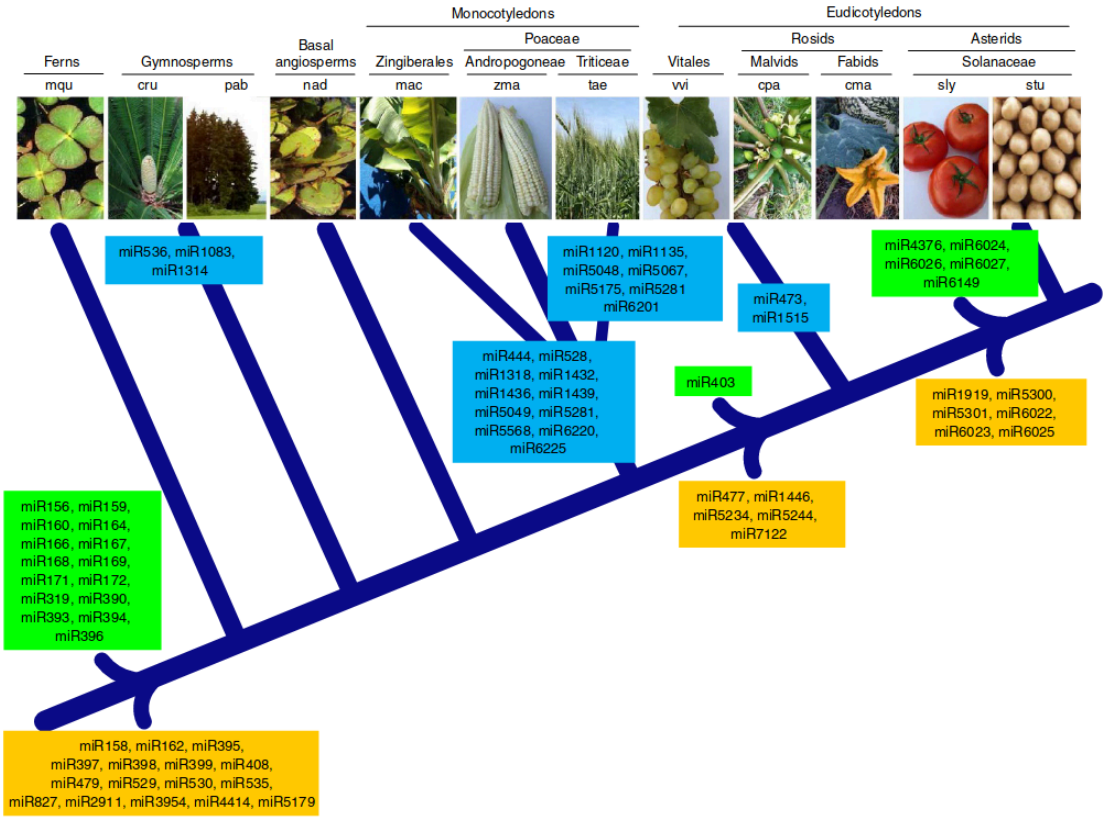


Figure 6: Emergence of miRNA families in terrestrial plants. Families present in all species analyzed are in green. Families in orange are conserved, but are absent in some

species of the group. Families in blue are specific to particular groups of species. (Chávez Montes et al., 2014).

3.3 Evolutionary constraints on miRNA genes

Over the course of evolution of miRNA genes, natural selection acts on genetic variation created by mutations. The strength and type of natural selection can be inferred by analyzing sequence divergence across species and sequence polymorphisms within species. Studies in humans (Quach et al., 2009, Saunders et al., 2007), in Brassicaceae including *A. thaliana* (Ehrenreich et al., 2008; de Meaux et al, 2008; Fahlgren et al., 2010; Ma et al., 2010; Smith et al., 2015), in Drosophila (Lu et al., 2008) and in *Caenorhabditis remanei* (Jovelin et al., 2014) revealed lower divergence and/or polymorphism in miRNA precursors compared to their flanking regions. The level of nucleotide divergence was particularly low in the mature miRNA sequence, indicating strong purifying selection to maintain interaction with targets (Quach et al., 2009, Saunders et al., 2007; Ehrenreich et al., 2008; de Meaux et al., 2008; Fahlgren et al., 2010; Ma et al., 2010; Smith et al., 2015). In humans, polymorphisms within mature miRNA sequences are predominantly localized at the 3' end (*i.e.* outside seed regions) where they would have a limited impact on mRNA targeting (Chen and Rajewsky, 2006; Saunders et al., 2007). Smith et al., (2015) also observed in *A. thaliana* a bias toward higher interspecific divergence of the 3' end relative to the rest of the mature miRNA sequence, but only for the class of most conserved miRNAs. In contrast, the divergence of the evolutionarily young miRNAs was more uniform along their entire length of their sequence. Patterns of polymorphism showed the same bias. The sequence of the terminal loop and stem of miRNA precursors was also more divergent and polymorphic than that of the duplex (Fahlgren et al., 2010; Ma et al., 2010; Jovelin et al., 2014), possibly indicating a lower functional constraint. However, in *A. thaliana,* DCL1 recognizes the loop to cut the precursor stem at a distance of 16-17nt, indicating a possible functional role for the loop in miRNA biogenesis (Zhu et al., 2013). Some plant miRNA precursors also have a conserved 15-17 bp region on the stem close to the miRNA-miRNA* duplex, which could guide DCL1 to direct precursor processing (Chorostecki et al., 2017). Thus, specific regions of the terminal loop or stem could have a role for miRNA biosynthesis. However, the extent to which these parts of the precursor are actually constrained by natural selection remains to be investigated.

Highly conserved miRNA genes generally target genes involved in crucial cellular processes in development and stress responses (Song *et al.,* 2019). In contrast, recently emerged (evolutionarily “young”) miRNA genes more frequently regulate genes related to adaptation to local environments, suggesting that many of these interactions have a minor functional significance leading to weaker constraints through natural selection (Wen *et al.,* 2016 ; Bradley *et al.,* 2017). Studies of young miRNA genes in Drosophila and *C. remanei* have shown that divergence and polymorphism was higher in young miRNAs compared to more conserved ones (Lu et al., 2008; Jovelin et al., 2014). In plants, Fahlgren et al., (2010) and Ma et al., (2010) examined miRNA genes in *A. thaliana* and *A. lyrata*, and observed that deeply conserved miRNA genes exhibit lower sequence divergence, suggesting that they undergo stronger purifying selection than those found exclusively in *A. thaliana* or *A. lyrata*. Analyzing two young miRNA genes (miR824 and miR856) in *A. thaliana*, de Meaux et al., (2008) observed distinct patterns of polymorphism and divergence. miR856 displayed low polymorphism but high divergence, whereas miR824 exhibited high levels of polymorphism and low levels of divergence, suggesting differences in the strength and kind of natural selection acting on them.

Prior studies have examined selectives constraints at the intraspecific level in *A thaliana* (e.g. 66 miRNAs in 23 individuals (Ehrenreich et al., 2008), 16 miRNAs in 40 individuals (Meaux et al., 2008), 327 miRNAs in 80 individuals (Smith et al., 2015)). However, these studies mainly included conserved miRNA genes (Ehrenreich et al., 2008; de Meaux et al., 2008) or only considered miRNA and miRNA* sequences (Smith et al., 2015). Thus, while it is clear that natural selection plays a crucial role in molding the evolution of young miRNA genes, there is still a notable dearth of comprehensive genome-scale investigations focusing on its impact at an intraspecific level.

3.4 The process by which new miRNA genes emerge

The rapid evolution of miRNA genes leads to a high proportion of species-specific miRNA genes. This opens up the possibility to better understand the processes responsible for their origin, as recently emerged genes may be expected to have retained a trace of their original genomic loci. Four hypotheses concerning the origin of miRNA genes have been proposed (Reviewed in Cui et al., 2017; Baldrich et al., 2018) (Figure 7): 1) duplication of an existing miRNA gene that expands the family from which it originated (Maher et al., 2006; Zhao et al., 2015); 2) inverted duplication of a portion of a protein-coding gene, resulting in a stem-loop

structure (Allen et al., 2004; Rajagopalan et al., 2006; Fahlgren et al., 2007, 2010); 3) an origin from a loop sequence derived from a transposable element with inverted repeat ends (Piriyapongsa and Jordan, 2008; Li et al., 2011; Poretti et al., 2019). In particular, miniature inverted repeat transposable elements (MITEs) seem to be a privileged class of transposons that can give rise to new miRNA genes in angiosperm (Guo et al., 2022; Pegler et al., 2023); 4) a *de novo* origin from a region of the genome that previously presented a stem-loop structure and acquired the ability to be transcribed (Felippes et al., 2008). Nozawa et al., assessed the contributions of these different sources of novel miRNA genes in 12 Drosophila species (Nozawa et al., 2010) and 11 plants including species such as *A. thaliana*, *Populus truncatula*, *Glycine max*, *O. sativa* and *Chlamydomonas reinhardtii* (Nozawa et al., 2012). While the genomic hairpin origin seems to be favored in Drosophila species, the contribution of protein-coding genes and TEs appears to be negligible (Nozawa et al., 2010). On the contrary, in plants, protein-coding genes, TEs and the duplication of new miRNA genes seem to participate more or less equally in the production of new miRNA genes (Nozawa et al., 2012) (Figure 7). The differences between plants and animals can be explained by the fact that target recognition in animals occurs via the seed region (6-8 bp) of the mature miRNA, facilitating the possibility of their generation by chance in the hairpin structures present in the genome, whereas in plants, target recognition occurs via the almost entire miRNA sequence, making it more difficult their generation by chance (Nozawa et al., 2012).

Beside the source of miRNA genes, a model for the functional evolution of miRNA genes after they originated was proposed by Allen et al., (2004) (Figure 8a). This model posits that the reverse duplication produces an initially perfect stem-loop structure. When transcribed, it is recognised by a DCL2,3,4 proteins producing multiple duplexes of siRNAs, capable of regulating target genes. Over the course of evolution, the stem-loop structures accumulate mutations, disrupting their complementarity and eventually facilitating the recognition and production of a miR/miR* duplex by DCL1 (Allen et al., 2004; Pegler et al., 2023). However, this verbal model has been rarely tested experimentally.

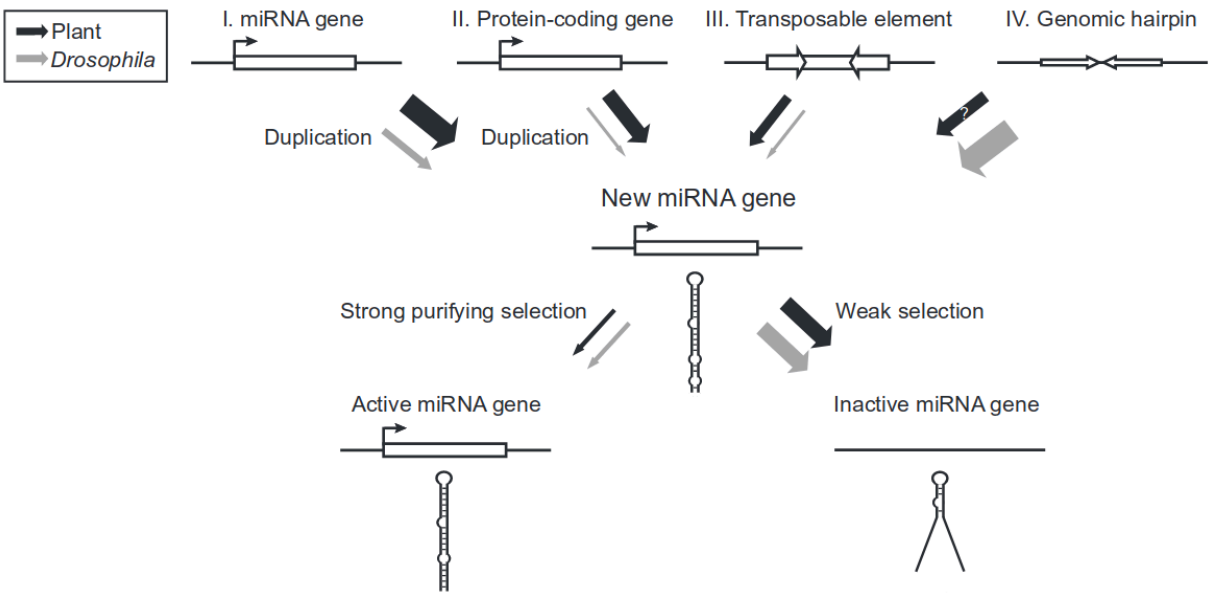


FIGURE 7: Different scenarios for the origin of the miRNA genes and their contributions in Drosophila and Arabidopsis species. miRNA genes can emerge from different genomic sources: 1) from a miRNA gene preexisting; 2) from a protein coding gene; 3) from transposable elements; 4) from genomic hairpin (Nozawa et al., 2012).

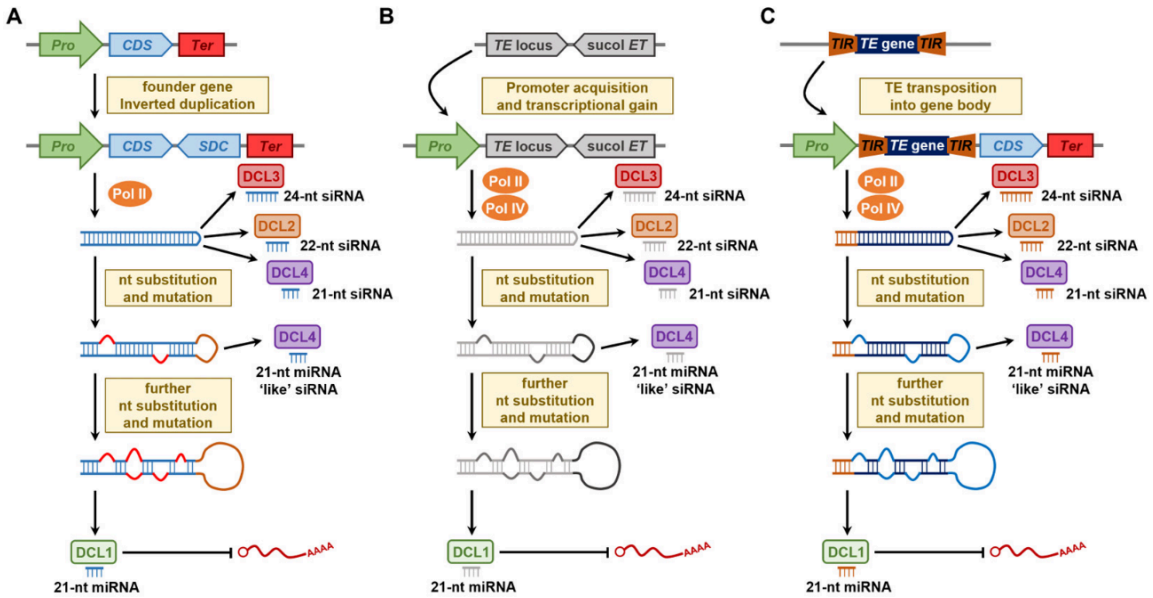


FIGURE 8: Models of miRNA gene emergence. The newly emerged miRNA gene forms a perfect stem-loop that is not recognised by the miRNA biosynthetic machinery and leads to the production of siRNAs. In the course of evolution, the gene will acquire mutations that create mismatches in the stem-loop. After a while, it has accumulated mutations that allow it to be recognised by DCL1 and produces canonical miRNAs that are taken up by the AGO1 protein. (Pegler et al., 2023).

4. Integration of miRNAs in the regulatory network

4.1 Natural selection on miRNA targets

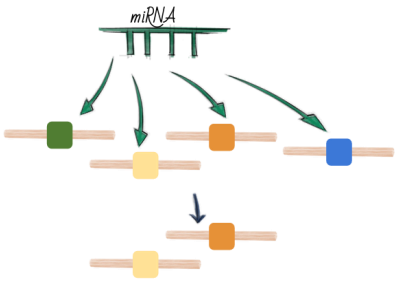
When a new miRNA appears in a genome, it may be more or less constrained by natural selection, depending on the number of targets or their level of essentiality, *i.e.* whether or not a gene is essential for the proper functioning of the cell. Within the targeted mRNA, a mutation occurring at the miRNA binding site can destabilize the interaction and affect the individual. For example, in humans, single nucleotide polymorphism (SNPs) in miRNA seed sequence and target sequences can lead to severe disease such as diabetes or cancer (Chhichholiya et al., 2021). Early population genetics studies on humans showed that miRNA binding sites in 3'UTR were less polymorphic than other conserved motifs in 3'UTR, suggesting strong purifying selection (Chen et Rajewsky, 2006; Saunders et a. 2007). However, more recent study investigated the selective pressures on miRNA binding site in humans, joints to miRNA expression and showed that the selectives constraints on miRNA binding site in targets were stronger when the miRNA that were targeting them were highly expressed, suggesting that the interaction between miRNA and target is a crucial factor determining selection against miRNA target site (Hatlen and Marco, 2020). In *A. thaliana*, the miRNA binding sites are less polymorphic and less divergent than the rest of the mRNA sequence, indicating purifying selection at this region (Ehrenreich and Purugganan, 2008). In addition, Smith et al., (2015) showed in *A. thaliana* that the polymorphism in the miRNA binding site in the target was lower than the polymorphism of mature miRNA sequences, suggesting a stronger constraint on the target site than on the mature miRNA sequence. However, only 52 binding sites of conserved miRNAs in 23 individuals of *A.thaliana* have been studied by Ehrenreich et al., (2008) and Smith et al. (2015) analyzed only targets of miRNA with a size of 21nt. Hence, the effect of selection on the binding site of the genes targeted by recently evolved miRNA genes has been poorly studied in plants.

4.2 Acquisition of new miRNA-targets

The network of miRNAs and their targets can change over the course of evolution by the apparition of mutations that can create or destroy miRNA binding sites. Smith et al., (2015) analyzed the conservation of mRNA targeting by miRNAs in the Camelineae, and showed that while some miRNA-target pairs were highly conserved across species, many mRNAs have become miRNA targets in a single species, suggesting that miRNA-target pairs are evolutionarily transient. When considering the evolution of miRNA networks, it is crucial to understand both the quantity of connections and the intensity of interactions between miRNAs and their target genes. In animals, every miRNA potentially regulates a large number of targets because recognition requires only "seed" sequence. In contrast, plant miRNA-target binding requires a nearly-full complementarity between the miRNA and its target, often resulting in a strong effect on a small number of targets (Liu et al., 2014). In addition, older miRNAs tend to have higher expression levels compared to miRNA that have emerged more recently (Lu et al., 2008; Ma et al., 2010). This suggests that older miRNAs may more effectively suppress their targets, possibly leading to a strengthening of the connection between miRNAs and their target genes over evolutionary time. The number of connections between a miRNA and its targets may also change over evolutionary time, but the way this happens remains controversial. Two models for the acquisition of targets have been proposed. In the so-called "decay model", new miRNAs initially have many targets, most of which are either neutral or deleterious, while only a few are beneficial. Over time, deleterious interactions are removed by natural selection and beneficial interactions are retained. Furthermore, when the miRNA first appears, it tends to have low levels of expression, and the more beneficial or deleterious it is, the faster selection can lead to an increase or decrease of its expression level (Chen et al., 2007; Roux *et al.,* 2012). In the "growth model", proposed in Drosophila, older miRNAs tend to have a greater number of targets than newly acquired miRNAs, suggesting an increase of the average number of miRNA targets during evolution (Nozawa et al., 2016). In this model, new miRNAs initially possess few targets, and they are gradually acquired through time. These changes of the repertoire of targets can proceed either through modification of the miRNA sequence itself, or through mutations of the CDS sequences across the genome that can either create a new target, or abolish an existing target. The existence of a sequence similarity threshold above which a given sequence can act as a *bona fide* target (Schwab et al., 2005 Burghgraeve et al., 2020) suggests that single mutations can create or abolish targets, but a more quantitative model, whereby changes introduced to a given sequence of the CDS can lead to

more or less efficient regulation (Liu et al., 2014) is also possible. The relevance of these two models has been tested in animals, but only to a limited extent in plants.

Negative selection against target sites
(Chen *et al.* 2007 in human and Roux *et al.* 2012 in mice)



Selection for the acquisition of new targets over time
(Nozawa *et al.* 2016 in Drosophila)

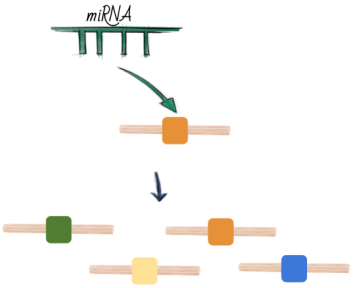


Figure 8: Two models of evolution of miRNA target network. The “decay model” suggests that newly emerged miRNA have a high number of targets, and the deleterious interactions are eliminated over time by natural selection. The “growth model” suggested that in course of evolution new miRNA genes gain targets with beneficial or neutral effects.

5. Objectives and structure of the thesis

While the biosynthetic pathways and mode of action of miRNAs have been widely studied and are now well known, the evolutionary history of miRNA genes remains poorly understood in plants. Several studies showed that the miRNA repertoire contains some deeply conserved miRNAs, but also a majority of more recently emerged ones whose evolutionary significance is still debated. My PhD project aims at improving our understanding of the origin of miRNA genes and the processes by which they become functionally specialized. In order to identify very recently emerged miRNA genes, I focused on the genus Arabidopsis, particularly by studying the two closely related species *A. halleri* and *A. lyrata* (that diverged about 1 Myr ago, Roux et al., 2011).

My PhD thesis is structured in three main chapters :

- In the first chapter, I performed a deep annotation of miRNA genes in a new reference genome of *A. halleri*, and I evaluated the level of conservation of the miRNA genes at increasing levels of phylogenetic divergence. By doing this, I investigated the processes by which proto-miRNAs eventually acquire features of “canonical” miRNA genes over short and long evolutionary times. This chapter is being finalized, and we aim at submitting the article in the coming weeks.

Contribution: I wrote the bioinformatic scripts to annotate the miRNA genes and predict their targets. I contributed to the plant collection for the AGO immunoprecipitations that I perform with the collaboration of Jacinthe Azevedo Favory. I set up the hydroponic culture to obtain root material for the AGO-IP. Then, I wrote the bioinformatic scripts and analyzed the conservation of the miRNA genes, their characteristics and the characteristics of their targets. Finally, I performed the variant calling analysis with the help of Chloé Beaumont, Sophie Gallina and Mathieu Genete to write part of the scripts.

- The second chapter is a preliminary investigation of the mutational origin of the category of the most recent miRNA genes identified in chapter 1. Based on the comparison between their genomic sequences and databases of their putative evolutionary progenitors, I aim at evaluating the relative contribution of these different sources (other miRNA genes, protein-coding genes, transposable elements, non-coding intergenic DNA).

Contribution: I wrote all the bioinformatic scripts. Eléanore Lacoste and Jean-Marc Aury (Génoscope) annotated the transposable elements in the *A. halleri* Auby-1 reference genome.

- The last chapter contains two sections. In the conclusion section, I synthesize my main findings and I evaluate how they change the scientific questions I addressed. In the perspective section, I describe how the results of my thesis open new avenues for research. I present in particular in some details the advancement of a pilot study I initiated, aimed at experimentally evaluating the regulatory potential of the recent miRNA genes identified in chapter 1. It relies on a *A. halleri* x *A. lyrata* backcross population for which we *de novo* assembled the parental genomes and performed sRNA-seq and RNA-seq on parental individuals and a number of backcrosses. The aim is to collectively measure the effect of the *A. halleri*-specific miRNA loci segregating in this backcross population on the level of transcripts of their putative targets. This experiment was initially intended to be included as a full chapter in my PhD thesis, but due to unforeseen circumstances a whole first batch of samples I collected for this chapter was lost during transport, resulting in delays that prevented completion of this chapter. It is thus presented as a perspective stemming from my PhD work.

Contribution: I participated in the plant material collection. I extracted the RNAs (total and small) of the three parents and ten individuals of the backcross population, and constructed the libraries with the help of Christelle Blassiau. I extracted the total RNA of a hundred individuals of the backcross population with the help of Laurence Debacker.

6. References

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.-H., Spatafora, J.W., and Carrington, J.C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat Genet **36**: 1282–1290.

Alles, J., Fehlmann, T., Fischer, U., Backes, C., Galata, V., Minet, M., Hart, M., Abu-Halima, M., Grässer, F.A., Lenhof, H.-P., Keller, A., and Meese, E. (2019). An estimate of the total number of true human miRNAs. Nucleic Acids Research **47**: 3353–3364.

Ambros, V. et al. (2003). A uniform system for microRNA annotation. RNA **9**: 277–279.

Arteaga-Vázquez, M., Caballero-Pérez, J., and Vielle-Calzada, J.-P. (2007). A Family of MicroRNAs Present in Plants and Animals. The Plant Cell **18**: 3355–3369.

Axtell, M.J., Westholm, J.O. & Lai, E.C. (2011) Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol* **12**, 221.

Axtell, M.J. (2013). Classification and Comparison of Small RNAs from Plants. Annu. Rev. Plant Biol. **64**: 137–159.

Axtell, M.J. and Meyers, B.C. (2018). Revisiting Criteria for Plant MicroRNA Annotation in the Era of Big Data. Plant Cell **30**: 272–284.

Baldrich, P., Beric, A., and Meyers, B.C. (2018). Despacito: the slow evolutionary changes in plant microRNAs. Current Opinion in Plant Biology **42**: 16–22.

Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell **136**: 215–233.

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell **173**: 20–51.

Boeva, V. (2016). Analysis of Genomic Sequence Motifs for Deciphering Transcription Factor Binding and Transcriptional Regulation in Eukaryotic Cells. Front. Genet. **7**.

Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C. (1998) AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J. **17**:170-80.

Bologna, N.G., Schapire, A.L., Zhai, J., Chorostecki, U., Boisbouvier, J., Meyers, B.C., and Palatnik, J.F. (2013). Multiple RNA recognition patterns during microRNA biogenesis in plants. Genome Res. **23**: 1675–1689.

Bradley, D. et al. (2017). Evolution of flower color pattern through selection on regulatory small RNAs. Science **358**: 925–928.

Bråte, J., Neumann, R.S., Fromm, B., Haraldsen, A.A.B., Tarver, J.E., Suga, H., Donoghue, P.C.J., Peterson, K.J., Ruiz-Trillo, I., Grini, P.E., and Shalchian-Tabrizi, K. (2018). Unicellular Origin of the Animal MicroRNA Machinery. Current Biology **28**: 3288-3295.e5.

Burghgraeve, N., Simon, S., Barral, S., Fobis-Loisy, I., Holl, A.-C., Ponitzki, C., Schmitt, E., Vekemans, X., and Castric, V. (2020). Base-Pairing Requirements for Small RNA-Mediated Gene Silencing of Recessive Self-Incompatibility Alleles in *Arabidopsis halleri*. Genetics **215**: 653–664.

Cambiagno, D.A., Giudicatti, A.J., Arce, A.L., Gagliardi, D., Li, L., Yuan, W., Lundberg, D.S., Weigel, D., and Manavella, P.A. (2021). HASTY modulates miRNA biogenesis by linking pri-miRNA transcription and processing. Molecular Plant **14**: 426–439.

Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Gilbert, K.B., Montgomery, T.A., Nguyen, T., Cuperus, J.T., and Carrington, J.C. (2012). Functional Analysis of Three *Arabidopsis* ARGONAUTES Using Slicer-Defective Mutants. The Plant Cell **24**: 3613–3629.

Carthew, R.W. and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell **136**: 642–655.

Chávez Montes, R.A., Rosas-Cárdenas, D.F.F., De Paoli, E., Accerbi, M., Rymarquis, L.A., Mahalingam, G., Marsch-Martínez, N., Meyers, B.C., Green, P.J., and De Folter, S. (2014). Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. Nat Commun **5**: 3722.

Chen, X. M., Liu, J., Cheng, Y. L., and Jia, D. X. (2002). HEN1 functions pleiotropically in arabidopsis development and acts in c function in the flower. Development **129**, 10851094.

Chen, K. and Rajewsky, N. (2006). Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet **38**: 1452–1456.

Chen, K. and Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet **8**: 93–103.

Chen, H.-M., Chen, L.-T., Patel, K., Li, Y.-H., Baulcombe, D.C., and Wu, S.-H. (2010). 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc. Natl. Acad. Sci. U.S.A. **107**: 15269–15274.

Chen, Q., Zhang, X., Shi, J., Yan, M., and Zhou, T. (2021). Origins and evolving functionalities of tRNA-derived small RNAs. Trends in Biochemical Sciences **46**: 790–804.

Chen, X. and Rechavi, O. (2022). Plant and animal small RNA communications between cells and organisms. Nat Rev Mol Cell Biol **23**: 185–203.

Chhichholiya, Y., Suryan, A.K., Suman, P., Munshi, A., and Singh, S. (2021). SNPs in miRNAs and Target Sequences: Role in Cancer and Diabetes. Front. Genet. **12**: 793523.

Chorostecki, U., Moro, B., Rojas, A.M.L., Debernardi, J.M., Schapire, A.L., Notredame, C., and Palatnik, J.F. (2017). Evolutionary Footprints Reveal Insights into Plant MicroRNA Biogenesis. Plant Cell **29**: 1248–1261.

Crick, F. (1970). Central Dogma of Molecular Biology.

Cui, J., You, C., and Chen, X. (2017). The evolution of microRNAs in plants. Current Opinion in Plant Biology **35**: 61–67.

D'Ario, M., Griffiths-Jones, S., and Kim, M. (2017). Small RNAs: Big Impact on Plant Development. Trends in Plant Science **22**: 1056–1068.

De Meaux, J., Hu, J.-Y., Tartler, U., and Goebel, U. (2008). Structurally different alleles of the ath- *MIR824* microRNA precursor are maintained at high frequency in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U.S.A. **105**: 8994–8999.

Devi, K., Dey, K.K., Singh, S., Mishra, S.K., Modi, M.K., and Sen, P. (2019). Identification and validation of plant miRNA from NGS data—an experimental approach. Briefings in Functional Genomics **18**: 13–22.

Doolittle, W.F. (2013). Is junk DNA bunk? A critique of ENCODE. Proc. Natl. Acad. Sci. U.S.A. **110**: 5294–5300.

Ehrenreich, I.M. and Purugganan, M.D. (2008). Sequence Variation of MicroRNAs and Their Binding Sites in Arabidopsis. Plant Physiology **146**: 1974–1982.

Fang, W. and Bartel, D.P. (2015). The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. Molecular Cell **60**: 131–145.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., and Carrington, J.C. (2007). High-Throughput Sequencing of Arabidopsis microRNAs: Evidence for Frequent Birth and Death of MIRNA Genes. PLoS ONE **2**: e219.

Fahlgren, N., Jogdeo, S., Kasschau, K.D., Sullivan, C.M., Chapman, E.J., Laubinger, S., Smith, L.M., Dasenko, M., Givan, S.A., Weigel, D., and Carrington, J.C. (2010). MicroRNA Gene Evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. The Plant Cell **22**: 1074–1089.

Fenselau De Felippes, F., Schneeberger, K., Dezulian, T., Huson, D.H., and Weigel, D. (2008). Evolution of *Arabidopsis thaliana* microRNAs from random sequences. RNA **14**: 2455–2459.

Fromm, B., Billipp, T., Peck, L.E., Johansen, M., Tarver, J.E., King, B.L., Newcomb, J.M., Sempere, L.F., Flatmark, K., Hovig, E., and Peterson, K.J. (2015). A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. Annu. Rev. Genet. **49**: 213–242.

Gloss, B.S. and Dinger, M.E. (2016). The specificity of long noncoding RNA expression. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms **1859**: 16–22.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature **455**: 1193–1197.

Guerra-Assunção, J. and Enright, A.J. (2012). Large-scale analysis of microRNA evolution. BMC Genomics **13**: 218.

Guo, Z. et al. (2020). PmiREN: a comprehensive encyclopedia of plant miRNAs. Nucleic Acids Research **48**: D1114–D1121.

Guo, Z., Kuang, Z., Deng, Y., Li, L., and Yang, X. (2022). Identification of Species-Specific MicroRNAs Provides Insights into Dynamic Evolution of MicroRNAs in Plants. IJMS **23**: 14273.

Guo, Z., Kuang Z., Tao Y., Wang H., Wan M., Hao C., Shen F., Yang X., Li L. (2022). Miniature Inverted-repeat Transposable Elements Drive Rapid MicroRNA Diversification in Angiosperms.

Hangauer, M.J., Vaughn, I.W., and McManus, M.T. (2013). Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. PLoS Genet **9**: e1003569.

Hatlen, A. and Marco, A. (2020). Pervasive Selection against MicroRNA Target Sites in Human Populations. Molecular Biology and Evolution **37**: 3399–3408.

Havecker, E.R., Wallbridge, L.M., Hardcastle, T.J., Bush, M.S., Kelly, K.A., Dunn, R.M., Schwach, F., Doonan, J.H., and Baulcombe, D.C. (2010). The *Arabidopsis* RNA-Directed DNA Methylation Argonautes Functionally Diverge Based on Their Expression and Interaction with Target Loci. The Plant Cell **22**: 321–334.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA Polymerase IV Directs Silencing of Endogenous DNA. Science **308**: 118–120.

Iwakawa, H. and Tomari, Y. (2013). Molecular Insights into microRNA-Mediated Translational Repression in Plants. Molecular Cell **52**: 591–601.

Jacob, F. and Monod, J. Genetic regulatory mechanisms in the synthesis of proteins.

Jodder J. Regulation of pri-MIRNA processing: mechanistic insights into the miRNA homeostasis in plant. Plant Cell Rep. 2021 May;**40**:783-798

Johnson, N.R., Larrondo, L.F., Álvarez, J.M., and Vidal, E.A. (2022). Comprehensive re-analysis of hairpin small RNAs in fungi reveals loci with conserved links. eLife **11**: e83691.

Jovelin, R. and Cutter, A.D. (2014). Microevolution of Nematode miRNAs Reveals Diverse Modes of Selection. Genome Biology and Evolution **6**: 3049–3063.

King, M.-C. and Wilson, A.C. (1975). Evolution at Two Levels in Humans and Chimpanzees. Science, New Series **188**: 107–116.

Kozomara, A. and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucl. Acids Res. **42**: D68–D73.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Research **47**: D155–D162.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell **172**: 650–665.

Lee RC, Feinbaum RL, Ambros V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell **75**:843-54.

Li, Y., Li, C., Xia, J., and Jin, Y. (2011). Domestication of Transposable Elements into MicroRNA Genes in Plants. PLoS ONE **6**: e19212.

Liu, J. and Robinson-Rechavi, M. (2020). Robust inference of positive selection on regulatory sequences in the human brain. Sci. Adv. **6**: eabc9863.

Liu, Q., Wang, F., and Axtell, M.J. (2014). Analysis of Complementarity Requirements for Plant MicroRNA Targeting Using a *Nicotiana benthamiana* Quantitative Transient Assay. The Plant Cell **26**: 741–753.

Lu, J., Fu, Y., Kumar, S., Shen, Y., Zeng, K., Xu, A., Carthew, R., and Wu, C.-I. (2008). Adaptive Evolution of Newly Emerged Micro-RNA Genes in Drosophila. Molecular Biology and Evolution **25**: 929–938.

Ma, Z., Coruh, C., and Axtell, M.J. (2010). *Arabidopsis lyrata* Small RNAs: Transient *MIRNA* and Small Interfering RNA Loci within the *Arabidopsis* Genus. Plant Cell **22**: 1090–1103.

Maher, C., Stein, L., and Ware, D. (2006). Evolution of *Arabidopsis* microRNA families through duplication events. Genome Res. **16**: 510–519.

Mallory, A. and Vaucheret, H. (2010). Form, Function, and Regulation of ARGONAUTE Proteins. Plant Cell **22**: 3879–3889.

Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol **8**: 209–220.

Mattick, J.S. et al. (2023). Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol **24**: 430–447.

Matzke, M.A. and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet **15**: 394–408.

Maxwell, E.K., Ryan, J.F., Schnitzler, C.E., Browne, W.E., and Baxevanis, A.D. (2012). MicroRNAs and essential components of the microRNA processing machinery are not encoded in the genome of the ctenophore Mnemiopsis leidyi. BMC Genomics **13**: 714.

Meyers, B.C. et al. (2008). Criteria for Annotation of Plant MicroRNAs. Plant Cell **20**: 3186–3190.

Mi, S. et al. (2008). Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5′ Terminal Nucleotide. Cell **133**: 116–127.

Moran, Y., Praher, D., Fredman, D., and Technau, U. (2013). The Evolution of MicroRNA Pathway Protein Components in Cnidaria. Molecular Biology and Evolution **30**: 2541–2552.

Moran, Y., Agron, M., Praher, D., and Technau, U. (2017). The evolutionary origin of plant and animal microRNAs. Nat Ecol Evol **1**: 0027.

Morgunova, E. and Taipale, J. (2017). Structural perspective of cooperative transcription factor binding. Current Opinion in Structural Biology **47**: 1–8.

Mukherjee, K., Campos, H., and Kolaczkowski, B. (2013). Evolution of Animal and Plant Dicers: Early Parallel Duplications and Recurrent Adaptation of Antiviral RNA Binding in Plants. Molecular Biology and Evolution **30**: 627–641.

Nanbo, A., Furuyama, W., and Lin, Z. (2021). RNA Virus-Encoded miRNAs: Current Insights and Future Challenges. Front. Microbiol. **12**: 679210.

Nozawa, M., Fujimi, M., Iwamoto, C., Onizuka, K., Fukuda, N., Ikeo, K., and Gojobori, T. (2016). Evolutionary Transitions of MicroRNA-Target Pairs. Genome Biol Evol **8**: 1621–1633.

Nozawa, M., Miura, S., and Nei, M. (2010). Origins and Evolution of MicroRNA Genes in Drosophila Species. Genome Biology and Evolution **2**: 180–189.

Nozawa, M., Miura, S., and Nei, M. (2012). Origins and Evolution of MicroRNA Genes in Plant Species. Genome Biology and Evolution **4**: 230–239.

Pegler, J.L., Oultram, J.M.J., Mann, C.W.G., Carroll, B.J., Grof, C.P.L., and Eamens, A.L. (2023). Miniature Inverted-Repeat Transposable Elements: Small DNA Transposons That Have Contributed to Plant MICRORNA Gene Evolution. Plants **12**: 1101.

Piriyapongsa, J. and Jordan, I.K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. RNA **14**: 814–821.

Poretti, M., Praz, C.R., Meile, L., Kälin, C., Schaefer, L.K., Schläfli, M., Widrig, V., Sanchez-Vallet, A., Wicker, T., and Bourras, S. (2020). Domestication of High-Copy Transposons Underlays the Wheat Small RNA Response to an Obligate Pathogen. Molecular Biology and Evolution **37**: 839–848.

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

Quach, H., Barreiro, L.B., Laval, G., Zidane, N., Patin, E., Kidd, K.K., Kidd, J.R., Bouchier, C., Veuille, M., Antoniewski, C., and Quintana-Murci, L. (2009). Signatures of Purifying and Local Positive Selection in Human miRNAs. The American Journal of Human Genetics **84**: 316–327.

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. (2006). A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev. **20**: 3407–3425.

Rebeiz, M., Patel, N.H., and Hinman, V.F. (2015). Unraveling the Tangled Skein: The Evolution of Transcriptional Regulatory Networks in Development. Annu. Rev. Genom. Hum. Genet. **16**: 103–131.

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

Romani, F. and Moreno, J.E. (2021). Molecular mechanisms involved in functional macroevolution of plant transcription factors. New Phytologist **230**: 1345–1353.

Roux, C., Castric, V., Pauwels, M., Wright, S.I., Saumitou-Laprade, P., and Vekemans, X. (2011). Does Speciation between Arabidopsis halleri and Arabidopsis lyrata Coincide with Major Changes in a Molecular Target of Adaptation? PLoS ONE **6**: e26872.

Roux, J., Gonzàlez-Porta, M., and Robinson-Rechavi, M. (2012). Comparative analysis of human and mouse expression data illuminates tissue-specific evolutionary patterns of miRNAs. Nucleic Acids Research **40**: 5890–5900.

Samad, A.F.A., Sajad, M., Nazaruddin, N., Fauzi, I.A., Murad, A.M.A., Zainal, Z., and Ismail, I. (2017). MicroRNA and Transcription Factor: Key Players in Plant Regulatory Network. Front. Plant Sci. **8**.

Saunders, M.A., Liang, H., and Li, W.-H. (2007). Human polymorphism at microRNAs and microRNA target sites. Proc. Natl. Acad. Sci. U.S.A. **104**: 3300–3305.

Schauer, S. E., Jacobsen, S. E., Meinke, D. W., and Ray, A. (2002). DICER-LIKE1: blind men and elephants in arabidopsis development. Trends Plant Sci. **7**, 487–491.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., and Weigel, D. (2005). Specific Effects of MicroRNAs on the Plant Transcriptome. Developmental Cell **8**: 517–527.

Sielemann, J., Wulf, D., Schmidt, R., and Bräutigam, A. (2021). Local DNA shape is a general principle of transcription factor binding specificity in Arabidopsis thaliana. Nat Commun **12**: 6549.

Smith, L.M., Burbano, H.A., Wang, X., Fitz, J., Wang, G., Ural-Blimke, Y., and Weigel, D. (2015). Rapid divergence and high diversity of miRNAs and miRNA targets in the Camelineae. Plant J **81**: 597–610.

Song, X., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and Their Regulatory Roles in Plant–Environment Interactions. Annu. Rev. Plant Biol. **70**: 489–525.

Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol **22**: 96–118.

Sun, Z. et al. (2021). Integrated genomic analysis reveals regulatory pathways and dynamic landscapes of the tRNA transcriptome. Sci Rep **11**: 5226.

Swarts, D.C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R.F., Koonin, E.V., Patel, D.J., and Van Der Oost, J. (2014). The evolutionary journey of Argonaute proteins. Nat Struct Mol Biol **21**: 743–753.

Tang, J. and Chu, C. (2017). MicroRNAs in crop improvement: fine-tuners for complex traits. Nature Plants **3**: 17077.

Tarver, J.E., Donoghue, P.C.J., and Peterson, K.J. (2012). Do miRNAs have a deep evolutionary history? BioEssays **34**: 857–866.

Tarver, J.E., Taylor, R.S., Puttick, M.N., Lloyd, G.T., Pett, W., Fromm, B., Schirrmeister, B.E., Pisani, D., Peterson, K.J., and Donoghue, P.C.J. (2018). Well-Annotated microRNAomes Do Not Evidence Pervasive miRNA Loss. Genome Biology and Evolution **10**: 1457–1470.

Tsai,H.L.,Li,Y.H.,Hsieh,W.P.,Lin,M.C.,Ahn,J.H.,andWu,S.H.(2014).HUA ENHANCER1 is involved in posttranscriptional regulation of positive and negative regulators in arabidopsis photomorphogenesis. Plant Cell **26**, 2858–2872.

Valli, A.A., Santos, B.A.C.M., Hnatova, S., Bassett, A.R., Molnar, A., Chung, B.Y., and Baulcombe, D.C. (2016). Most microRNAs in the single-cell alga *Chlamydomonas reinhardtii* are produced by Dicer-like 3-mediated cleavage of introns and untranslated regions of coding RNAs. Genome Res. **26**: 519–529.

Vaucheret, H. (2008). Plant ARGONAUTES. Trends in Plant Science **13**: 350–358.

Wang, F., Polydore, S., and Axtell, M.J. (2015). More than meets the eye? Factors that affect target selection by plant miRNAs and heterochromatic siRNAs. Current Opinion in Plant Biology **27**: 118–124.

Wang, M., Xiao, Y., Su, N., and Song, Y. (2023). Editorial: Functional analysis of species-specific non-coding RNAs in plants. Front. Genet. **13**: 1105433.

Wang, S., Liang, H., Xu, Y., Li, L., Wang, H., Sahu, D.N., Petersen, M., Melkonian, M., Sahu, S.K., and Liu, H. (2021a). Genome-wide analyses across Viridiplantae reveal the origin and diversification of small RNA pathway-related genes. Commun Biol **4**: 412.

Wen, M., Lin, X., Xie, M., Wang, Y., Shen, X., Liufu, Z., Wu, C.-I., Shi, S., and Tang, T. (2016). Small RNA transcriptomes of mangroves evolve adaptively in extreme environments. Sci Rep **6**: 27551.

Wiberg, R.A.W., Halligan, D.L., Ness, R.W., Necsulea, A., Kaessmann, H., and Keightley, P.D. (2015). Assessing Recent Selection and Functionality at Long Noncoding RNA Loci in the Mouse Genome. Genome Biol Evol **7**: 2432–2444.

Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell. **75**:855-62.

Wu, L., Zhou, H., Zhang, Q., Zhang, J., Ni, F., Liu, C., and Qi, Y. (2010). DNA Methylation Mediated by a MicroRNA Pathway. Molecular Cell **38**: 465–475.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C. (2005). Expression of Arabidopsis *MIRNA* Genes. Plant Physiology **138**: 2145–2154.

You, C., Cui, J., Wang, H., Qi, X., Kuo, L.-Y., Ma, H., Gao, L., Mo, B., and Chen, X. (2017). Conservation and divergence of small RNA pathways and microRNAs in land plants. Genome Biol **18**: 158.

Zhan, J. and Meyers, B.C. (2023). Plant Small RNAs: Their Biogenesis, Regulatory Roles, and Functions. Annu. Rev. Plant Biol. **74**: 21–51.

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

Zhang, P. and Dai, M. (2022). CircRNA: a rising star in plant biology. Journal of Genetics and Genomics **49**: 1081–1092.

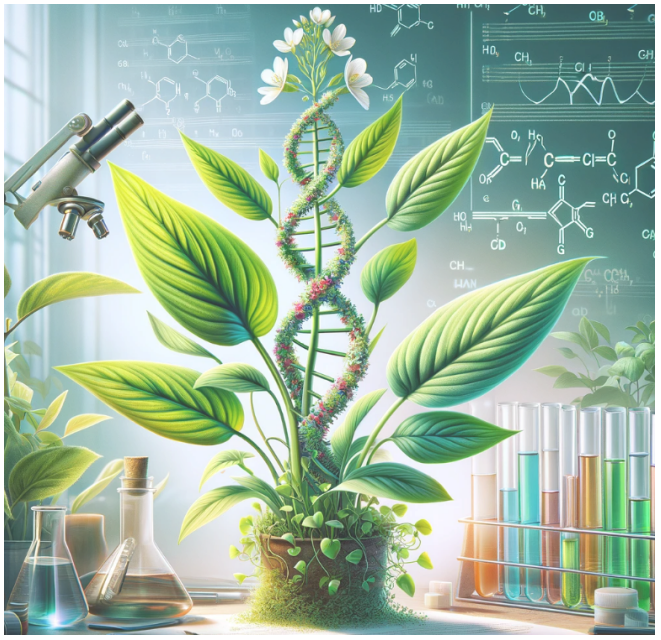
Zhang, X. et al. (2017). The Transcription Factor MYB29 Is a Regulator of *ALTERNATIVE OXIDASE1a*. Plant Physiol. **173**: 1824–1843.

Zhang, X., Niu, D., Carbonell, A., Wang, A., Lee, A., Tun, V., Wang, Z., Carrington, J.C., Chang, C.A., and Jin, H. (2014). ARGONAUTE PIWI domain and microRNA duplex structure regulate small RNA sorting in Arabidopsis. Nat Commun **5**: 5468.

Zhao, M., Meyers, B.C., Cai, C., Xu, W., and Ma, J. (2015). Evolutionary Patterns and Coevolutionary Consequences of *MIRNA* Genes and MicroRNA Targets Triggered by Multiple Mechanisms of Genomic Duplications in Soybean. Plant Cell **27**: 546–562.

Zhu, H., Zhou, Y., Castillo-González, C., Lu, A., Ge, C., Zhao, Y.-T., Duan, L., Li, Z., Axtell, M.J., Wang, X.-J., and Zhang, X. (2013). Bidirectional processing of pri-miRNAs with branched terminal loops by Arabidopsis Dicer-like1. Nat Struct Mol Biol **20**: 1106–1115.

CHAPTER I



The evolutionary history and functional specialization of microRNA genes in *Arabidopsis halleri* and *A. lyrata*.

Flavia Pavan[1], Jacinthe Azevedo Favory[2], Eléanore Lacoste[3], Chloé Beaumont[1], Firas Louis[1], Christelle Blassiau[1], Corinne Cruaud[4], Sophie Gallina[1], Mathieu Genete[1], Vinod Kumar[5], Ute Kramer[5], Rita A. Batista[1], Claire Patiou[1], Laurence Debacker[1], Chloé Ponitzki[1], Esther Houzé[1], Eléonore Durand[1], Jean-Marc Aury[3], Vincent Castric[1], Sylvain Legrand[1].

[1] Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France

[2] Laboratoire Génome et Développement des Plantes, UMR5096 CNRS/UPVD, Perpignan, France

[3] Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France

[4] Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, 91057, France

[5] Faculty of Biology and Biotechnology, Ruhr University Bochum, D-44801 Bochum, Germany

Author for correspondence : vincent.castric@univ-lille.fr

Abstract

MicroRNAs (miRNAs) are a class of small non-coding RNAs that play important regulatory roles in plant genomes. While some miRNA genes are deeply conserved, the majority appear to be species-specific, raising the question of how they emerge and integrate into cellular regulatory networks. We first performed a detailed annotation of miRNA genes in the closely related plants *Arabidopsis halleri* and *A. lyrata* and evaluated their phylogenetic conservation across 87 plant species. We then characterized the process by which newly emerged miRNA genes progressively acquire the properties of "canonical" miRNA genes, in terms of size and stability of the hairpin precursor, loading of their cleavage products into Argonaute proteins, and potential to regulate downstream target genes. Nucleotide polymorphism was lower in the mature miRNA sequence than in the other parts of the hairpin (stem, terminal loop), and the regions of coding sequences targeted by miRNAs also had reduced diversity as compared to their neighboring regions along the genes. These patterns were less pronounced for recently emerged than for evolutionarily conserved miRNA genes, suggesting a weaker selective constraint on the most recent miRNA genes. Our results illustrate the rapid birth-and-death of miRNA genes in plant genomes, and provide a detailed picture of the evolutionary processes by which a small fraction of them eventually integrate into "core" biological processes.

Key words: miRNAs, evolution, polymorphism, species-specific genes, Arabidopsis

1. Introduction

The origins of evolutionary novelties has been a topic of considerable interest in biology (Wagner, 2011). Following Francois Jacob's (1977) seminal concept of molecular tinkering (Jacob, 1977), the emergence of novel biological functions "from scratch" has long been considered an unlikely evolutionary event. Instead, evolution was believed to proceed through the modification of existing structures (such as protein-coding genes) following various forms of rearrangements such as small or large-scale duplications, fusions or fissions. Recently, however, it has become apparent that new genes do actually arise relatively readily as a result of a variety of processes including pervasive transcription throughout the genome, and the field has moved from "whether" new genes can arise to "how" they arise (Van Oss and Carvunis, 2019). A particularly debated question is whether the newly emerged "proto-genes" become gradually optimized by natural selection and eventually acquire the "canonical" gene-like characteristics (as per the "continuum model", Carvunis et al., 2012), or whether they result from the immediate apparition of rare "hopeful monsters", *i.e.* DNA sequences that are already pre-adapted and immediately exhibit gene-like characteristics with essentially no further optimization (as per the "preadaptation model", Wilson et al., 2017; McLysaght and Guerzoni, 2015). This process has been mostly studied in the particular case of protein-coding genes, and requires the broad-scale comparison of genes of different evolutionary ages that were formed at different times in the past along a phylogeny. However, not all genes are coding for proteins, and the study of the other sorts of genetic elements populating the genome is necessary for a comprehensive understanding of phenotypic evolution.

Regulatory RNAs are an important class of regulators of gene expression, and among them microRNAs (miRNAs) are key post-transcriptional regulators of gene expression in plants, animals, fungi and some viruses (Dexheimer and Cochella, 2020; Nanbo et al., 2021). miRNAs are expressed from genes that do not encode for proteins but are transcribed by RNA polymerase II into primary miRNAs (pri-miRNAs). These pri-miRNAs possess a hairpin-like structure recognized by DICER-LIKE (DCL) proteins in plants. DCL proteins cleave the pri-miRNA once to generate the pre-miRNA, and a second time to further release the miRNA/miRNA* duplex. The mature miRNA, typically a single 21 nucleotides-long RNA, is loaded into ARGONAUTE1 (AGO1) proteins, forming the RNA-induced silencing complex (RISC) in association with other proteins. The RISC recognizes messenger RNA (mRNA) targets through near-complete sequence complementarity with the mature miRNA, leading to negative regulation through mRNA degradation or translation inhibition (Reviewed in Zhan and Meyers, 2023; Ding and Zhang, 2023).

The recent availability of genome assemblies together with massive small RNA sequencing data has enabled broad-scale comparisons of the repertoire of miRNA genes across a growing number of plant and animal species, albeit with a strong bias toward model organisms. These comparisons revealed striking quantitative variation, with the total number of annotated miRNA genes ranging from just a few dozens to hundreds of miRNA genes per genome (miRBase v22, Kozomara et al., 2019). However, properly interpreting these variations has remained challenging because annotation of miRNA genes in plant and animal genomes is notoriously difficult due to their small size and the abundance of short inverted repeats, the heterogeneity of annotation methods, of the quality of genome assemblies, of molecular methodologies employed for small RNA sequencing, and of tissue types being compared. In spite of these caveats, the data available clearly indicate that, while some miRNAs are deeply conserved, lineage-specific miRNAs are also common, even between closely related species, suggesting a rapid evolutionary dynamics (Fahlgren et al., 2007; Cuperus et al., 2011; Nozawa et al., 2012; Chávez Montes et al., 2014).

The current model posits that “proto-miRNAs” genes originate from a variety of sources, including the inverted duplication of protein-coding genes, transposable element-related sequences, or regions of the genome that happened to contain inverted repeats and acquired the ability to be transcribed (reviewed in Cui et al., 2017). However, the abundance of proto-miRNAs relative to canonical miRNAs and the processes by which proto-miRNAs transition into canonical miRNAs have rarely been characterized in detail. Previous studies suggested that the process initially starts from stem-loops exhibiting near perfect complementarity that are the preferred substrate for DCL2, DCL3 or DCL4 proteins, imprecisely generating multiple duplexes of 24-nt-long small interfering RNAs (siRNAs) that are then loaded into AGO4 proteins. As the hairpin structure accumulates mutations over evolutionary time its complementarity is progressively disrupted, facilitating recognition by DCL1 and leading to the precise production of a single 21-nt-long mature miRNA preferentially loaded in AGO1, hence acquiring features of “canonical” miRNA genes (Allen et al., 2004; Voinnet et al., 2009; Baldrich et al., 2018; Pegler et al., 2023). Recent miRNA genes were also suggested to have weaker and spatio-temporally more limited expression territories than the more conserved miRNA genes, and that they also tend to be processed less precisely by DCL proteins leading to the production of a more diverse population of mature miRNAs (Fahlgren et al., 2007, 2010; Ma et al., 2010; Chávez Montes et al., 2014). Young miRNA genes tend to target genes associated with adaptation to local environments (Wen et al., 2016; Bradley et al., 2017), while highly conserved miRNA genes more often target genes involved in crucial cellular processes in plant development and stress responses (Dong et al., 2022). Two models have been proposed for the evolution of

miRNA-target interactions. In the "decay model" (Chen and Rajewsky, 2007; Roux et al., 2012), new miRNAs initially have many targets, most of which are deleterious, while only a few are beneficial. Over time, deleterious interactions are removed by natural selection and advantageous interactions are retained. In the "growth model" in contrast, older miRNAs have a greater number of targets than newer miRNAs, and the number of miRNA targets instead increases over the course of evolution (Nozawa et al., 2016). In this model, new miRNAs initially possess few targets, most of which are neutral with only a few being beneficial. This allows the level of expression of the miRNA gene to eventually increase and gradually acquire new targets over the course of evolution. While the target acquisition model has received some support in humans and mice (Chen and Rajewsky, 2007; Roux et al., 2012), the "growth model" has been favored in Drosophila (Nozawa et al., 2016). Hence, the relevance of these models, and the overall evolutionary significance of the newly acquired miRNA genes and their potential regulatory across genomes, has not been tested widely. Finally, young miRNA genes also seem to diverge more rapidly between related species, suggesting weaker functional constraints than that applying to older miRNA genes (Fahlgren et al., 2010; Ma et al., 2010). In *Arabidopsis thaliana,* the binding site within the genes targeted by miRNAs exhibited low polymorphism, indicating strong purifying selection (Ehrenreich and Purugganan, 2008; Smith et al., 2015). However, little is known about the microevolution of the binding site of the genes targeted by recently evolved miRNA genes.

In the genus Arabidopsis, a total of 221 miRNA genes have been annotated in the plant model *A. thaliana* (PmiREN 2.0, Guo et al., 2022b), and the companion papers by Ma et al., (2010) and Fahlgren et al., (2010) identified 154 and 164 miRNA genes, respectively in *A. lyrata*, from which *A. thaliana* diverged about 5 Myrs ago (Koch et al., 2000; Ossowski et al., 2010). These comparisons revealed a series of miRNA genes specific to either species, but given the rapid evolutionary dynamics of miRNA genes such broad-scale phylogenetic comparisons are inherently limited, and the comparison of even more closely related species are needed, as they represent a powerful way to reveal the most recently formed miRNA genes. *A. halleri* diverged from *A. lyrata* only one million years ago (Roux et al., 2011) and is a promising model, but the genome assemblies published for this sister species are highly fragmented (Briskine et al., 2017; Legrand et al., 2019), and the repertoire of annotated miRNAs is very incomplete (only 18 miRNAs have been deposited in the PmiREN 2.0 database, Guo et al., 2022b).

In this study, we explored the recent evolutionary dynamics of miRNA genes by focusing on *A. halleri* and *A. lyrata*. We first obtained a high-quality chromosome-level reference genome assembly for *A. halleri* and used sRNA-seq data from a variety of accessions to provide the

first comprehensive annotation of miRNA genes for this species and followed the same approach to compare them to those in the closely related *A. lyrata* genome. Immunoprecipitation of AGO1 and AGO4 proteins confirmed the validity of the majority of our miRNA gene predictions, including a substantial fraction of those specific to either *A. halleri* or *A. lyrata*, and analysis of the conservation patterns across the Viridiplantae provided a detailed picture of their evolutionary progression along the proto-miRNA - canonical miRNA continuum. Finally, we analyzed whole-genome resequencing data from natural *A. halleri* and *A. lyrata* accessions and showed that the functional constraint varied along the miRNA sequence in a manner that differed according to the evolutionary age of miRNA genes.

2. Results

2.1 Reference-level assembly of a *A. halleri* genome

We first produced a chromosome-level reference genome assembly for an individual from Northern France (Auby-1, from the Auby population, 50.40624°N, 3.08265°E) based on a combination of long Oxford Nanopore Technology reads, short Illumina reads and Hi-C data. Briefly, high molecular weight DNA from leaf tissue was extracted and a total of 29 Gb of sequence were obtained using a PromethION (Oxford Nanopore Technology). The 3.32 million reads had a N50 of 18.9 kbp (Supplemental Table S1). The high quality long reads were assembled using NECAT (Chen et al., 2021) and then polished first using all long reads with Racon (Vaser et al., 2017) and Medaka (https://github.com/nanoporetech/medaka) and then using Illumina short-reads with Hapo-G (Aury and Istace, 2021) (Supplemental Table S2). The resulting assembly was composed of 175 contigs and had a cumulative size of 227 Mbp with an N50 of 25.9Mb (Supplemental Table S2). The eight largest contigs covered 90% of the total length and had a size compatible with complete chromosomes (ranging from 22.2 to 31.7 Mbp). The remaining unanchored scaffolds represented only 8.4% of the assembly (Supplemental Figure S1). Hi-C (omni-C) sequencing data were generated to facilitate the chromosome-level assembly and were used to further orientate and anchor contigs to scaffolds (Supplemental Figure S1). We assessed the completeness of the reference genome using BUSCO and found 99.1% complete universal single-copy orthologs, 0.2% fragmented universal single-copy orthologs and 0.7% missing universal single-copy orthologs from the Brassica dataset odb10 (Supplemental Table S2). Overall, the resulting assembly has a sharply higher contiguity

than the one published by Legrand et al., (2019) with a 18-times lower number of scaffolds and 93-times higher N50 (Supplemental Table S3).

We used two approaches to annotate protein-coding genes in the genome. First, we aligned the protein sequences of *A. lyrata* and *A. thaliana* against the genome assembly using GeneWise (Birney et al., 2004) to search for homologs. Second, RNA-sequencing data were mapped to the reference genome using Hisat2 (Kim et al., 2019) and assembled by Stringtie (Shumate et al., 2022). Finally, we used Gmove (Dubarry et al., 2016) to combine these two sets of predictions. Overall, a total of 34,721 protein-coding genes were predicted. We used OrthoFinder (Emms and Kelly, 2019) to analyze orthology relationships between the predicted genes of *A. halleri*, *A. lyrata* and *A. thaliana*. After removing orthogroups containing paralogs, we identify 20,306 orthologous genes between *A. halleri* and *A. lyrata*, 13,082 orthologous genes between *A. lyrata* and *A. thaliana* and 13,977 orthologous genes between *A. halleri* and *A. thaliana*.

2.2 Annotation of the miRNA genes in the *A. halleri* Auby1 individual

To obtain a comprehensive set of miRNA genes in the *A. halleri* reference genome, we first generated ultra-deep small RNA sequencing (sRNA-seq) data from two tissues (leaves and a mix of flower buds at different stages of development) of the accession used to obtain the reference genome (Auby-1). We obtained a total of 206 and 159 million Illumina reads for the two sRNA-seq libraries (leaves and buds, respectively) (Supplemental Table S4). To enhance our ability to annotate miRNA genes, we combined predictions from miRkwood (Guigon et al., 2019) and Shortstack (Johnson et al., 2016), two algorithms that are adapted for plant genomes. While Shorstack is more conservative and predicts fewer miRNAs, miRkwood includes less reliable miRNA predictions but still with a majority of the miRNAs predicted in *A. thaliana* loaded in AGO1 or AGO4, which is considered high-level evidence for their regulatory potential (Guigon et al., 2019). Overall, after merging the predictions from the two tissues, we obtained a total of 332 predicted miRNA genes in the *A. halleri* reference genome (Supplemental Table S4).

To investigate whether sequencing depth could be a limiting factor for the discovery of miRNA genes, we randomly sub-sampled sequencing reads from the library with the highest number of reads (the one obtained from leaves, comprising 206 million reads), and newly predicted the miRNA genes using the exact same procedure in ten independent replicates for each sample size. We observed that a depth of 165 million reads is required to predict

90% of the total set of miRNA genes (Figure 1a), and observed no clear plateau of the number of predicted miRNA genes, indicating that even such a high sequencing depth remains a limiting factor, and that more miRNA genes with low abundance remain to be discovered. However, we note a clear inflection of the saturation curve once the first 86 miRNA genes have been discovered, suggesting that a limited set of miRNAs with relatively high abundance can already be revealed with a lower sequencing depth (around 20 million reads, as is classically done in many sRNA sequencing experiments).

2.3 Core and accessory miRNA genes in the *A. halleri* and *A. lyrata* reference genomes

To evaluate the variation of the repertoire of miRNA genes, we then aligned sRNA-seq data that we either generated ourselves (*n* = 6 libraries) or retrieved from the Sequence Read Archive (SRA) at the NCBI (*n* = 13 libraries) onto the *A. halleri* reference genome. For *A. lyrata*, we used the recently updated reference genome (accession MN47, Kolesnikova et al., 2023) and aligned reads from *n* = 3 sRNA-seq libraries that we generated and *n* = 10 sRNA-seq publicly available libraries. These data originate from a diversity of geographical accessions, plant tissues (leaves, buds and roots), developmental stages, sample preparation (such as True-seq, Nextflex Small RNA-Seq, SOLiD Total RNA-Seq, ION total RNA-seq), sequencing methods (SOLID, PROTON, Illumina) and sequencing depths (from two to 206 million reads) (Supplemental Table S4). Given this heterogeneity, the results are expected to buffer the inherent technical biases associated with individual sRNA sequencing experiments (Wright et al., 2019). Our analysis predicted between 46 and 267 miRNA genes per sample (Supplemental Table S4). After merging the predictions across samples, we identified a total of 463 and 276 miRNA genes in *A. halleri* and *A. lyrata*, respectively (Supplemental Table S5). The higher number detected in *A. halleri* is expected because of the larger number of sequencing datasets analyzed. Because a given miRNA precursor could produce different mature miRNA molecules in different accessions, these miRNA genes together resulted in a total of 678 and 521 mature miRNAs in *A. halleri* and *A. lyrata*, respectively (*i.e.* on average, a miRNA gene produced 1.5 and 1.9 mature miRNAs across all accessions in *A. halleri* and *A. lyrata*) (Supplemental Table S5). About a third of these miRNA genes were predicted by both software (287 in *A. halleri* and 87 in *A. lyrata*), while 145 and 176 genes were unique to miRkwood and 31 and 13 were unique to Shortstack in *A. halleri* and *A. lyrata,* respectively. The higher number of predictions made by miRkwood is in line with Li et al., (2021), who showed that miRkwood is able to predict substantially more miRNAs than other plant miRNA prediction tools.

2.4 Completeness of the repertoires

While some miRNA genes were broadly shared and predicted in at least 80% of the samples (8.6% and 9.8% in *A. halleri* and *A. lyrata*), a large proportion was private to a single sample (52.3% and 42.8% in *A. halleri* and *A. lyrata*) (Figure 2a, b). Hence, the number of "core" miRNA genes was relatively limited as compared to the accessory miRNome, noting that different individuals from the same species can carry or express different miRNA genes because of genetic or environmental variation. To evaluate the completeness of the set of miRNA genes we predicted in the *A. halleri* and *A. lyrata* genomes, we performed a saturation analysis by randomly subampling within the 21 and 13 individual samples from each species. We performed 1,000 replicates for each sample size and evaluated how the number of miRNA predictions increased with the number of samples upon which they are based. For *A. halleri,* we found that 18 of the 21 samples were needed to reach 90% of the total number of predictions (Figure 1b). Similarly, in *A. lyrata* 8 of the 11 samples needed to be included to reach 90% of the total number of predictions (Supplemental Figure S2). Hence, it is clear that the repertoire of miRNA genes in these two species was not saturated and was limited by the number of accessions that have been sequenced so far. In particular, our results show that analyses based on a single sequencing experiment in a single reference accession (as commonly performed) are likely underestimating the number of miRNA genes in a species by at least an order of magnitude. Altogether, our results suggest that our ability to discover miRNA genes remains limited both by the number of accessions and the sequencing depth.

2.5 A majority of miRNA predictions are validated by AGO-IP

We then performed immunoprecipitation of AGO1 and AGO4 proteins to provide formal experimental validation of our predictions. To broaden the set of miRNA genes we could discover, we analyzed three tissues (leaves, buds and roots) from a pool of six *A. halleri* or *A. lyrata* individuals per populations (Auby, France and I9, Italy for *A. halleri* and Plech, Germany for *A. lyrata*), and sequenced the small RNAs associated with these proteins as well as the input material (total cellular fraction). Out of the total set of miRNA genes predicted above, 314 and 147 were present in the *A. halleri* and *A. lyrata* input samples. A large majority of these predicted miRNA genes (88.2% and 83.4%, respectively) produced mature miRNAs associated with either AGO1 or AGO4 proteins (Figure 1c). Consistent with previous findings (Mi et al., 2008), the sRNAs loaded in AGO1 were predominantly 21-nucleotides-long with a 5' uridine (38.6% in *A. halleri* and 33.05% in *A. lyrata*), while the

sRNAs loaded in AGO4 were predominantly 24-nucleotides-long with a 5' adenosine (56.2% in *A. halleri* and 60.6% in *A. lyrata*) (Supplemental Figure S3). Therefore, our bioinformatic annotation strategy identifies *bona fide* miRNAs with a substantial number of canonical miRNA genes, including a large number of those that are accession-specific.

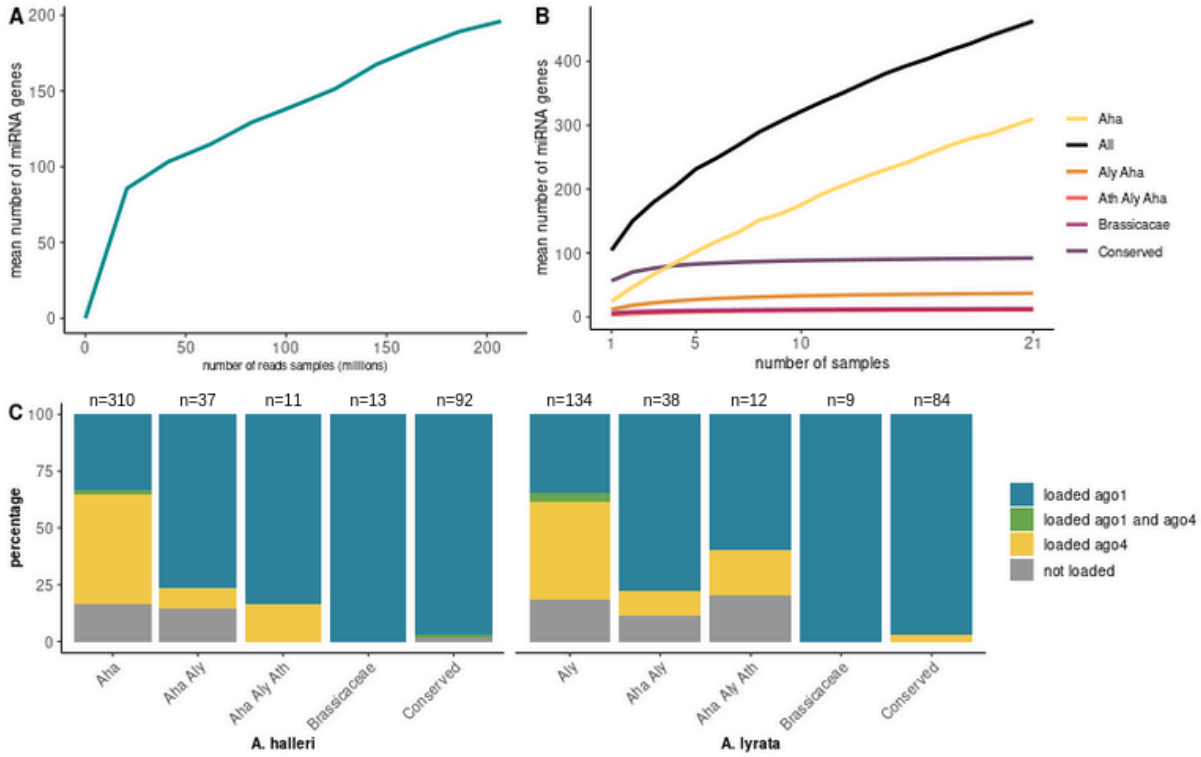


Figure 1. Annotation of miRNA gene repertoires. (a) Sequencing depth saturation curve for the most deeply sequenced accession (*A. halleri* Auby-1 accession, leaves). (b) Samples saturation curve for miRNA gene repertoires in *A. halleri* according to their conservation. (c) Percentage of miRNA genes loaded in AGO1 and/or AGO4 proteins out of 314 and 147 miRNA genes expressed in the input fractions in *A. halleri* and *A. lyrata*.

2.6 A minority of miRNA genes are conserved at a large phylogenetic scale

To evaluate the evolutionary age of miRNA genes, we combined three different strategies at increasingly divergent phylogenetic scales. First, we aimed at a fine-scale comparison between *A. halleri* and *A. lyrata.* To do this, we considered miRNA genes as syntenic if their hairpin sequences were reciprocal best-hits and they were flanked by syntenic genes. Second, we took advantage of the availability of assembled genomes and sRNA sequencing experiments in eleven Brassicaceae species to apply the exact same discovery pipeline we used in *A. halleri* and *A. lyrata.* Details of the genome assemblies and sRNA-seq experiments included are reported in Supplemental Table S6. Note that because of divergent genome structures and variable quality of the genome assemblies we did not attempt to

recover synteny relationships for this phylogenetic level. Finally, we extended the analysis to the broad set of Viridiplantae species included in the PmiREN 1.0. database, which was constructed by uniformly processing sRNA sequencing datasets and uses a recent set of criteria to identify miRNA genes (Guo et al., 2020). Because precursor sequences can diverge rapidly, we aligned the mature miRNAs predicted in *A. halleri* and *A. lyrata* (rather than full precursor sequences) to the mature miRNAs predicted in the 87 species of the database (Figure 2a), and considered mature miRNAs as homologous if they shared ⩾ 85% sequence similarity. Combining the results of the three analyses, we observed very similar patterns of conservation in both species (Figure 2b, c). We defined five groups of conservation for which we associated an age based on the divergence time in the phylogeny: 1) deeply conserved miRNAs shared by very distant species. This category represents 20% (*n* = 92) and 30% (*n* = 84) of the predicted miRNAs genes in *A. halleri* and *A. lyrata*, respectively (Figure 2b, c), and includes well-studied miRNA families such as miR156/miR157, miR166/miR161, miR169 and miR395. 2) miRNAs shared across the Brassicaceae family. This category represents 3% (*n* = 13) in *A. halleri* and 3% (*n* = 9) in *A. lyrata* (Figure 2b, c), and also includes some well-studied miRNA families such as miR158, miR845, miR400. 3) miRNAs shared between *A. thaliana*, *A. halleri* and *A. lyrata*. This category represents 2% (*n* = 11) in *A. halleri* and 4% (*n* = 12) in *A. lyrata* (Figure 2b, c), and also includes some well-studied miRNA families such as miR822, miR823 and miR842. 4) miRNAs shared only between *A. halleri* and *A. lyrata*. This category represents 8% (*n* = 37) in *A. halleri* and 14% (*n* = 38) in *A. lyrata*, including previously annotated families such as miR3433 and miR3443. 5) the *A. halleri*-specific miRNAs represent 67% of the *A. halleri* repertoire (*n* = 310) and the *A. lyrata*-specific miRNAs represents 51% of the *A. lyrata* repertoire (*n* = 134) (Figure 2b, c). Based on the divergence time between these two closely related species (Roux et al., 2011), we estimate that this last category of miRNAs appeared at most one million years ago. Overall, in both species we found that the vast majority of annotated miRNAs were either broadly conserved or species-specific, with only a small fraction of miRNAs showing intermediate levels of phylogenetic conservation.

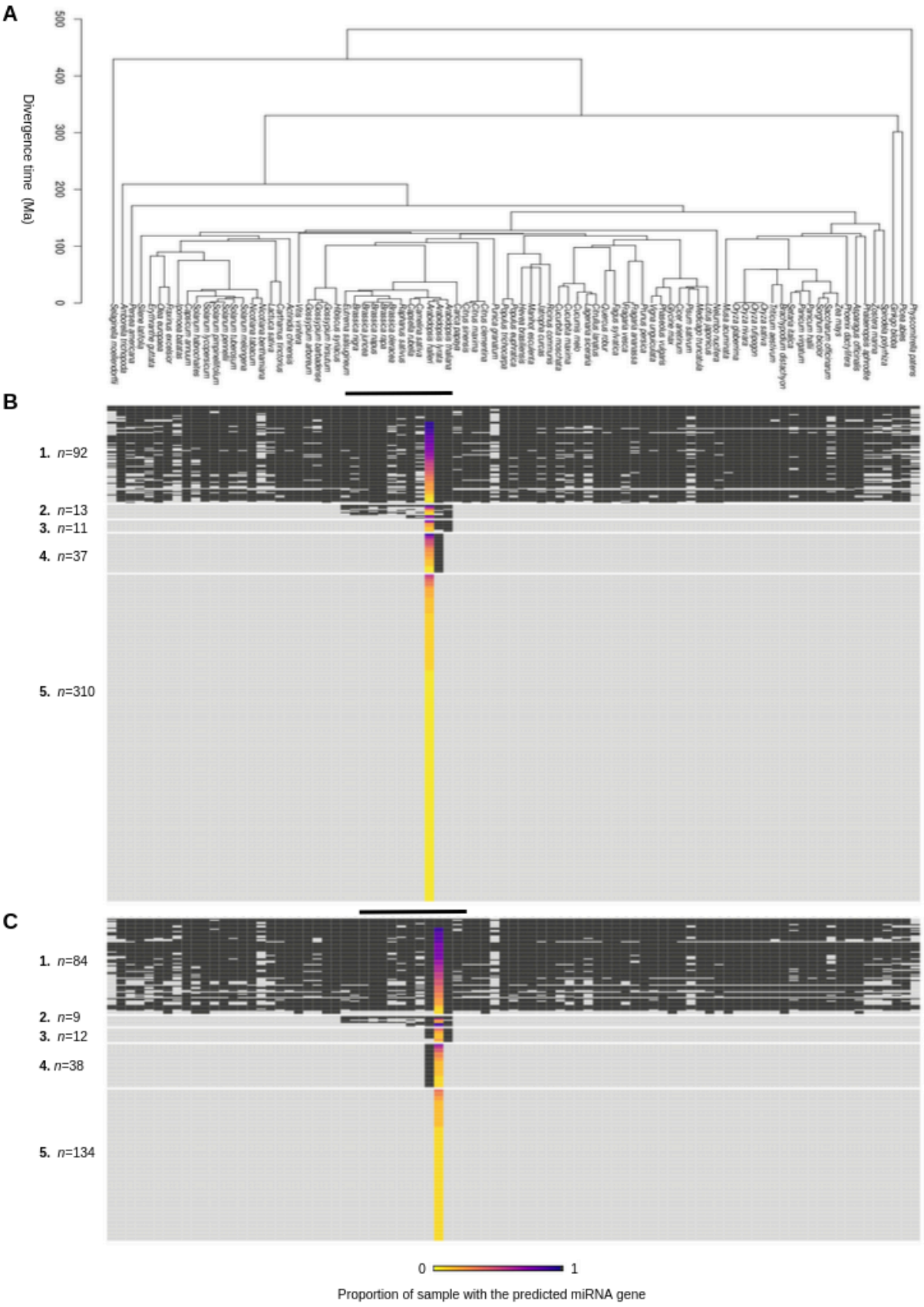


Figure 2: The majority of miRNA genes is either deeply conserved or species-specific. (a) Phylogenetic tree based on TimeTree v.5 of 87 Viridiplantae species present in the PmiREN database. The black bars indicate the Brassicaceae family. Overview of the miRNA

gene conservation (b) in *A. halleri* and (c) in *A. lyrata*. Each line corresponds to one miRNA gene, and species are represented in rows. Black squares indicate the presence of an homolog/ortholog, and gray squares its absence in the corresponding species. The number of miRNA genes in the five groups of conservation are indicated on the left part of the figure (1: deeply conserved, 2: shared with the Brassicaceae family, 3: shared with the Arabidopsis family, 4: shared between *A. halleri* and *A. lyrata,* 5: species-specifics). The proportion of accessions in which the miRNA gene was predicted is indicated by the colored bars from yellow (unique sample) to black (all samples).

2.7 Natural variation of the repertoire of deeply conserved and species-specific miRNAs

We then determined how the set of miRNA genes in each group of conservation varied with the number of samples included in the analysis. The species-specific miRNA genes tended to be detected in a smaller number of samples (8.2% and 10.8% of the samples in *A. halleri* and *A. lyrata,* respectively) than the deeply conserved genes (detected in 61.8% and 59.6% of the sample in *A. halleri* and *A. lyrata,* Figure 2b,c). Specifically, in *A. halleri*, 90% of the total number of predictions of the most deeply conserved miRNA genes were already annotated with only five of the 21 samples. Similarly, only six samples were needed to annotate 90% of the total number of miRNA genes shared within the Brassicaceae family, respectively. In contrast, up to nine and 14 samples were needed to annotate miRNA genes shared with *A. lyrata* and the *A. halleri*-specific genes (Figure 1b). Similarly, in *A. lyrata*, only five and three individuals were needed to identify 90% of the deeply conserved and the miRNA genes shared with the Brassicaceae family, while up to eight and nine individuals were needed for the miRNA genes shared with *A. halleri* and the *A. lyrata*-specific miRNA genes (Supplemental Figure S2). Overall, these results indicate that our analysis of multiple samples probably represents a comprehensive set of the deeply conserved miRNA genes, while the repertoire of species-specific miRNA genes is not saturated even with a large number of samples. Hence, including more samples would probably mostly increase the number of species-specific miRNA genes.

2.8 How young miRNAs become canonical miRNAs

Based on our evaluation of the evolutionary age of miRNA genes, we then sought to characterize how the more recent miRNAs genes differ from the more ancient ones. For this analysis, we merged the orthologous miRNA genes between *A. halleri* and *A. lyrata,* for which we took the average of each character value. Our dataset was composed of 97 deeply

conserved miRNA genes, 14 genes shared within the Brassicaceae family, 14 genes shared between *A. thaliana*, *A. halleri* and *A. lyrata,* 38 genes shared between *A. halleri* and *A. lyrata*, and 444 *A. halleri*- or *A. lyrata*-specific miRNA genes. We used linear regression models to evaluate how a series of molecular properties evolved with age of the miRNA genes. The mean normalized level of expression of the miRNA genes increased from 74.4 reads per million mapped reads (RPM) for the species-specific miRNA genes to 278.9 RPM for the most deeply conserved (adjusted R^2=0.28; p-value=2.66e-55) (Figure 3a; Supplemental Figure S5). Similarly, expression of the mature miRNA increased from 2.8 to 35.9 RPM (adjusted R^2=0.34; p-value=1.58e-70) (Supplemental Figure S4). These results are consistent with previous studies that showed that conserved miRNA genes tend to be expressed broadly at higher levels than the more recent miRNA genes (Cuperus et al., 2011). We note that in spite of this general trend, there is a strong variance within categories (as evidenced by the low adjusted R^2), and some of the most recent miRNA genes could still be expressed quite substantially, at levels comparable to those of some of the most conserved miRNAs.

Second, we analyzed the evolution of the size, stability and processing precision of the hairpins produced by the miRNA genes. We found that the average hairpin length tended to decrease over the course of evolution, with relatively long hairpins for species-specific and Arabidopsis-specific miRNA genes (mean size of 213 nt and 261 nt, respectively), but a shorter mean size of only 155 nt for the deeply conserved miRNAs (adjusted R^2= 0.05; p-value=1.57e-10) (Figure 3b; Supplemental Figure S5). We then estimated the minimal free energy index (MFEI) of each hairpin as an indicator of stability. The average MFEI increased from the species-specific (-1.21) to the deeply conserved miRNA genes (-0.96; adjusted R^2=0.07; p-value=3.86e-13) (Figure 3c; Supplemental Figure S5), corresponding to a decrease of the stability of the hairpin structure as miRNAs became more ancient. The hairpin structure, in particular the presence of bulges, is an important factor for cleavage by DCL proteins (Bologna et al., 2013). Following Ma et al., (2010), we defined the DCL processing precision of each miRNA gene as the abundance of mature miRNA sequences divided by the abundance of all the reads mapping to the hairpin. A score close to one indicates a high processing precision by DCL, while a score close to zero indicates an imprecise processing. The average processing precision increased from the species-specific miRNA genes (0.24) to the deeply conserved miRNA genes (0.67; adjusted R^2=0.38; p-value=4.24e-78) (Figure 3d; Supplemental Figure S5). Altogether, our results show that over the course of evolution, the hairpin produced by miRNA gene decreases in length, becomes more unstable and is processed more precisely by DCL proteins.

Third, we examined the size and 5’ nucleotide of miRNAs, as these features are known to be important for miRNA biogenesis and functions (Mi et al., 2008). The proportion of 21-nucleotides miRNAs with a uridine as the first 5’ nucleotide increased from the species-specific miRNAs (15%) to the deeply conserved (37%), while the proportion of 24-nucleotides miRNAs with an adenosine as the first 5' nucleotide decreased from 32% (species-specifics) to 0.3% (deeply conserved) (Figure 3e). AGO1 proteins select mainly 21-nucleotides miRNAs with a 5’ uridine, while AGO4 proteins select mainly 24-nucleotides miRNAs with a 5' adenosine (Mi et al., 2008), and accordingly we found that the vast majority of the conserved miRNAs were loaded in AGO1. This was especially true for the most conserved miRNAs (71/72, 99% and 66/68, 97% in *A. halleri* and *A. lyrata* respectively), but also for the miRNAs shared across the Brassicaceae family (100% for both species, 8/8 and 7/7 in *A. halleri* and *A. lyrata* respectively), those shared across the Arabidopsis genus (5/6, 83% and 3/5, 60% in *A. halleri* and *A. lyrata*) and those shared by *A. halleri* and *A. lyrata* miRNAs (16/21, 89% and 14/18, 78% in *A. lyrata*). A substantial proportion of the *A. halleri*- and of the *A. lyrata*-specific miRNAs (68/207, 33% and 17/49, 35% respectively) were also loaded in AGO1 (Figure 1c). Loading into AGO4 followed the opposite trend, as 99 of the 207 (48%) *A. halleri*-specific and 21 of the 49 (43%) *A. lyrata*-specific miRNAs were found in the AGO4 fraction. This proportion decreased rapidly as miRNA genes became older, with only 2/21 (9%) and 2/18 (11%) for miRNA shared by *A. halleri* and *A. lyrata,* respectively, 1/6 (17%) and 1/5 (20%) for miRNAs shared across the three Arabidopsis species, in *A. halleri* and in *A. lyrata*, respectively. None of the deeply conserved miRNAs in *A. halleri* and only two of the 68 deeply conserved miRNAs in *A. lyrata* (3%) were associated with AGO4 (Figure 1c). Finally, dual loading in both AGO1 and AGO4 was relatively rare, with only 7/207 of the *A. halleri*-specific, 2/49 of the *A. lyrata*-specific and 2/72 of the deeply conserved miRNAs in *A. halleri* being almost equally loaded in AGO1 and AGO4 (Figure 1c). Hence, our results provide a clear picture, where miRNAs produced by nearly all ancient miRNA genes are almost exclusively loaded in AGO1, while miRNAs produced by the very young miRNA genes are mainly loaded in AGO4 and a substantial proportion in AGO1. Thus, in spite of their limited conservation, a substantial proportion of these species-specific miRNAs may already have some regulatory potential.

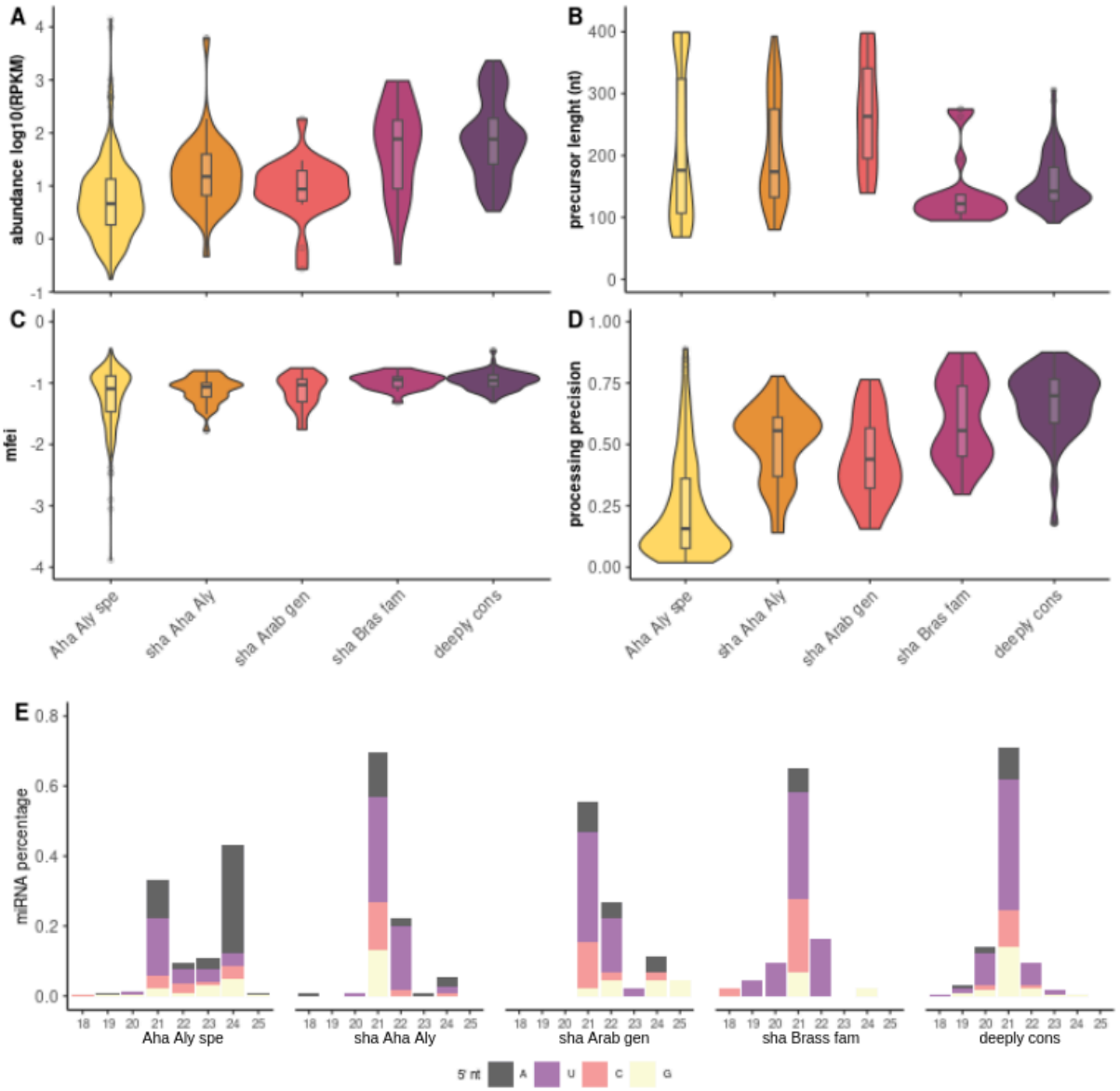


Figure 3: The characteristics of miRNA genes evolve slightly in the course of evolution. The orthologous genes between *A. halleri* and *A. lyrata* have been merged and categorized in five groups of conservation: the deeply conserved miRNA genes, those shared with the Brassicaceae family, those shared among *A. thaliana*, *A. halleri* and *A. lyrata*, those shared exclusively between *A. halleri* and *A. lyrata*, and the species-specific miRNA genes. (a) Expression level of the miRNA genes. (b) Length of the hairpin produced by the miRNA gene. (c) Hairpin stability, estimated using Minimum Free Energy Index (MFEI) calculation (d) DCL processing precision. (e) Mature miRNA size distribution and nature of the first 5' nucleotide.

2.9 The number of essential targets increases over the course of evolution.

To gain insight into the integration of miRNA genes in gene regulatory networks, we predicted the targets in the coding sequences (CDS) across the genome for each miRNA

gene using Targetfinder (Bo and Wang, 2005), and first evaluated the evolution of their number according to the age of the miRNA gene. The number of predicted targeted genes per miRNA gene was positively correlated with its age (adjusted R^2=0.17; p-value=3.09e-32) (Supplemental Figure S6), increasing from species-specific miRNA genes (0.87 targets on average per miRNA) to the most deeply conserved (5.42 targets on average per miRNA) (Figure 4a). Second, we determined the essentiality of the genes targeted by miRNAs using three proxies as in Legrand et al., (2019): 1) the size of the gene family (single-copy genes are predicted to be more essential due to the lack of functional redundancy); 2) the k_A/k_S ratio calculated from orthologous genes between *A. halleri, A. lyrata* and *A. thaliana*, for which lower values are expected for more essential genes; 3) the presence of loss-of-function (LOF) phenotype in *A. thaliana* mutant*s* (Lloyd and Meinke, 2012). After merging the orthologous miRNA genes, our dataset was composed of 262 genes targeted by the most deeply conserved miRNA genes, 40 by the miRNA genes shared across the Brassicaceae family, 17 by the miRNA genes shared between *A. thaliana*, *A. halleri* and *A. lyrata,* 129 by the miRNA genes shared between *A. halleri* and *A. lyrata* and 150 by the species-specific miRNA genes. The k_A/k_S ratios calculated from *A. halleri, A. lyrata* and *A. thaliana* divergence were negatively correlated with age of the miRNA gene (adjusted R^2=0.02; p-value=0.03), with a mean k_A/k_S of 0.22 and 0.33 for the genes targeted by species-specific miRNAs and miRNAs shared between *A. halleri* and *A. lyrata* respectively, and a lower value (k_A/k_S=0.18) for the genes targeted by the deeply conserved miRNAs (Figure 4b; Supplemental Figure S6). The average gene family size of the genes targeted was negatively correlated with age of the miRNA genes (adjusted R^2=0.01; p-value=0.003), decreasing from 1.47 and 5.01 for the genes targeted by species-specific and Brassicaceae-specific miRNAs to 1.38 for those targeted by deeply conserved miRNAs (Figure 4c; Supplemental Figure S6). The frequency of target genes with a LOF phenotype was correlated with age of the miRNA gene (p-value=0.009). However, the frequency of LOF genes initially decreased slightly (from 0.087 for the genes targeted by the species-specific genes, close to the genomic average, to 0.066 for the genes targeted by miRNAs shared by *A. halleri* and *A. lyrata*), but then increased sharply to 0.179 for those targeted by deeply conserved miRNAs (Figure 4d). Overall, our results indicate that the number of miRNA-target interactions increases over the course of evolution, along with the proportion of interactions involving essential genes.

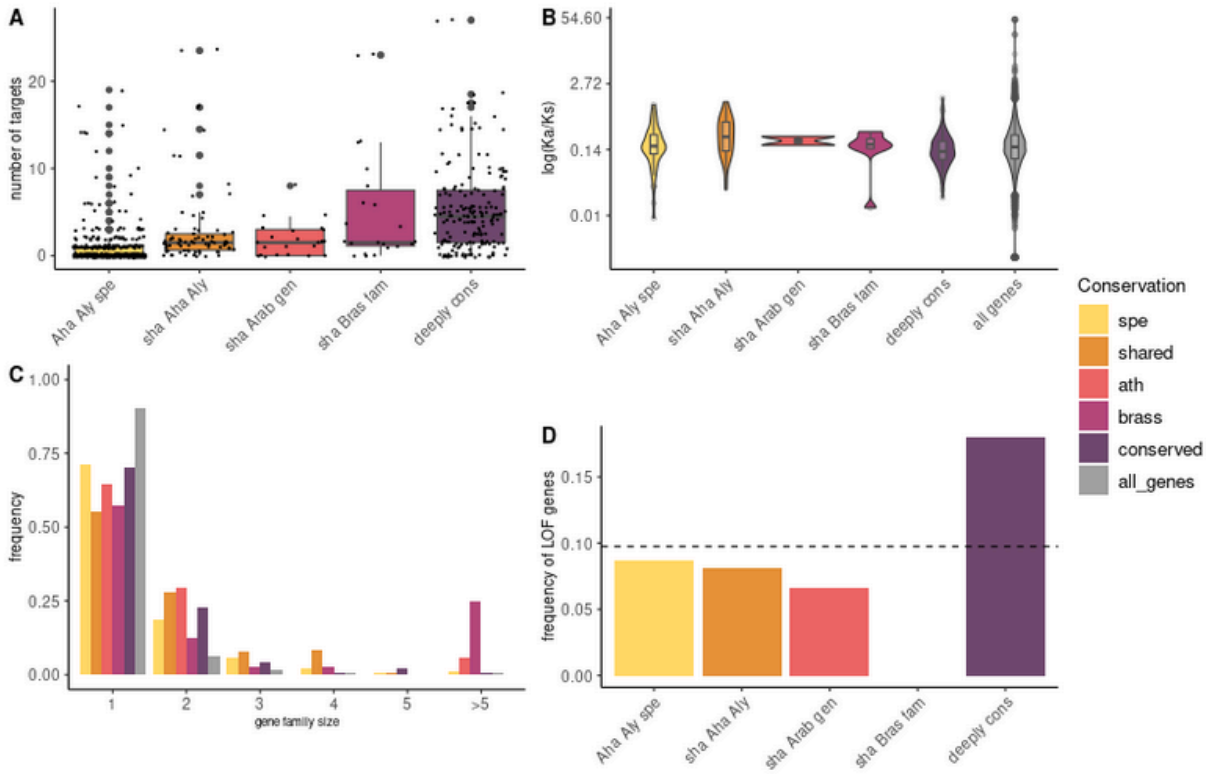


Figure 4: The old miRNA genes have more essential targets than the young ones. (a) Number of targets per miRNA gene according to its age. (b) Frequency of LOF phenotype genes in the miRNA genes targets. The frequency of LOF phenotype genes in all the genes present in the species is indicated by the dashed line. (c) k_A/k_S ratios of targeted genes calculated from *A. halleri, A. lyrata* and *A. thaliana* divergence. (d) Frequency of the targeted gene family size.

2.10 Functional constraint on the miRNA/miRNA* duplex over the course of evolution

To determine whether certain parts of the hairpin were more constrained by natural selection than others, we investigated nucleotide polymorphism of the 276 *A. lyrata* miRNA genes in 100 *A. lyrata* individuals from natural populations that we either newly sequenced or retrieved from published datasets. We determined the level of nucleotide polymorphism (π) for each part of the miRNA hairpins, including the miRNA, the miRNA*, the rest of the stem, the loop, as well as 200 bp of upstream and downstream flanking sequences (Figure 5a). Polymorphism of the miRNA/miRNA* duplex showed a decrease of about 53% in *A. lyrata* compared to the rest of the precursor (π=0.0062 vs. 0.0134), suggesting high selective constraint (Supplemental Figure S7). Strikingly, polymorphism of the duplex in the deeply conserved miRNA genes (mean π of 0.0062) was equivalent to the polymorphism of the 0-fold degenerate positions of protein-coding genes (mean π of 0.0047 for both species), suggesting that this part of the precursor evolves under considerable selective constraint

(Figure 5b). The overall level of polymorphism of the hairpin decreased from the species-specific (mean π of 0.0153) to the deeply conserved miRNA genes (0.0076) (Figure 5b). Polymorphism of the species-specific miRNA genes was similar to the polymorphism of the 4-fold degenerate positions across the genome (mean π of 0.0150). Thus, our results suggest that, collectively, the youngest miRNA genes tend to evolve close to neutrality, although we note that this conclusion does not preclude the possibility that some of them may be involved in the control of important biological functions. In contrast, the selective constraint on the more deeply conserved miRNAs is considerable, with levels of polymorphism of the miRNA/miRNA* duplex even lower than those of the most strongly constrained sites of protein-coding sequences.

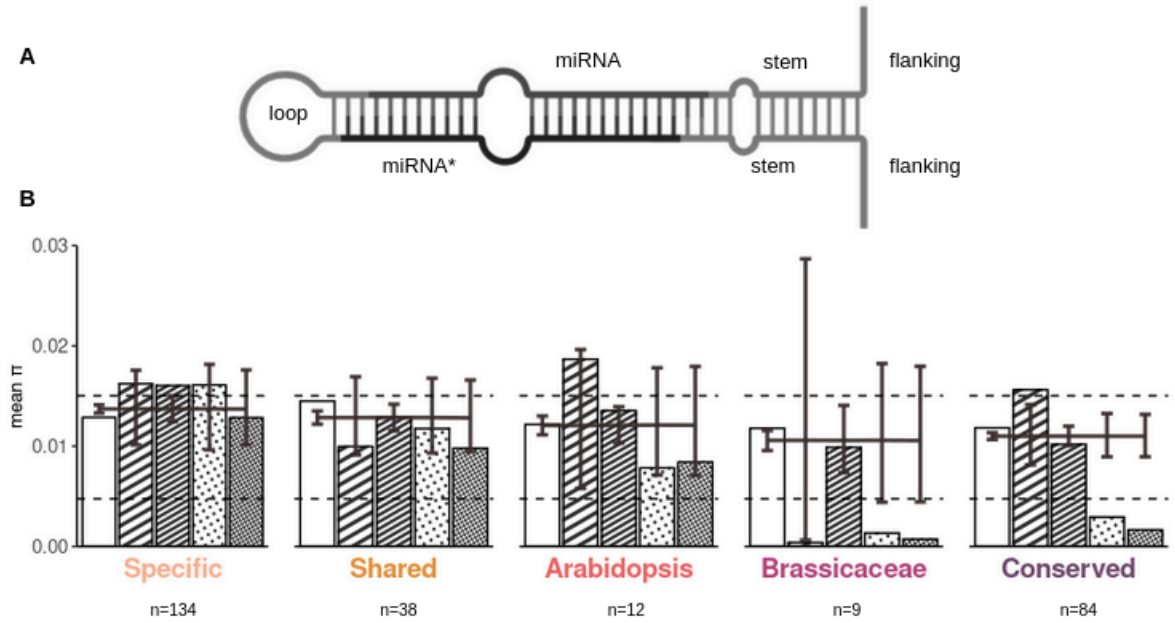


Figure 5: Selective constraints increase over the course of the evolution of miRNA genes. (a) Description of the miRNA hairpin regions with the upstream and downstream flanking regions (200 bp each). (b) *A. lyrata* average nucleotide diversity in the different parts of the miRNA hairpins. The dashed lines represent the mean π values for the 0 fold (lower line) and 4 fold (upper line) degenerate positions of all protein-coding genes of the genome. The bars represent the 95% confidence interval obtained by random permutation of nucleotides for 1,000 simulations.

2.11 Natural selection on the miRNA binding sites

We then asked whether the targeting by miRNAs could represent a detectable functional constraint along the coding sequence of their target genes. To test this hypothesis, we compared the polymorphism of 1,042 predicted binding sites in *A. lyrata* with that of their 300 bp upstream and downstream flanking regions along the target mRNAs. We observed slightly lower polymorphism of the binding site (average π 0.0090 in *A. lyrata*) as compared

to the flanking regions (average π 0.0098 in *A. lyrata*), *i.e.* a 8.8% reduction, suggesting that the presence of the miRNA binding site represents a detectable selective constraint on the CDS in addition to the original constraint of coding for a specific set of amino-acids (Figure 6).

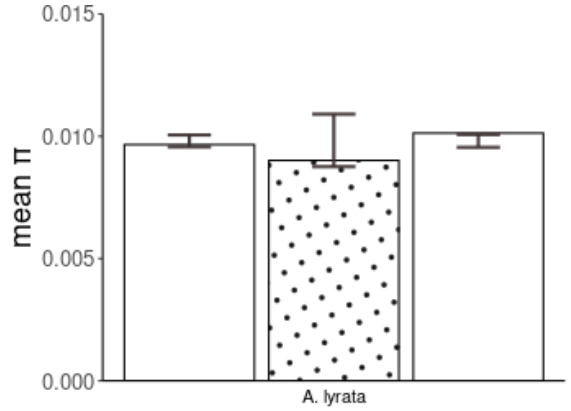


Figure 6: miRNA binding sites (dotted bar) **have lower** average nucleotide diversity in *A. lyrata* than their upstream and downstream flanking regions (300 bp each, white bars) . The whiskers represent the 95% confidence interval under the assumption of a random distribution of polymorphisms along the concatenated sequence and were obtained by random permutation of nucleotides for 1,000 simulations.

3. Discussion

3.1 Challenges in the identification of miRNAs in plant genomes

Proto-miRNAs have been proposed to emerge relatively readily, but studying their emergence and evolution has remained challenging because this requires the comparison of well-assembled and well-annotated closely related genomes, and high-quality deep sRNA sequencing data. Our deep miRNA annotation of the closely related *A. halleri* and *A. lyrata* genomes revealed both the long-term conservation and important evolutionary lability of these genetic elements. Identifying the complete set of miRNA genes in a species remains challenging for at least two reasons. First, in line with Ma et al., (2010) and Cuperus et al., (2011), we find that evolutionarily young miRNA genes tend to be expressed at low levels, and Chávez Montes et al., (2014) suggested that their expression territories might be limited in space and in time. Our results show that in spite of our extensive sequencing of a diversity of samples from diverse environmental conditions, tissues, or accessions of origin, the discovery of new miRNA genes is not yet exhausted in *A. halleri* and *A. lyrata*. Specifically, we observed a relatively limited “core” miRNAome, and the majority of miRNA genes belong to the “accessory” miRNAome, found in a single or a few samples only. To achieve an even

more comprehensive annotation, it will now be necessary to take into account the genomic variation among accessions by moving away from the alignment of sRNA-seq reads onto a single reference genome, and assembling individual genomes across a diverse set of natural accessions. While we took advantage of published data from a diversity of sources to maximize the number of accessions and environmental conditions, an important limitation is that we did not control for these factors. For a more detailed analysis it will also be necessary to compare miRNA annotations across accessions cultivated under a common garden environment. The second challenge is that annotation of miRNA genes relies on a set of criteria that have remained debated (reviewed in Axtell and Meyers, 2018). Here, we show that even though young miRNA genes tend to exhibit non-canonical features and can thus be hard to distinguish from false positives, in line with Guiguon et al., (2019) we observed that a vast majority of the predicted miRNA genes were experimentally validated by AGO1- and AGO4-IP, including a substantial fraction of the most evolutionarily recent ones. We conclude that the criteria we used for miRNA annotation were relatively stringent, and that the regulatory potential conferred by loading into AGO1 and/or AGO4 seems to be acquired rapidly, at least for some of them.

Fahlgren et al., (2010) compared miRNAs in *A. lyrata* and *A. thaliana* and estimated that 18% of them were *A. lyrata*-specific and 22% were *A. thaliana* specific. Here, with our deeper annotation, we found an even higher proportion, with up to 67% *A. halleri*- and 51% *A. lyrata*-specific miRNA genes. This difference illustrates the increased sequencing power over the last decade, and the effect of our strategy to multiply the number of accessions. In addition, Fahlgren et al., (2010) estimated that 134 miRNA genes were shared exclusively between *A. lyrata* and *A. thaliana*. Here, by including a much larger set of outgroup species (87 species in the Pmiren database), we restricted this set to only 11 or 12 miRNA genes specific to the Arabidopsis genus (depending on whether they were seen specifically in *A. halleri* or *A. lyrata*). This further illustrates that the vast majority of miRNA genes are either deeply conserved or species-specific, with very few showing intermediate levels of conservation. This small number might still be an overestimation, since the set of mature miRNA sequences for some species in the PmiREN database is probably incomplete (*e.g. Nicotiana benthamiana n*=73, *Punica granatum n*=33 or *Pisum sativum n*=51), possibly explaining the lack of homologs detected in the deeply conserved group for some species.

3.2 The evolutionary history of miRNA genes

A large proportion of the miRNA genes we identified are species-specific and have emerged recently, providing unprecedented power to explore the early steps of their evolution. Collectively, our results largely support the verbal model of emergence of miRNA genes proposed earlier (Allen et al., 2004; Voinnet et al., 2009; Baldrich et al., 2018; Pegler et al., 2023), whereby young miRNA genes start from near-perfect and relatively long hairpins, whose length and stability decrease by an accumulation of mutations creating bulges over the course of evolution together with a decrease of the diversity in size of the mature miRNA population, while the overall expression level and processing precision of the hairpin increases. Our findings parallel the observations made in the context of the *de novo* birth of protein-coding genes (Carnuvis et al., 2012; Wilson et al., 2017). Just like a large number of ORFs can be identified in a genome, we also identified a very large number of potential candidates being tested by natural selection, with possibly neutral or deleterious effects initially. Then only a very small fraction are retained over the long run, and eventually control essential cellular functions. The question of whether the miRNA genes that eventually become fixed have been slowly optimized by natural selection from imperfect progenitors, or rather represent “hopeful monsters” that were immediately beneficial when they arose is difficult to address directly. Yet, we note that the variance of molecular features among the group of the most recent miRNA genes is very large, so their distribution largely overlaps that of the canonical miRNA genes. Hence, our results are consistent with the idea that at least some of them may already exhibit features allowing them to function as efficiently as the highly conserved canonical miRNA genes.

3.3 Integration of young miRNA genes in the regulatory network

Although there are examples of young miRNA genes having important functional roles (Wen et al., 2016; Bradley et al., 2017), our results suggest most of them are unlikely to have essential biological functions, and are rapidly lost by genetic drift, mutation or natural selection (Fahlgren et al., 2010; Ma et al., 2010; Nozawa et al., 2012; Smith et al., 2015). Here, we found that although a substantial proportion of the young miRNA genes were loaded in AGO1 or AGO4, their expression level was low and their miRNA/miRNA* duplex evolved largely neutrally, suggesting that these genes may not have a significant effect on the cell or the individual. On the other hand, some miRNA genes are deeply conserved and the question of how a new miRNA integrates the functional regulatory network without impairing the fitness of the individual is still debated. Chen and Rajewsky, (2007) argued that

in humans young miRNA genes have many targets that appear at random in the genome, few of which are neutral or advantageous and many of which are slightly deleterious and will be lost. In contrast, Nozawa et al., (2016) argued that young miRNA genes have only few targets, most of which are neutral, and only a small fraction of which are beneficial. Neutral miRNA-targets interactions are rapidly lost through drift mutations, while beneficial ones are conserved under purifying selection. During this period, the expression level of the miRNA can increase to enable efficient suppression of its important targets and the miRNA may also acquire new targets because the chances of forming pairs with mRNAs is higher when it is more highly expressed itself (Nozawa et al., 2016). We observed that the young miRNA genes have fewer essential targets than older ones, supporting the “growth model”. Nonetheless, we found that the proportion of these interactions involving essential genes decreased before increasing again in the deeply conserved genes. This trend could result from natural selection initially removing deleterious interactions as the expression level of the miRNA gene increases.

A striking result of our analysis is the reduced nucleotide diversity of the miRNA binding site along the mRNA sequence. However, the extent of the reduction we observed is a lot weaker than that observed in *A. thaliana* by Ehrenreich and Purugganan, (2008). This study focused on miRNA binding sites that were validated by experimental data, so are probably enriched for the interactions with the strongest magnitude of regulatory effect. In addition, the annotation of miRNAs in this study relied on more limited data, and so were also enriched for the “low hanging”, most highly expressed miRNA genes that are easier to detect. It would be interesting to extend our analysis to evaluate the effect of the choice of miRNA-target interactions on the magnitude of the reduced diversity within the target sites.

3.4 Evolutionary significance of new miRNA genes

It is clear from our results that not all miRNA genes in a genome have the same evolutionary age. Some have been present for extended periods of time, while others emerged very recently. While it is clear that the most conserved miRNA genes fulfill essential biological functions, the evolutionary significance of the species-specific miRNA genes is harder to establish. This difficulty parallels that encountered for other genomic elements or cellular features. For instance, the evolution of long non-coding RNAs has been hotly debated. While key roles have been documented for some, (e.g. Statello et al., 2021), overall they seem to have little to no actual evolutionary importance, and most of them are largely dispensable (Goudarzi et al., 2019). Similarly, while alternative splicing is now recognized as a

widespread phenomenon, the fraction of alternative splicing events with actual adaptive role is possibly low, and the variation of this feature among species is best explained by the drift-barrier hypothesis (Benitière et al. 2023; Lynch 2007). Here, even though the species-specific miRNA are not conserved, we cannot exclude that some have important biological functions. One example of non-conserved miRNA genes obviously fulfilling an important biological function is given by the sRNA precursors controlling dominance interactions between self-incompatibility alleles in Arabidopsis (Durand et al., 2014). Similarly, sRNAs determining the patterns of adaptation to the local environments encountered by specific accessions would also not be expected to show strong conservation. Given the large number of new miRNA being tested by natural selection at a given time, it is possible that non-conserved miRNA may play an important role in the rapid adaptation of plants to changing environments. At the same time, it is also possible that the majority of species-specific miRNA genes may in fact be neutral, as suggested by their low number of predicted targets and the fact that the proportion of LOF genes among their predicted target genes closely matches that of a random draw across the genome.

To achieve a better understanding of the origin of new miRNA genes, it will now be necessary to investigate the molecular nature of their potential progenitors across the genome (see chapter II). In addition, their actual regulatory impact is currently hard to measure, and speculation can only be made on the basis of very indirect evidence. Designing experiments to determine whether at least some of them actually have the capacity to regulate their predicted target genes will be a challenging, yet fascinating next step (see Perspectives section).

4. Materials and methods

Plant material

A. halleri and *A. lyrata* plants were gathered from natural populations (see Supplementary table 7) and subsequently cultivated in standard greenhouse conditions. This cultivation aimed to produce leaves and buds for DNA and RNA extractions. For argonaute immunoprecipitation experiments, cuttings from six *A. halleri* Auby, ten *A. halleri* I9 and six *A. lyrata* Plech individuals were cultivated in hydroponic conditions in a growth medium composed of 1 mM $Ca(NO_3)_2$, 0.5 mM $MgSO_4$, 3 mM KNO_3, 0.5 mM $NH_4H_2PO_4$, 0.1 µM $CuSO_4$, 0.1 µM NaCl, 1 µM KCl, 2 µM $MnSO_4$, 25 µM H_3BO_3, 0.1 µM $(NH_4)_6Mo7O_{24}$, 20 µM

FeEDDHA, and 1 µM (*A. lyrata*) or 10 µM (*A. halleri*) $ZnSO_4$. The pH of the solution was maintained at 5.0 using MES acid buffer (2 mM). Roots were collected after six weeks.

A. halleri reference genome

High-molecular-weight DNA extraction, PromethION library preparation and sequencing

Two grams of fresh leaves were collected and flash-frozen. High molecular weight genomic DNA was extracted as described in (Belser et al., 2018). For Nanopore library preparation, the smallest genomic DNA fragments were first eliminated using the Short Read Eliminator Kit (Pacific Biosciences). Libraries were then prepared according to the protocol "1D Native barcoding genomic DNA (with EXP-NBD104 and SQK-LSK109)" provided by Oxford Nanopore Technologies. Depending of how many samples were pooled, 250ng (pool of 9 samples) to 1µg (pool of 4 samples) of genomic DNA fragments were repaired and end-prepped with the NEBNext FFPE DNA Repair Mix and the NEBNext Ultra II End Repair/dA-Tailing Module (New England Biolabs). Barcodes provided by ONT were then ligated using the Blunt/TA Ligase Master Mix (NEB). Barcoded fragments were purified with AMPure XP beads (Beckmann Coulter), then pooled and ONT adapters were added using the NEBNext Quick Ligation Module (NEB). After purification with AMPure XP beads (Beckmann Coulter), each library was mixed with the sequencing buffer (ONT) and the loading beads (ONT) and loaded on a PromethION R9.4.1 flow cell. In order to maintain the translocation speed, flow cells were refueled with 250µl Flush Buffer when necessary.

Illumina library preparation and sequencing

For Illumina PCR-free library preparation, 1.5 µg of genomic DNA was sonicated to a 100–1500-bp size range using a Covaris E220 sonicator (Covaris). The fragments (1µg) were end-prepped, and Illumina adapters (NEXTFLEX Unique Dual Index Barcodes, Perkin Elmer) were added using the Kapa Hyper Prep Kit (Roche). The ligation products were purified twice with 1X AMPure XP beads (Beckman Coulter). The libraries were then quantified by qPCR using the KAPA Library Quantification Kit for Illumina Libraries (Roche), and their profiles were assessed on an Agilent Bioanalyzer (Agilent Technologies). The libraries were sequenced on an Illumina NovaSeq 6000 instrument (Illumina) using 150 base-length read chemistry in a paired-end mode.

Assembly of the *Arabidopsis halleri* reference genome

We generated three sets of read samples: the complete set of reads, 30X coverage of the longest reads, and 30X coverage of the filtlong (https://github.com/rrwick/Filtlong)

highest-score reads (Supplemental Table S1). We then launched three different assemblers, Smartdenovo (Liu et al., 2021), Flye (Kolmogorov et al., 2019), and NECAT (Chen et al., 2021) on these three subsets of reads with the exception that NECAT was specifically run on the entire set of reads due to the implementation of a downsampling algorithm in its pipeline. Smartdenovo was launched with the parameters -k 17, as advised by the developers in case of larger genomes and -c 1 to generate a consensus sequence. Flye was launched with an estimated genome size of 240 Mbp and the -nano-raw option. NECAT was launched with a genome size of 240 Mbp and all other parameters set to their default values. Out of the 7 different assemblies obtained (Supplemental Table S8), we selected the Necat output for its higher contiguity (N50 > 1Mb) to continue our workflow. The Necat output was polished one time using Racon (Vaser et al., 2017) with Nanopore reads, then one time with Medaka (https://github.com/nanoporetech/medaka) (model r941_prom_hac_g507) and Nanopore reads, and two times with Hapo-G (1.3.4) (Aury and Istace, 2021) and Illumina short reads. We obtained an assembly of 518 contigs.

However the cumulative size of the assembly was higher than expected due to the high heterozygosity rate (320Mb vs 240Mb), and suggesting that the assembly size was currently inflated by the presence of allelic duplications. As indicated by BUSCO (Waterhouse et al., 2018) and KAT (Mapleson et al., 2017) (Supplemental Table S2 and Figure S9A), we observed the two alleles for many genes and a significant proportion of homozygous kmers were present twice in the assembly. We used HaploMerger2 (Huang et al., 2017) with default parameters and generated a haploid version of the assembly (Batch A twice to remove major misjoin and one run of Batch B). Haplomerger2 detected allelic duplications through all-against-all alignments and chose for each alignment the longest genomic regions (parameter -selectLongHaplotype), which may generate haplotype switches but ensure to maximize the gene content. We obtained two haplotypes: a reference version composed of the longer haplotype (when two haplotypes are available for a genomic locus) and a second version, named alternative, with the corresponding other allele of each duplicated genomic locus. At the end of the process, *A. halleri* haploid assembly has a cumulative size of 225 Mb, closer to the expected one, and KAT analysis showed a reduction of allelic duplications (Figure S9B). Additionally, the contig N50 benefited greatly from the separation and combination of the two haplotypes, rising to 3.3 Mb (Supplemental Table S2). Final assembly was polished one last time with Hapo-G and Illumina short reads to ensure that no allelic regions present twice in the diploid assembly have remained unpolished.

Chromosome-scale assembly was achieved using Hi-C data (Supplemental Table S1) and the 3D-DNA pipeline (version 180419) (https://github.com/aidenlab/3d-dna). Hi-C raw reads

were aligned against the assembly (-s none option) using Juicer (Durand et al., 2016). The resulting merged_nodups.txt file and the assembly were given to the run-asm-pipeline.sh script with the options "--editor-repeat-coverage 5 --splitter-coarse-stringency 30 --editor-coarse-resolution 100,000". Contact maps were visualized through the Juicebox tool (version 1.11.08) (https://github.com/aidenlab/Juicebox) and edited to adjust the construction of chromosomes or break misjoins (Supplemental Figure S1). After edition, the new.assembly file was downloaded from the Juicebox interface, filtered and converted into a fasta file using the juicebox_assembly_converter.py script. Finally, Hapo-G was run one last time on the chromosome-scale haploid assembly.

Genome annotation of the *Arabidopsis halleri* reference assembly

The *A. halleri* reference genome was masked using RepeatMasker (v.4.1.0, default parameters) (Smit AFA, Hubley R, Green P. RepeatMasker. http://repeatmasker.org/) and a home-made library of transposable elements (based on four *Arabidopsis* species) available on the *A. halleri* repository (see Data availability section). Using this procedure, 48.6% of the input assembly was masked.

Gene prediction was done using as input homologous proteins and RNA-Seq data. Proteins from *Arabidopsis thaliana* (TAIR10) and *Arabidopsis lyrata* (extracted from uniprot database) were aligned against *A. halleri* masked genome assembly in two steps. Firstly, BLAT (default parameters) (Kent, 2002) was used to quickly localize corresponding putative genes of the proteins on the genome. The best match and matches with a score ≥ 90% of the best match score were retained. Secondly, the alignments were refined using Genewise (default parameters) (Birney et al., 2004), which is more precise for intron/exon boundary detection. Alignments were kept if more than 50% of the length of the protein was aligned to the genome.

To allow the detection of expressed and/or specific genes, we also used short-read RNA-Seq data extracted from two tissues (leaves and flower buds) of the same *A. halleri* individual. Short-reads were mapped on the genome assembly using HiSat2 (version 2.2.1 with default parameters) (Kim et al., 2019). Bam files were then sorted and merged by tissue and Stringtie (version 2.2.1) (Shumate et al., 2022) was launched on each tissue with the following parameters (--rf -p 16 -v -m 150). At each genomic locus, we kept only the most expressed transcript.

Finally, we integrated the protein homologies and transcripts using a combiner called Gmove (-m 10000 -e 3 -score) (Dubarry et al., 2016). This tool can find CDSs based on genome

located evidence without any calibration step. Briefly, putative exons and introns, extracted from the alignments, were used to build a simplified graph by removing redundancies. Then, Gmove extracted all paths from the graph and searched for open reading frames (ORFs) consistent with the protein evidence. Completeness of the gene catalogs was assessed using BUSCO version 4.0.2 with the Brassica dataset odb10 and default parameters (Supplemental Table S2).

Identification of miRNAs

sRNA extraction, library preparation and sequencing

Total RNA from *A. halleri,* Auby1, PL22, I30 and *A. lyrata* CP99 and MN47 samples were extracted with the miRNeasy minikit (Qiagen). For *A. halleri* PL22, I30 and *A. lyrata* CP99 and MN47, 3 µg of total RNA were sent to LC Sciences for library construction and sequencing. For *A. halleri* Auby1, small RNAs (<200bp) were isolated from total RNA using the RNA clean and concentrator kit (ZymoReasearch). Library constructions were done with the NEXTFLEX Small RNA Sequencing Kit V3 (PerkinElmer). The libraries were sequenced on an Illumina NovaSeq 6000 instrument (Illumina) using 150 base-length read chemistry in a paired-end mode.

Additional data collection

Alongside the sRNA sequencing data produced in this study, various sets of sRNA sequencing data for *A. halleri*, *A. lyrata*, *A. thaliana, Camelina sativa, Capsella rubella, Raphanus sativus, Brassica oleracea, B. rapa, B. napus, B. juncea, B. nigra* and *Eutrema salsugineum* were obtained from the NCBI SRA database (https://www-ncbi-nlm-nih-gov.inee.bib.cnrs.fr/sra) (detailed information can be found in Supplemental Tables S4 and S6).

Identification of putative miRNA genes

The raw sRNA reads were processed according to miRkwood recommendations (https://bioinfo.univ-lille.fr/mirkwood/smallRNAseq/BED_file.php) using Python scripts performing adapter removal, trimming and quality filtering. Then, the sRNA reads were aligned to the reference genome of the respective species using Bowtie1 (Langmead et al., 2009), allowing for zero mismatch for the sample from *A. halleri* Auby1 and allowing for one mismatch for the other samples to be able detect isomirs (miRNA variants). The reference genomes used were those of the Auby1 (this present study) and MN47 accessions (Kolesnikova et al., 2023) for *A. halleri* and *A. lyrata*, respectively. For the remaining species, genome assemblies were downloaded from NCBI ASSEMBLY database

(https://www-ncbi-nlm-nih-gov.inee.bib.cnrs.fr/assembly/) (detailed information can be found in Supplemental Table S6).

Our annotation strategy consisted of combining miRNAs predicted by miRkwood (score >=5) (Guigon et al., 2019) and Shortstack 4.0.2 (Johnson et al., 2016). miRkwood include a set of filters defined in Axtell and Meyers, (2018) such as a threshold for the stability of the hairpin (MFEI < -0.8), for the reads mapping to each arm of the hairpin (at least ten), the accuracy of precursor cleavage, the existence of the mature miRNA (read frequency at least 33%), the presence of the miRNA/miRNA* duplex and its stability (Guignon et al., 2019). Then, we merged the common predictions between the different samples and removed the predictions that fell into small chromosomal contigs to obtain a unique repertory for each species. Finally, to gain higher confidence in these predictions we mapped our sRNA read data onto predicted miRNA precursors using structVis v0.4 for manual observation (https://github.com/MikeAxtell/strucVis).

Experimental validation of miRNA predictions

Deep-sequencing of Argonaute-associated small RNAs

Anti-AGO1 antibodies (AS09 527, Agrisera) and Anti-AGO4 antibodies (AS09 617, Agrisera). Inflorescence, leaf and root tissues from pooled individuals of *A. halleri* (Auby and I9) and *A. lyrata* (Plech) were ground in liquid nitrogen and were homogenized in extraction buffer EB (50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.2% v/v NP40, 10% glycerol, 10 µM MG132) containing the EDTA-free protease inhibitor cocktail (Roche). After 15 min of incubation at 4°C, cell debris were removed by centrifugation at 21,000*g* for 30 min at 4°C. The clarified lysate was incubated for 2 h at 4°C at 7-10 rpm, with AGO1 or AGO4 antibodies (from agrisera), and then 1 h at 4°C at 7-10 rpm with dynabeads protein A (Invitrogen), equilibrated with the EB . Beads were isolated using a magnetic rack, and washed once with 1 mL of EB and 4 times with 1 mL of PBS (Gibco). The sRNA were extracted from total/inputs and immunoprecipitated fractions using respectively Trizol and Trizol-LS, according to supplier instructions (Invitrogen) (Barre-Villeneuve et al., 2024). Subsequent sRNA libraries were performed and sequenced by the POPS platform from the plants science institute of Paris-Saclay (IPS2).

Bioinformatic analysis of AGO-IP libraries

After removal of adaptors, trimming and quality filtering, sequences were aligned onto the *A. halleri* and *A. lyrata* reference genomes with Bowtie1 allowing for one mismatch. We searched for an exact match between mature miRNA and sRNA read sequences and considered a miRNA loaded in AGO protein if more than 5 reads were found in the

immunoprecipitate data. For each sample, reads were normalized per million total mapped reads (RPM). Enrichment with respect to the immunoprecipitate was calculated as the ratio of reads in the immunoprecipitate to reads in the input.

Conservation analysis of miRNA genes

Synteny analysis of miRNA genes

The orthology maps of genes between *A. halleri* vs. *A. lyrata*, *A. halleri* vs. *A. thaliana* and *A. lyrata* vs. *A. thaliana* were constructed using protein sequences with OrthoFinder v2.5.4 (Emms and Kelly, 2019) using default parameters. Only orthogroups that contain one-to-one orthologues per species were kept for further comparison. Orthologous miRNAs between *A. halleri*, *A. lyrata* and *A. thaliana* were identified using the gene orthology maps described above. We selected miRNA genes located between upstream and downstream orthologous genes and restricted the size of the chromosomal fragment to 100 kb. The sequences of framed miRNA genes were aligned using the best-hit approach, commonly used to establish orthology relationships within genomes (Ward and Moreno-Hagelsieb, 2014). Two miRNA genes were considered syntenic if they were a reciprocal best match.

miRNA genes conservation across Viridiplantae

The miRNA families and the conservation across Viridiplantae were assigned based on similarity of mature miRNA sequences using the PmiREN 1.0 database (Guo et al., 2020). This database is specialized for plant miRNAs and is based on a standardized analysis of sRNA-seq data, which reduces the variability between predictions that would be due to the use of different tools. We filtered the database on mature miRNA sequence length requiring 18-nt to 25-nt sequences. In addition, we enriched the database with the predicted miRNAs from ten Brassicaceae species (*A. thaliana, Brassica juncea, B. napus, B. nigra, B. oleracea, B. rapa, Capsella rubella, Eutrema salsugineum, Camelina sativa, Raphanus sativus*), allowing us to be more precise about the conservation status of the miRNAs inside the Brassicaceae family. Then, the sequences of the mature miRNAs were aligned using Exonerate (Slater et al., 2005), allowing for three mismatch/gap/insertion. Alignments with a unique distant species (outside the Brassicaceae family) were considered as false positives.

Characterization of features of miRNA genes

We assessed the thermodynamic stability of the precursors using the Minimum Free Energy Index (MFEI) according to the equation MFEI = [MFE / sequence length x 100] / (G+C%) (Guignon et al., 2019). We determined the secondary structure MFE of the precursors using the RNAfold software (Lorenz et al., 2011) and used Python scripts to calculate the GC content.

From secondary structure, we further defined the different parts of the miRNA precursors (miRNA/miRNA* duplex, loop, stem and the flanking regions) using python scripts.
We calculated the abundance miRNAs in each sample where they were predicted and took the average value. The abundance of precursors was defined as the reads mapping the precursor normalized per 1,000,000 total mapped reads and the precursor length (RPKM). The abundance of mature miRNAs was normalized per 1,000,000 total mapped reads (RPM).
The associations between miRNA features and their age were examined with regression linear models using R (v4.1.2; R Core Team 2023).

Targets characterization

Targets prediction

We identified the potential miRNA targets in the CDS of *A. halleri* and *A. lyrata* using TargetFinder (Bo and Wang, 2005) with default parameters, which provide the best balance between specificity and sensitivity (Srivastava et al., 2014). We applied a cut-off penalty score of ≤ 3 as recommended in Fahlgren et al., (2007) for reliable miRNA-mRNA target interactions.

Proxies of essentiality of *A. halleri* and *A. lyrata* genes

Three proxies have been used as in Legrand et al., (2019) to assess gene essentiality. Briefly, the all-against-all Blast method was employed using the CDS to estimate the size of the gene family. The hits with a query coverage inferior to 50% and/or an e-value superior to 1e-30 were discarded. Ka/Ks was estimated using KaKs_Calculator2.0 (Wang et al., 2010) with the Goldman and Yang method (Goldman et al., 1994) from the alignments of pairs of orthologous CDS between *A. halleri* vs. *A. thaliana* and *A. lyrata* vs. *A. thaliana* obtained using Water from the EMBOSS package (Rice et al., 2000). Finally, loss of function genes were identified using a dataset composed of 2400 Arabidopsis genes with a loss-of-function mutant phenotype (Lloyd and Meinke, 2012).
The associations between target gene features and miRNA gene ages were examined with regression linear models using R (v4.1.2; R Core Team 2023), except for loss-of-function genes proxy for which we used a Chi-squared test on all conservation groups.

Polymorphism analysis

Data collection

To assess the genomic diversity, we analyzed 100 *A. lyrata* individuals from natural accessions. In addition to the genomic data produced, we downloaded WGS data obtained

by Takou et al., (2021) and Mattila et al., (2017). The set was composed of 39 individuals from Michigan, USA (this study); Spiterstulen, Norway (24 individuals) (Mattila et al., 2017; Takou et al., 2021); Stubbsand, Sweden (6 individuals) (Mattila et al., 2017); Plech, Germany (18 individuals) (Mattila et al., 2017; Takou et al., 2021); Austria (7 individuals) (Takou et al., 2021); Mayodan, USA (6 individuals) (Mattila et al., 2017).

Variant calling and pi calculation

After adapters removal, the reads were mapped to the reference genomes of *A. halleri* and *A. lyrata* using bowtie2 (Langmead et al., 2012) and PCR duplicated reads were removed with picard MarkDuplicates version 2.21.4 (available at http://broadinstitute.github.io/picard). GATK version 4.1.9.0 (McKenna et al., 2010) was used to call and annotate single nucleotide polymorphisms (SNPs) using haplotypecaller. Individual GVCF files were subjected to joint genotyping to obtain a .vcf file with information on all sites, both variant and invariant. We extracted the precursors, targets and flanking regions and filtered the resulting .vcf files with VCFtools version 0.1.16 (Danecek et al., 2011) using the following options --remove-indels --min-alleles 1 --max-alleles 2 --max-missing 0.75 --minDP 5. The average number of nucleotide differences between genotypes (π) was calculated using VCFtools version 0.1.16 (Danecek et al., 2011). Additionally, we carried out permutation tests to assess the probability that the differences we observed could be due to our result being different from chance alone and thus determining its significance. Specifically, For example, each nucleotide associated with its nucleotide diversity value (π) was permuted within the hairpin, and then the average π of each region of the hairpin was calculated. This was repeated a large number of times (n=1000), allowing us to define a confidence interval. If the average π observed for the hairpin part was outside the confidence interval, this meant that the observed value was different from chance and therefore significant.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Availability of supporting data

The Illumina and Oxford Nanopore sequencing data are available in the European Nucleotide Archive under the following project PRJEB70878.

Acknowledgments

This project was funded by Région Nord Pas de Calais (MICRO^2 project) to VC and SL, ERC (NOVEL project, grant #648321) to VC, ANR (project TE-MoMa, grant

ANR-18-CE02-0020-01) to VC and SL. We thank the high performance computing service and Bilille at the University of Lille for providing computing resources. This work was performed using the infrastructure and technical support of the “Plateforme Serre, cultures et terrains expérimentaux – Université de Lille” for the greenhouse/field facilities. We thank Anamaria Nesculesca, Filipe Borges for taking part in FP’s PhD committee, and Blake Meyers, Noah Fahlgren, Patricia Baldrich and Xavier Vekemans for discussions.

5. References

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.-H., Spatafora, J.W., and Carrington, J.C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat Genet **36**: 1282–1290.

Aury, J.-M. and Istace, B. (2021). Hapo-G, haplotype-aware polishing of genome assemblies with accurate reads. NAR Genomics and Bioinformatics **3**: lqab034.

Axtell, M.J. and Meyers, B.C. (2018). Revisiting Criteria for Plant MicroRNA Annotation in the Era of Big Data. Plant Cell **30**: 272–284.

Baldrich, P., Beric, A., and Meyers, B.C. (2018). Despacito: the slow evolutionary changes in plant microRNAs. Current Opinion in Plant Biology **42**: 16–22.

Barre-Villeneuve, C., Laudié, M., Carpentier, M.-C., Kuhn, L., Lagrange, T., and Azevedo-Favory, J. (2024). The unique dual targeting of AGO1 by two types of PRMT enzymes promotes phasiRNA loading in *Arabidopsis thaliana*. Nucleic Acids Research: gkae045.

Belser, C. et al. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nature Plants **4**: 879–887.

Bénitière, F., Necsulea, A., and Duret, L. (2024). Random genetic drift sets an upper limit on mRNA splicing accuracy in metazoans.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res. **14**: 988–995.

Bo, X. and Wang, S. (2005). TargetFinder: a software for antisense oligonucleotide target site selection based on MAST and secondary structures of target mRNA. Bioinformatics **21**: 1401–1402.

Bologna, N.G., Schapire, A.L., Zhai, J., Chorostecki, U., Boisbouvier, J., Meyers, B.C., and Palatnik, J.F. (2013). Multiple RNA recognition patterns during microRNA biogenesis in plants. Genome Res. **23**: 1675–1689.

Bradley, D. et al. (2017). Evolution of flower color pattern through selection on regulatory small RNAs. Science **358**: 925–928.

Briskine, R.V., Paape, T., Shimizu-Inatsugi, R., Nishiyama, T., Akama, S., Sese, J., and Shimizu, K.K. (2017). Genome assembly and annotation of *Arabidopsis halleri* , a model for heavy metal hyperaccumulation and evolutionary ecology. Molecular Ecology Resources **17**: 1025–1036.

Carvunis, A.-R. et al. (2012). Proto-genes and de novo gene birth. Nature **487**: 370–374.

Chávez Montes, R.A., Rosas-Cárdenas, D.F.F., De Paoli, E., Accerbi, M., Rymarquis, L.A., Mahalingam, G., Marsch-Martínez, N., Meyers, B.C., Green, P.J., and De Folter, S. (2014). Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. Nat Commun **5**: 3722.

Chen, K. and Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet **8**: 93–103.

Chen, Y. et al. (2021). Efficient assembly of nanopore reads via highly accurate and intact error correction. Nat Commun **12**: 60.

Cui, J., You, C., and Chen, X. (2017). The evolution of microRNAs in plants. Current Opinion in Plant Biology **35**: 61–67.

Cuperus, J.T., Fahlgren, N., and Carrington, J.C. (2011). Evolution and Functional Diversification of *MIRNA* Genes. Plant Cell **23**: 431–442.

Danecek, P. et al. (2011). The variant call format and VCFtools. Bioinformatics **27**: 2156–2158.

Dexheimer, P.J. and Cochella, L. (2020). MicroRNAs: From Mechanism to Organism. Front. Cell Dev. Biol. **8**: 409.

Ding, N. and Zhang, B. (2023). microRNA production in Arabidopsis. Front. Plant Sci. **14**: 1096772.

Dong, Q., Hu, B., and Zhang, C. (2022). microRNAs and Their Roles in Plant Development. Front. Plant Sci. **13**: 824240.

Dubarry, M. et al., *Gmove a Tool for Eukaryotic Gene Predictions Using Various Evidences* (F1000Research, 2016).

Durand, E. et al. (2014). Dominance hierarchy arising from the evolution of a complex small RNA regulatory network. Science **346**: 1200–1205.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Systems **3**: 95–98.

Ehrenreich, I.M. and Purugganan, M.D. (2008). Sequence Variation of MicroRNAs and Their Binding Sites in Arabidopsis. Plant Physiology **146**: 1974–1982.

Emms, D.M. and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol **20**: 238.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., and Carrington, J.C. (2007). High-Throughput Sequencing of Arabidopsis microRNAs: Evidence for Frequent Birth and Death of MIRNA Genes. PLoS ONE **2**: e219.

Fahlgren, N., Sullivan, C.M., Kasschau, K.D., Chapman, E.J., Cumbie, J.S., Montgomery, T.A., Gilbert, S.D., Dasenko, M., Backman, T.W.H., Givan, S.A., and Carrington, J.C. (2009). Computational and analytical framework for small RNA profiling by high-throughput sequencing. RNA **15**: 992–1002.

Fahlgren, N., Jogdeo, S., Kasschau, K.D., Sullivan, C.M., Chapman, E.J., Laubinger, S., Smith, L.M., Dasenko, M., Givan, S.A., Weigel, D., and Carrington, J.C. (2010). MicroRNA Gene Evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. The Plant Cell **22**: 1074–1089.

François Jacob (1977). Evolution and Tinkering. Science **196**: 1661–1666.

Goldman N. and Yang Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biology and Evolution.

Goudarzi, M., Berg, K., Pieper, L.M., and Schier, A.F. (2019). Individual long non-coding RNAs have no overt functions in zebrafish embryogenesis, viability and fertility. eLife **8**: e40815.

Guigon, I., Legrand, S., Berthelot, J.-F., Bini, S., Lanselle, D., Benmounah, M., and Touzet, H. (2019). miRkwood: a tool for the reliable identification of microRNAs in plant genomes. BMC Genomics **20**: 532.

Guo, Z. et al. (2020). PmiREN: a comprehensive encyclopedia of plant miRNAs. Nucleic Acids Research **48**: D1114–D1121.

Guo, Z., Kuang, Z., Deng, Y., Li, L., and Yang, X. (2022a). Identification of Species-Specific MicroRNAs Provides Insights into Dynamic Evolution of MicroRNAs in Plants. IJMS **23**: 14273.

Guo, Z., Kuang, Z., Zhao, Y., Deng, Y., He, H., Wan, M., Tao, Y., Wang, D., Wei, J., Li, L., and Yang, X. (2022b). PmiREN2.0: from data annotation to functional exploration of plant microRNAs. Nucleic Acids Research **50**: D1475–D1482.

Huang, S., Kang, M., and Xu, A. (2017). HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. Bioinformatics **33**: 2577–2579.

Johnson, N.R., Yeoh, J.M., Coruh, C., and Axtell, M.J. Improved Placement of Multi-mapping Small RNAs. *G3 Genes*|*Genomes*|*Genetics* **6:**2103–2111.

Kent, W.J. (2002) BLAT—The BLAST-Like Alignment Tool.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol **37**: 907–915.

Koch, M.A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative Evolutionary Analysis of Chalcone Synthase and Alcohol Dehydrogenase Loci in Arabidopsis, Arabis, and Related Genera (Brassicaceae). Molecular Biology and Evolution **17**: 1483–1498.

Kolesnikova, U.K., Scott, A.D., Van De Velde, J.D., Burns, R., Tikhomirov, N.P., Pfordt, U., Clarke, A.C., Yant, L., Seregin, A.P., Vekemans, X., Laurent, S., and Novikova, P.Y. (2023). Transition to Self-compatibility Associated With Dominant *S* -allele in a Diploid Siberian Progenitor of Allotetraploid *Arabidopsis kamchatica* Revealed by *Arabidopsis lyrata* Genomes. Molecular Biology and Evolution **40**: msad122.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol **37**: 540–546.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Research **47**: D155–D162.

Kubota, S., Iwasaki, T., Hanada, K., Nagano, A.J., Fujiyama, A., Toyoda, A., Sugano, S., Suzuki, Y., Hikosaka, K., Ito, M., and Morinaga, S.-I. (2015). A Genome Scan for Genes Underlying Microgeographic-Scale Local Adaptation in a Wild Arabidopsis Species. PLoS Genet **11**: e1005361.

Kumar, S., Suleski, M., Craig, J.M., Kasprowicz, A.E., Sanderford, M., Li, M., Stecher, G., and Hedges, S.B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. Molecular Biology and Evolution **39**: msac174.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods **9**: 357–359.

Legrand, S. et al. (2019). Differential retention of transposable element-derived sequences in outcrossing Arabidopsis genomes. Mobile DNA **10**: 30.

Li, J., Reichel, M., Li, Y., and Millar, A.A. (2014). The functional scope of plant microRNA-mediated silencing. Trends in Plant Science **19**: 750–756.

Li, Q., Liu, G., Bao, Y., Wu, Y., and You, Q. (2021). Evaluation and application of tools for the identification of known microRNAs in plants. Appl Plant Sci **9**.

Liu, H., Wu, S., Li, A., and Ruan, J. (2021). SMARTdenovo: a de novo assembler using long noisy reads. Gigabyte **2021**: 1–9.

Lloyd, J. and Meinke, D. (2012). A Comprehensive Dataset of Genes with a Loss-of-Function Mutant Phenotype in Arabidopsis. Plant Physiology **158**: 1115–1129.

Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms Mol Biol **6**: 26.

Lynch, M. (2007). The evolution of genetic networks by non-adaptive processes. Nat Rev Genet **8**: 803–813.

Ma, Z., Coruh, C., and Axtell, M.J. (2010). *Arabidopsis lyrata* Small RNAs: Transient *MIRNA* and Small Interfering RNA Loci within the *Arabidopsis* Genus. Plant Cell **22**: 1090–1103.

Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J., and Clavijo, B.J. (2017). KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics **33**: 574–576.

Mattila, T.M., Tyrmi, J., Pyhäjärvi, T., and Savolainen, O. (2017). Genome-Wide Analysis of Colonization History and Concomitant Selection in Arabidopsis lyrata. Molecular Biology and Evolution **34**: 2665–2677.

McLysaght, A. and Guerzoni, D. (2015). New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. Phil. Trans. R. Soc. B **370**: 20140332.

Mi, S. et al. (2008). Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5′ Terminal Nucleotide. Cell **133**: 116–127.

Nanbo, A., Furuyama, W., and Lin, Z. (2021). RNA Virus-Encoded miRNAs: Current Insights and Future Challenges. Front. Microbiol. **12**: 679210.

Nozawa, M., Fujimi, M., Iwamoto, C., Onizuka, K., Fukuda, N., Ikeo, K., and Gojobori, T. (2016). Evolutionary Transitions of MicroRNA-Target Pairs. Genome Biol Evol **8**: 1621–1633.

Nozawa, M., Miura, S., and Nei, M. (2012). Origins and Evolution of MicroRNA Genes in Plant Species. Genome Biology and Evolution **4**: 230–239.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D., and Lynch, M. (2010). The Rate and Molecular Spectrum of Spontaneous Mutations in *Arabidopsis thaliana*. Science **327**: 92–94.

Pegler, J.L., Oultram, J.M.J., Mann, C.W.G., Carroll, B.J., Grof, C.P.L., and Eamens, A.L. (2023). Miniature Inverted-Repeat Transposable Elements: Small DNA Transposons That Have Contributed to Plant MICRORNA Gene Evolution. Plants **12**: 1101.

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. (2006). A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev. **20**: 3407–3425.

Rice, P. EMBOSS: The European Molecular Biology Open Software Suite.

Roux, C., Castric, V., Pauwels, M., Wright, S.I., Saumitou-Laprade, P., and Vekemans, X. (2011). Does Speciation between Arabidopsis halleri and Arabidopsis lyrata Coincide with Major Changes in a Molecular Target of Adaptation? PLoS ONE **6**: e26872.

Roux, J., Gonzàlez-Porta, M., and Robinson-Rechavi, M. (2012). Comparative analysis of human and mouse expression data illuminates tissue-specific evolutionary patterns of miRNAs. Nucleic Acids Research **40**: 5890–5900.

Shumate, A., Wong, B., Pertea, G., and Pertea, M. (2022). Improved transcriptome assembly using a hybrid of long and short reads with StringTie. PLoS Comput Biol **18**: e1009730.

Slater, G. and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics **6**: 31.

Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol **22**: 96–118.

Smith, L.M., Burbano, H.A., Wang, X., Fitz, J., Wang, G., Ural-Blimke, Y., and Weigel, D. (2015). Rapid divergence and high diversity of miRNAs and miRNA targets in the Camelineae. Plant J **81**: 597–610.

Srivastava, P.K., Moturu, T.R., Pandey, P., Baldwin, I.T., and Pandey, S.P. (2014). A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. BMC Genomics **15**: 348.

Takou, M., Hämälä, T., Koch, E.M., Steige, K.A., Dittberner, H., Yant, L., Genete, M., Sunyaev, S., Castric, V., Vekemans, X., Savolainen, O., and Meaux, J.D. (2021). Maintenance of Adaptive Dynamics and No Detectable Load in a Range-Edge Outcrossing Plant Population. Molecular Biology and Evolution **38**: 1820–1836.

Van Oss, S.B. and Carvunis, A.-R. (2019). De novo gene birth. PLoS Genet **15**: e1008160.

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. **27**: 737–746.

Voinnet, O. (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. Cell **136**: 669–687.

Wagner, A. (2011). The molecular origins of evolutionary innovations. Trends in Genetics **27**: 397–410.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. Genomics, Proteomics & Bioinformatics **8**: 77–80.

Ward, N. and Moreno-Hagelsieb, G. (2014). Quickly Finding Orthologs as Reciprocal Best Hits with BLAT, LAST, and UBLAST: How Much Do We Miss? PLoS ONE **9**: e101850.

Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2018). BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. Molecular Biology and Evolution **35**: 543–548.

Wong, G.Y. and Millar, A.A. (2023). Target Landscape of Conserved Plant MicroRNAs and the Complexities of Their Ancient MicroRNA-Binding Sites. Plant Cell Physiol **64**:604-621.

Wen, M., Lin, X., Xie, M., Wang, Y., Shen, X., Liufu, Z., Wu, C.-I., Shi, S., and Tang, T. (2016). Small RNA transcriptomes of mangroves evolve adaptively in extreme environments. Sci Rep **6**: 27551.

Wilson, B.A., Foy, S.G., Neme, R., and Masel, J. (2017). Young genes are highly disordered as predicted by the preadaptation hypothesis of de novo gene birth. Nat Ecol Evol **1**: 0146.

Wright, C., Rajpurohit, A., Burke, E.E., Williams, C., Collado-Torres, L., Kimos, M., Brandon, N.J., Cross, A.J., Jaffe, A.E., Weinberger, D.R., and Shin, J.H. (2019). Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. BMC Genomics **20**: 513.

Zhan, J. and Meyers, B.C. (2023). Plant Small RNAs: Their Biogenesis, Regulatory Roles, and Functions. Annu. Rev. Plant Biol. **74**: 21–51.

6. Supplementary data

Table S1. Comparison of nanopore readset statistics.

	Full	Filtlong	Longest
Number of Bases (GB)	29	7.2	7.2
Coverage	122	30	30
Number of Reads	3 322 114	230 775	173 520
N50	18 929	33 132	40 619
Average Size	8 854	31 199	41 494

Table S2. Assembly statistics of the reference accession (Auby1) throughout the process.

	Necat	Polished	Haplomerger2	Chromosome Scale
Total length (Mb)	323	323	227	227
Number of scaffolds/contigs	509	518	284	175
N50 (Kb)	1 597	1 600	3 347	25 922
L50	57	57	20	5
Average contig size (Kb)	634	625	800	1 298
Merqury score	24.4648	31.8868	32.1278	32.7199
Complete universal single-copy orthologs	C:98.4% S:59.4% D:39.0%	C:99.3% S:57.9% D:41.4%	C:99.0% S:97.0% D:2.0%	C:99.1% S:97.2% D:1.9%
Fragmented universal single-copy orthologs	0.6%	0.2	0.3%	0.2%
Missing universal single-copy orthologs	1.0%	0.5%	0.7%	0.7%

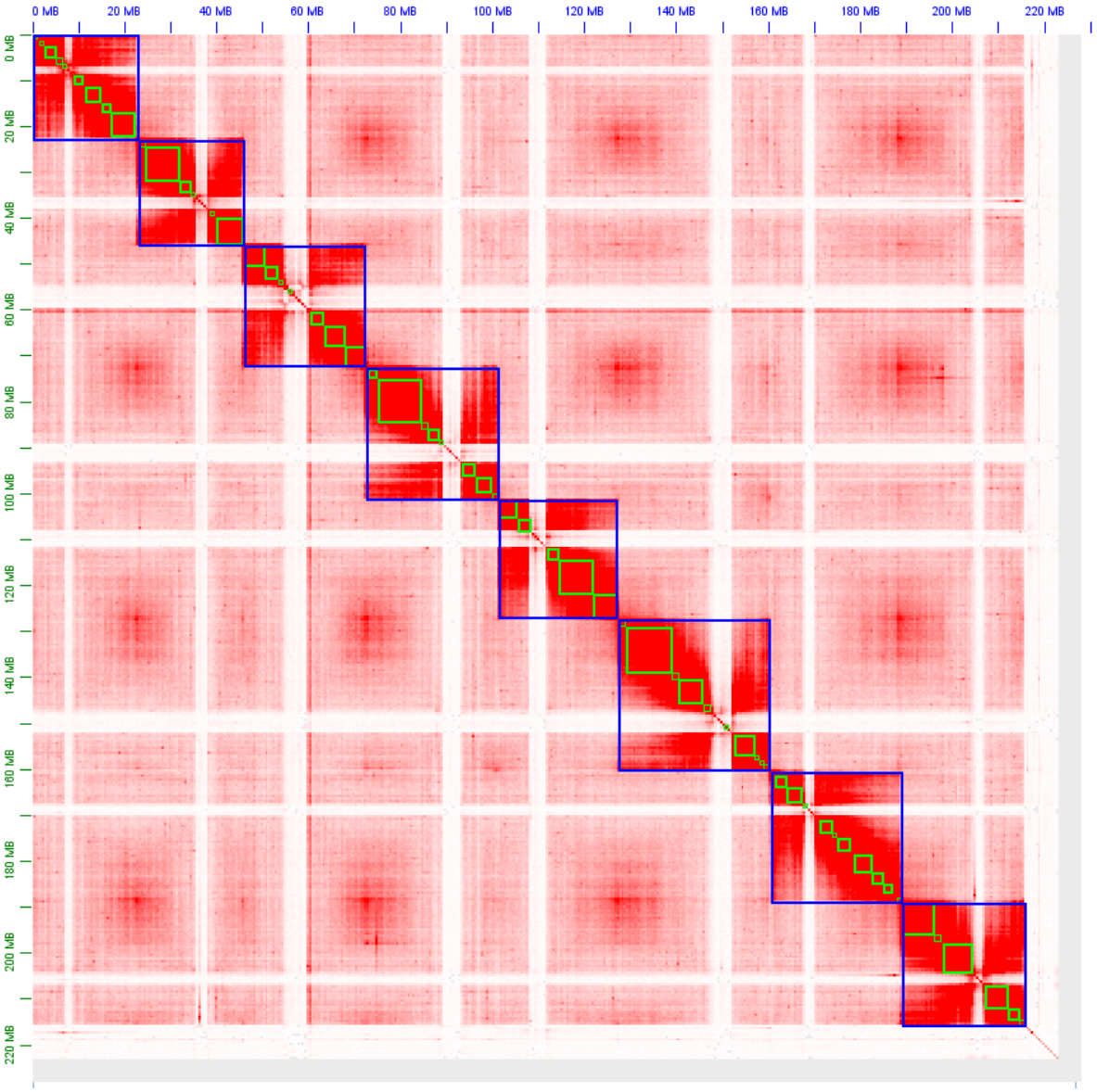


Figure S1. Curated chromosome-scale assembly of a reference *A. halleri* accession (Auby-1). The red dots correspond to Hi-C contacts. The green squares correspond to contigs from the PACBIO assembly, and are assembled into chromosome-level scaffolds represented by blue squares.

Table S3. Comparison of *A. halleri* PL22 and Auby1 genome assemblies.

Genome assembly metrics	*A. halleri* PL22 (Legrand et al. 2019)	*A. halleri* Auby1 (this study)
Number of contigs/ scaffolds	3152	175
Total length (Mb)	174	227
N50	279,389	25,922,902
L50	177	5
Longest contig/scaffold (Mb)	1.5	31
Complete universal single-copy orthologs	95.3%	99.1%
Fragmented universal single-copy orthologs	1.5%	0.2%
Missing universal single-copy orthologs	3.2%	0.7%

Table S4. sRNAseq datasets for miRNA predictions.

Species	Accession	Tissue	Library preparation	Sequencing technology	Total reads[a]	Number of miRNA genes[b]	Reference	SRA-NCBI
A.halleri	Auby1	Leaves	Nextflex® Small RNA-Seq	Illumina	206,983,903	196	This study	NA
A.halleri	Auby1	Buds	Nextflex® Small RNA-Seq	Illumina	159,726,202	267	This study	NA
A.halleri	Auby	Roots	Nextflex® Small RNA-Seq	Illumina	41,470,720	80	This study	NA
A.halleri	Auby	Buds	Nextflex® Small RNA-Seq	Illumina	40,304,810	121	This study	NA
A.halleri	Auby	Leaves	Nextflex® Small RNA-Seq	Illumina	40,236,011	72	This study	NA
A.halleri	I9	Roots	Nextflex® Small RNA-Seq	Illumina	40,564,542	71	This study	NA
A.halleri	I9	Buds	Nextflex® Small RNA-Seq	Illumina	39,370,067	100	This study	NA
A.halleri	I9	Leaves	Nextflex® Small RNA-Seq	Illumina	41,102,882	84	This study	NA
A.halleri	PL22	Leave	Nextflex® Small RNA-Seq	Illumina	9,403,700	124	This study	NA
A.halleri	I30	Leaves	Nextflex® Small RNA-Seq	Illumina	13,304,075	109	This study	NA
A.halleri	HF70	Buds	SOLiD Total RNA-Seq	SOLiD	31,926,829	74	Durand et al. (2014)	SRR1271746
A.halleri	I5	Buds	SOLiD Total RNA-Seq	SOLiD	27,201,005	91	Durand et al. (2014)	SRR1271747
A.halleri	I5	Buds	SOLiD Total RNA-Seq	SOLiD	28,166,676	96	Durand et al. (2014)	SRR1271748
A.halleri	I5	Buds	SOLiD Total RNA-Seq	SOLiD	27,521,772	79	Durand et al. (2014)	SRR1271749
A.halleri	I9	Buds	SOLiD Total RNA-Seq	SOLiD	27,760,235	77	Durand et al. (2014)	SRR1271750
A.halleri	Nivelle	Buds	SOLiD Total RNA-Seq	SOLiD	24,590,979	82	Durand et al. (2014)	SRR1271751
A.halleri	Nivelle	Buds	SOLiD Total RNA-Seq	SOLiD	26,216,099	69	Durand et al. (2014)	SRR1271752
A.halleri	Nivelle	Buds	SOLiD Total RNA-Seq	SOLiD	30,136,006	77	Durand et al. (2014)	SRR1271753
A.halleri	BC01	Buds	ION total RNA-seq	Proton	23,441,621	99	Durand et al. (2014)	SRR1271755
A.halleri	BC02	Buds	ION total RNA-seq	Proton	21,907,916	98	Durand et al. (2014)	SRR1271756
A.halleri	BC03	Buds	ION total RNA-seq	Proton	23,779,085	93	Durand et al. (2014)	SRR1271757
A.lyrata	MN47	Leaves	TruSeq Small RNA	Illumina	9,012,391	102	Legrand et al. (2019)	SRR9665460
A.lyrata	CP99	Leaves	NA	Illumina	28,995,456	95	This study	NA
A.lyrata	CP99	Buds	NA	Illumina	10,424,454	73	This study	NA
A.lyrata	AI14	Buds	SOLiD Total RNA-Seq	SOLiD	36,006,064	69	Durand et al. (2014)	SRR1271754
A.lyrata	MN47	Leaves	Ma et al. (2010)	Illumina	2,012,409	46	Ma et al. (2010)	SRR034856
A.lyrata	MN47	Buds	SOLiD Small RNA Expression	SOLiD	90,518,311	126	Ma et al. (2010)	SRR040401
A.lyrata	MN47	Buds	SOLiD Small RNA Expression	SOLiD	10,321,920	67	Ma et al. (2010)	SRR040402
A.lyrata	MN47	Leaves	Fahlgren et al. (2009)	Illumina	5,093,642	62	Fahlgren et al. (2010)	SRR051926
A.lyrata	MN47	Buds	Fahlgren et al. (2010)	Illumina	4,876,824	61	Fahlgren et al. (2010)	GSM518430[c]
A.lyrata	MN47	Buds	Fahlgren et al. (2010)	Illumina	4,229,395	64	Fahlgren et al. (2010)	GSM518431[c]
A.lyrata	Plech	Roots	Nextflex® Small RNA-Seq	Illumina	43,024,185	64	This study	NA
A.lyrata	Plech	Buds	Nextflex® Small RNA-Seq	Illumina	41,899,496	128	This study	NA
A.lyrata	Plech	Leaves	Nextflex® Small RNA-Seq	Illumina	36,480,074	110	This study	NA

[a] Total number of reads in the sequencing experiment.

[b] Number of miRNA genes predicted in the sequencing experiment.

[c] Data in gff3 format while all the others are in fastq format.

Table S5. *Arabidopsis halleri* predicted miRNA genes.

The colored cases represent the score of miRkwood prediction (from 1 to 6) (Guigon et al., 2019). The letter “S” indicates that the miRNA has only been predicted by Shortstack 4.0.2 (Johnson et al., 2016).

Loci	strand	Auby1_lea	Auby1_bud	TOT1	TOT2	TOT4	TOT5	TOT7	TOT8	PL22	I30	BC01	BC02	BC03	Niv001	Niv045	Niv060	HF70	I5_53	I5_67	I5_77	I9_47	miRNA lenght	5' nt	AGO1 enrichment	AGO4 enrichment	Main loading	Family	Conservation
ddAraHall_1.3_1:531924-532073	+	6	6	6	6	.	6	6	6	6	6	6	.	6	6	6	6	6	6	6	5	6	20, 21	U, U	36,29	0,08	AGO1	miR166	conserved
ddAraHall_1.3_2:16126383-16126515	-	6	6	.	6	.	6	6	6	6	6	6	6	6	.	.	6	6	6	6	.	.	.	21, 20	U, U	19,00	0,17	AGO1	miR156	conserved
ddAraHall_1.3_2:16139528-16139662	+	6	6	.	6	.	6	6	6	6	6	6	6	6	.	.	6	6	6	6	.	.	.	21, 20	U, U	19,00	0,17	AGO1	miR156	conserved
ddAraHall_1.3_3:20338645-20338815	-	6	.	.	.	6	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	U, G	NA	NA	NA	miR396	conserved
ddAraHall_1.3_7:6671318-6671491	-	6	6	6	6	6	6	.	6	6	6	6	6	6	6	5	6	5	6	6	6	5	21	U	NA	NA	NA	miR396	conserved
ddAraHall_1.3_7:15280713-15280867	+	6	6	6	6	6	6	.	6	6	6	6	6	6	6	6	6	6	6	6	6	6	21, 21	U, C	1,29	0,10	AGO1	miR168	conserved
ddAraHall_1.3_8:1965495-1965627	+	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	21	U	2,30	0,03	AGO1	miR168	conserved
ddAraHall_1.3_3:11214767-11214913	+	6	6	6	6	6	6	6	6	.	.	6	6	6	6	6	6	6	6	6	6	6	19, 21, 21	U, U, G	6,12	0,05	AGO1	miR167	conserved
ddAraHall_1.3_2:20688789-20689010	-	6	6	6	.	.	.	6	.	6	6	6	6	6	6	6	6	6	6	6	6	6	21	U	NA	NA	NA	miR159	conserved
ddAraHall_1.3_6:108992-109193	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	U, C	40,81	0,04	AGO1	miR164	conserved
ddAraHall_1.3_4:20385307-20385423	+	6	6	6	6	6	.	.	6	6	6	6	.	6	6	6	6	6	6	6	6	6	21, 21	A, C	1,25	0,30	AGO1	miR390	conserved
ddAraHall_1.3_8:17714927-17715063	+	6	6	6	6	6	6	6	6	.	6	6	6	6	.	6	.	.	6	6	6	6	21, 20	A, C	0,33	0,30	AGO1 AGO4	miR390	conserved
ddAraHall_1.3_6:1525896-1526024	-	6	6	6	6	6	6	6	6	6	6	6	6	6	.	.	.	.	.	.	.	.	21, 21	A, G	NA	NA	NA	miR172	conserved
ddAraHall_1.3_6:3545660-3545941	-	5	5	5	5	5	.	.	5	4	4	5	5	5	5	5	5	5	5	5	5	5	21, 21	U, G	NA	NA	NA	miR162	conserved
ddAraHall_1.3_6:10419569-10419710	-	.	6	6	6	6	.	.	6	6	6	6	6	6	5	6	5	5	6	6	6	5	21, 21	U, G	NA	NA	NA	miR162	conserved
ddAraHall_1.3_8:4338321-4338451	-	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	.	6	21, 21, 21	G, U, U	23,14	6,82	AGO1	miR166	conserved
ddAraHall_1.3_3:8395124-8395388	-	5	5	.	5	5	5	5	5	5	5	5	5	.	5	5	5	5	5	.	5	5	21, 20	U, U	19,00	0,17	AGO1	miR156	conserved
ddAraHall_1.3_4:25662470-25662588	+	6	6	6	6	6	6	6	6	.	.	6	6	6	.	.	.	.	6	.	6	6	21, 21	U, C	21,00	0,25	AGO1	miR164	conserved
ddAraHall_1.3_8:18474093-18474252	+	6	6	6	6	.	6	.	6	6	6	6	6	.	6	6	6	6	6	6	6	6	21, 21, 20	G, G, G	3,71	0,20	AGO1	miR172	conserved
ddAraHall_1.3_2:21852196-21852375	+	.	6	6	6	.	.	5	5	5	4	6	6	.	6	6	6	6	6	6	5	6	20, 21	U, U	9,19	0,05	AGO1	miR394	conserved
ddAraHall_1.3_7:17889326-17889560	+	6	6	6	6	6	6	.	6	6	5	6	6	6	5	5	5	4	5	6	5	5	21, 21	U, A	5,95	0,03	AGO1	miR319	conserved
ddAraHall_1.3_4:25534171-25534306	+	6	6	.	.	6	6	6	6	6	6	.	.	.	.	.	.	6	.	.	.	.	21	U	25,78	0,09	AGO1	miR403	conserved
ddAraHall_1.3_1:8284259-8284481	+	6	6	6	6	6	6	6	6	.	.	6	6	6	.	.	.	.	.	.	.	.	21, 20, 21	U, U, G	16,96	0,07	AGO1	miR159	conserved
ddAraHall_1.3_5:6058023-6058287	-	5	5	.	.	5	.	.	.	.	4	5	.	.	.	.	.	.	.	.	.	.	21, 21	U, U	3,52	0,01	AGO1	miR171	conserved
ddAraHall_1.3_1:9312053-9312230	+	5	6	6	5	.	5	5	5	5	4	6	.	6	6	5	5	6	5	5	5	5	20, 21	U, U	9,19	0,05	AGO1	miR394	conserved
ddAraHall_1.3_5:1746981-1747093	+	.	6	.	.	.	.	.	.	6	6	6	6	6	.	.	.	.	.	.	.	6	21	U	19,11	0,02	AGO1	miR827	conserved
ddAraHall_1.3_1:15438744-15439051	+	6	6	6	6	.	6	6	6	6	6	.	.	6	5	6	6	6	6	6	5	6	20, 22, 19, 21	U, U, U, A	9,16	0,21	AGO1	miR167	conserved
ddAraHall_1.3_5:3936445-3936577	-	.	6	6	.	6	6	.	6	.	.	6	6	6	.	.	.	.	.	.	.	.	21, 21, 20	A, G, A	NA	NA	NA	miR172	conserved
ddAraHall_1.3_4:25228506-25228750	+	6	6	6	6	6	.	.	6	6	6	6	.	5	.	.	.	.	.	.	.	.	21, 21, 21	G, G, U	15,44	4,69	AGO1	miR166	conserved
ddAraHall_1.3_5:689053-689215	-	6	6	6	6	6	.	.	6	6	6	6	.	6	.	.	.	.	2	2	5	.	21, 22, 21, 21	G, U, G, U	49,73	0,86	AGO1	miR166	conserved
ddAraHall_1.3_6:3816032-3816181	+	6	6	6	6	6	6	6	6	6	6	6	6	6	.	.	.	.	6	6	6	6	21, 21, 22, 21	G, C, U, U	24,98	0,53	AGO1	miR166	conserved
ddAraHall_1.3_6:3817863-3818003	+	6	6	6	6	6	6	6	6	6	6	6	6	6	.	.	.	.	6	6	6	6	21, 21, 21	G, C, U	10,47	0,17	AGO1	miR166	conserved
ddAraHall_1.3_6:23427222-23427395	-	5	6	6	6	6	5	.	.	5	.	.	.	.	5	5	5	5	5	5	5	5	21, 20	U, U	36,29	0,08	AGO1	miR166	conserved
ddAraHall_1.3_1:26688586-26688805	-	.	6	.	.	6	.	.	.	.	.	6	6	6	6	6	6	6	6	6	6	6	21, 20	C, U	20,94	0,17	AGO1	miR156	conserved
ddAraHall_1.3_7:14301742-14301905	+	6	6	6	6	6	6	.	6	6	6	6	6	6	.	.	.	.	6	6	.	.	21, 22	U, G	2,54	0,04	AGO1	miR160	conserved
ddAraHall_1.3_6:6251884-6252039	+	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21, 21, 21	G, U, A	2,58	0,11	AGO1	miR398	conserved
ddAraHall_1.3_7:246251-246423	+	6	6	6	6	6	6	6	6	6	6	6	6	6	.	.	.	.	6	.	.	.	21, 21, 22, 21	G, G, U, U	22,93	3,54	AGO1	miR166	conserved
ddAraHall_1.3_2:2103532-2103656	+	6	6	.	.	6	.	6	.	6	6	6	.	6	.	.	.	.	.	.	.	.	21, 21, 21	U, A, U	5,28	0,04	AGO1	miR171	conserved
ddAraHall_1.3_5:51359-51485	-	6	6	5	5	6	.	.	6	6	6	2	.	2	6	6	6	6	6	6	6	6	21	U	9,36	0,04	AGO1	miR167	conserved
ddAraHall_1.3_4:20867659-20867784	+	6	6	6	6	6	.	.	6	6	6	6	.	6	.	.	.	.	6	6	6	.	21, 21	G, U	2,53	0,04	AGO1	miR160	conserved
ddAraHall_1.3_6:11696509-11696686	+	6	6	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21, 22, 22	C, G, C	0,95	0,03	AGO1	miR169	conserved
ddAraHall_1.3_4:13262558-13262700	-	6	6	6	6	6	.	.	6	6	6	6	.	6	2	6	6	.	6	6	.	2	21, 21, 21	A, G, G	NA	NA	NA	miR172	conserved
ddAraHall_1.3_5:3784091-3784269	-	6	6	.	.	6	6	6	6	6	6	6	.	6	.	.	.	.	.	.	.	.	21	A	NA	NA	NA	miR393	conserved
ddAraHall_1.3_6:4624807-4624935	-	.	6	.	.	.	6	6	.	.	.	.	6	6	.	.	.	.	.	.	.	.	21, 24, 21, 20	U, G, G, U	19,05	0,11	AGO1	miR156	conserved
ddAraHall_1.3_8:698316-698570	+	.	.	5	5	5	5	.	.	.	2	.	.	.	.	.	.	.	5	.	.	.	21, 21, 21, 21	U, G, C, A	1,47	0,02	AGO1	miR160	conserved
ddAraHall_1.3_3:4678299-4678441	-	.	6	6	6	.	6	.	6	.	6	6	6	6	.	.	.	.	.	.	.	.	21, 20, 21	A, A, G	NA	NA	NA	miR172	conserved
ddAraHall_1.3_6:6255692-6255835	+	6	6	.	.	6	.	6	.	6	6	6	.	.	6	.	.	.	6	6	.	.	21, 21	U, G	3,58	0,03	AGO1	miR398	conserved
ddAraHall_1.3_8:22559784-22559886	+	.	6	.	.	6	6	6	.	6	6	.	.	.	.	.	.	.	.	.	.	.	21, 21, 21	U, U, U	5,39	0,33	AGO1	miR171	conserved
ddAraHall_1.3_7:553128-553337	+	.	6	6	6	6	6	6	6	6	.	6	6	6	.	.	.	.	.	.	.	.	21, 21, 21	U, C, A	6,22	0,78	AGO1	miR319	conserved
ddAraHall_1.3_8:19783913-19784046	+	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	.	.	.	21, 21, 21	U, G, U	8,24	0,03	AGO1	miR399	conserved
ddAraHall_1.3_6:12870613-12870944	+	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	.	.	.	.	.	.	22, 20, 21	U, U, U	23,80	0,07	AGO1	miR156	conserved
ddAraHall_1.3_7:16460067-16460288	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	.	21, 21, 20	U, G, G	1,86	0,02	AGO1	miR169	conserved
ddAraHall_1.3_3:6325396-6325579	-	5	.	6	2	.	.	6	.	4	5	.	.	.	.	.	.	.	.	.	.	.	21, 21, 20, 23, 21	U, G, G, G, U	2,98	0,01	AGO1	miR169	conserved
ddAraHall_1.3_7:2182470-2182798	+	.	6	.	.	.	.	.	.	.	.	6	6	6	.	6	6	6	6	.	6	6	21, 23	C, C	2,17	0,00	AGO1	miR169	conserved
ddAraHall_1.3_8:20549875-20550002	+	6	6	.	6	6	6	6	.	6	6	6	6	6	.	.	.	.	.	.	.	.	22, 21, 21	U, C, U	25,60	0,08	AGO1	miR166	conserved
ddAraHall_1.3_1:30319903-30320022	+	.	6	6	.	.	6	.	.	.	.	6	6	6	.	.	.	.	.	.	.	.	21, 21, 20	G, U, G	1,86	0,02	AGO1	miR169	conserved
ddAraHall_1.3_1:5199527-5199652	-	.	.	.	.	6	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	A, U	9,41	0,00	AGO1	miR171	conserved
ddAraHall_1.3_2:1450500-1450652	+	.	.	.	.	6	.	.	.	.	.	6	6	6	.	.	.	.	.	.	.	.	21	U	16,47	0,06	AGO1	miR399	conserved
ddAraHall_1.3_8:16073540-16073631	-	6	6	.	.	.	6	.	.	.	6	6	6	.	.	6	.	.	6	6	6	.	20, 21, 21	U, U, G	NA	NA	NA	miR156	conserved
ddAraHall_1.3_4:21950709-21950928	+	.	6	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21, 21, 21	U, U, U	10,62	0,02	AGO1	miR319	conserved
ddAraHall_1.3_4:24974756-24974973	+	5	5	5	5	5	.	.	5	5	5	6	.	5	.	.	.	.	.	.	.	.	19, 20, 18, 22	U, U, U, U	19,36	0,21	AGO1	miR159	conserved
ddAraHall_1.3_6:13544614-13544739	-	.	.	.	.	.	.	.	.	6	6	.	.	.	.	.	.	.	.	.	.	.	21, 20, 21	U, U, C	46,99	0,05	AGO1	miR164	conserved
ddAraHall_1.3_3:5717413-5717676	-	.	.	.	.	.	.	.	.	6	.	6	6	.	.	6	6	.	6	6	6	.	21	C	NA	NA	NA	miR169	conserved
ddAraHall_1.3_7:21825024-21825218	+	.	6	5	5	6	.	.	5	5	6	6	6	.	.	.	.	.	.	.	.	.	23	U	100,87	0,19	AGO1	miR156	conserved
ddAraHall_1.3_1:13882285-13882422	-	.	6	.	.	.	6	.	.	.	.	.	.	6	6	6	6	6	6	6	.	.	22, 21	G, U	NA	NA	NA	miR399	conserved
ddAraHall_1.3_2:18521137-18521255	-	6	6	6	.	6	6	6	6	.	.	.	.	.	.	.	.	.	.	.	6	.	21	C	NA	NA	NA	miR395	conserved
ddAraHall_1.3_2:18537085-18537199	-	.	6	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21	C	NA	NA	NA	miR395	conserved
ddAraHall_1.3_7:22291869-22292133	-	5	5	.	.	5	.	.	5	5	5	.	.	.	.	.	.	.	.	.	.	.	22, 23, 20	G, U, U	19,05	0,11	AGO1	miR156	conserved
ddAraHall_1.3_3:1669685-1669873	-	.	.	6	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23, 21	U, U	NA	NA	NA	miR167	conserved
ddAraHall_1.3_1:13875892-13876044	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	6	6	6	6	.	.	21, 22	U, G	NA	NA	NA	miR399	conserved
ddAraHall_1.3_5:20919162-20919320	-	6	6	.	.	5	.	.	.	5	5	.	.	5	.	.	.	.	.	5	.	.	19, 20	U, U	11,74	0,12	AGO1	miR398	conserved
ddAraHall_1.3_4:21370432-21370617	+	6	6	.	.	6	.	.	.	6	6	.	.	6	.	.	.	.	.	.	.	.	21, 22	A, U	9,65	0,49	AGO1	miR393	conserved
ddAraHall_1.3_2:18518151-18518392	-	5	5	.	.	5	5	.	5	.	.	.	.	.	.	.	.	.	.	.	5	.	21	C	NA	NA	NA	miR395	conserved
ddAraHall_1.3_4:3820042-3820256	-	5	.	.	.	5	.	5	.	4	4	.	.	.	.	.	.	.	.	.	.	.	22, 20	G, U	19,05	0,11	AGO1	miR156	conserved
ddAraHall_1.3_5:19654562-19654789	+	.	.	6	.	.	.	.	.	.	.	.	.	.	.	6	.	.	6	6	.	.	19, 21	G, U	AGO1	0,00	AGO1	miR169	conserved
ddAraHall_1.3_5:19658493-19658751	+	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	.	.	19, 21	G, U	AGO1	0,00	AGO1	miR169	conserved
ddAraHall_1.3_1:11644703-11644847	-	6	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21	C	19,55	0,02	AGO1	miR395	conserved
ddAraHall_1.3_2:18537909-18538042	+	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21	C	19,55	0,02	AGO1	miR395	conserved
ddAraHall_1.3_8:18878225-18878338	+	.	.	.	.	.	.	.	.	6	6	.	.	.	.	.	.	.	.	.	.	.	21, 21	A, U	7,88	1,97	AGO1	miR391	conserved
ddAraHall_1.3_1:5468525-5468729	-	2	2	.	.	.	2	2	6	6	6	.	.	.	.	4	2	.	5	5	.	.	21, 21	A, U	NA	NA	NA	miR482	conserved
ddAraHall_1.3_5:19658876-19659078	+	.	.	6	.	.	.	.	.	3	.	.	.	.	.	6	.	.	.	5	.	.	19, 21	G, U	AGO1	0,00	AGO1	miR169	conserved
ddAraHall_1.3_4:17953858-17954104	-	.	5	.	.	.	.	.	.	.	.	.	.	5	.	.	.	.	5	5	.	.	21	U	NA	NA	NA	miR399	conserved

ddAraHall_1.3_2:18522313-18522447	+	6	21	C	19,55	0,02	AGO1	miR395	conserved
ddAraHall_1.3_1:30316894-30317047	-	. . . 2 . . . . 4 4 6 6	21	U	5,96	0,00	AGO1	miR169	conserved
ddAraHall_1.3_6:5186629-5186771	+	. . . . S	20	U	19,05	0,11	AGO1	miR156	conserved
ddAraHall_1.3_6:25702347-25702474	+	6 5	21	U	9,45	0,00	AGO1	miR397	conserved
ddAraHall_1.3_7:10240811-10240946	-	. 6	21	U	9,45	0,00	AGO1	miR397	conserved
ddAraHall_1.3_1:8859994-8860140	-	. . . . S	21	U	AGO1	NA	AGO1	miR169	conserved
ddAraHall_1.3_5:19646831-19647021	+	. . . . S	21	U	AGO1	0,00	AGO1	miR169	conserved
ddAraHall_1.3_5:19647167-19647335	+	. . . . S	21	U	AGO1	0,00	AGO1	miR169	conserved
ddAraHall_1.3_1:11667878-11667983	+	. . . . . 5	21	C	3,21	AGO4	AGO1 and AGO4	miR395	conserved
ddAraHall_1.3_6:530252-530370	+	. . . . 5 .	24	A	0,00	0,00	not loaded	miR2111	conserved
ddAraHall_1.3_3:4367856-4367999	-	6 6 6 6 6 6 6 6 6 6 6 6 . 6 6 6 6 6 6 6 6	20	U	NA	NA	NA	miR158	brass
ddAraHall_1.3_1:31015665-31015767	+	. S	19, 20	U, U	16,96	0,09	AGO1	miR158	brass
ddAraHall_1.3_7:17687982-17688164	-	. . . . 6 . 6	21	C	NA	NA	NA	miR845	brass
ddAraHall_1.3_7:17685025-17685131	-	. 6 . . 6 6 . . . . 6 6 6 6 6 6 6 6 . 6 .	22, 21	U, U	NA	NA	NA	miR845	brass
ddAraHall_1.3_3:857732-858125	-	5 4 . . . . . . 2 6 . . . . 6 . .	21, 21, 21	U, A, A	NA	NA	NA	miR3435	brass
ddAraHall_1.3_1:16467359-16467532	+	6 6 . . . . . . 6 5	21, 21	G, U	15,23	0,15	AGO1	miR400	brass
ddAraHall_1.3_6:12771979-12772190	+	. 2 . . . . . . 6 6 . . . . . . 5 . . . 5	21, 24	U, G	4,32	0,00	AGO1	miR860	brass
ddAraHall_1.3_1:16508440-16508548	-	6 6	21	U	NA	NA	NA	miR5654	brass
ddAraHall_1.3_1:16508697-16508959	-	. 5	21	U	24,29	0,24	AGO1	miR400	brass
ddAraHall_1.3_4:25408860-25408988	+	. . . . S . . .	21	C	0,09	0,05	AGO1	new	brass
ddAraHall_1.3_1:26384433-26384562	+	. . . 6 6 6 . 6 . . 6	21, 21, 21, 21	U, C, U, C	4,88	0,03	AGO1	miR161	brass
ddAraHall_1.3_4:6041153-6041300	+	. . . . 6 6 . .	22, 21	U, U	7,32	0,04	AGO1	miR825	brass
ddAraHall_1.3_3:11430855-11430981	+	6 6 6 6 6 6 6 6 6 6 6 6 6	22	U	11,67	0,05	AGO1	miR173	brass
ddAraHall_1.3_2:16048600-16048993	+	5 5 . . . . 5 . . . 6 . 6 . . .	23, 21, 21	U, A, U	4,17	0,23	AGO1	miR163	ath
ddAraHall_1.3_2:2755301-2755435	-	. . . . 6 . . . . 6 6 . 6 6	21, 21, 22, 21, 21	C, U, C, C, U	NA	NA	NA	miR842	ath
ddAraHall_1.3_7:19145542-19145827	+	6 . . . 5 5 . . . . 2 5 6 5 5 . 5 .	22, 22, 21	U, A, U	17,20	0,04	AGO1	new	ath
ddAraHall_1.3_3:5925219-5925476	-	. . 6 . 6 . 6 . 6 6	21, 21	U, C	NA	NA	NA	miR823	ath
ddAraHall_1.3_8:13501763-13502074	+	. . . . 5 5 . 6	21, 24, 21	A, A, U	1,93	32,35	AGO4	miR4245	ath
ddAraHall_1.3_2:2754822-2755128	-	2 2 5 6 . 2 5 6 4 . . 5	21	U	NA	NA	NA	new	ath
ddAraHall_1.3_6:1161174-1161572	-	. 2 5 5 5 . . . 4 . . . . 2 5 . .	22, 21	U, G	NA	NA	NA	miR822	ath
ddAraHall_1.3_4:21149863-21150156	-	. . . . 4 5 . . . . . . 5	25	G	0,87	0,51	AGO1	new	ath
ddAraHall_1.3_3:8738281-8738456	+	. 4 . . . . 4 4 . 5 . . 3 . 4 . .	21, 21	U, U	25,94	0,00	AGO1	new	ath
ddAraHall_1.3_1:283422-283605	+	. . . . 6 6	21, 21	C, C	1,26	0,00	AGO1	miR2112	ath
ddAraHall_1.3_1:2326910-2327287	-	5 2 . . . . . . 2 2	24	A	NA	NA	NA	miR5640	ath
ddAraHall_1.3_5:20099087-20099225	-	6 6 6 . . . . 6 5 5 6 6 6 . 6 6 6 6 6 6 6	21, 21	U, G	NA	NA	NA	miR3443	shared
ddAraHall_1.3_3:8338133-8338388	+	. 6 6 6 6 6 6 6 6 6 5 5 . 6 6 6 5 6 5 6 6	22, 22	U, U	8,80	0,06	AGO1	miR3440	shared
ddAraHall_1.3_2:3648545-3648763	-	. 6 . . 6 6 . . . . 6 6 6 6 . 5 . 6 6 5 6	21	U	NA	NA	NA	miR859	shared
ddAraHall_1.3_1:25715149-25715285	-	. . . . 6 6 6 6 6 . . . . 6 6 . 6	21	C	NA	NA	NA	miR3436	shared
ddAraHall_1.3_2:22450520-22450793	+	6 6 6 6 6 6 6 6 6 6 . . . . . . 6 . 6 . .	22, 21	U, A	13,75	0,09	AGO1	miR402	shared
ddAraHall_1.3_2:22984689-22984796	+	. 6 . . . . . . 6 . . . 6 . . 6 . . . 6 6	21, 21, 21	U, G, G	27,53	0,85	AGO1	miR3433	shared
ddAraHall_1.3_7:15457376-15457593	-	. 6 . . . . 6 6 6 6 6 6 6 . . 6 6	22, 21	A, G	NA	NA	NA	miR4240	shared
ddAraHall_1.3_7:22259415-22259588	+	6 6 . . 6 6 . . 2 6 2 6 5 . . . 6 5 5 . .	22, 21, 21	U, C, A	1,67	0,02	AGO1	miR3448	shared
ddAraHall_1.3_4:1716859-1717207	+	. . . . 4 . . . . 5	21	U	8,19	0,00	AGO1	new	shared
ddAraHall_1.3_5:16411558-16411950	-	. . . . 5 5	24	G	0,00	5,08	AGO4	new	shared
ddAraHall_1.3_2:16929416-16929712	-	2 . . . 5 . . 5 . 5 . . . 5 2 5 5 5 6 5 .	21, 22, 21	A, U, U	NA	NA	NA	miRN127	shared
ddAraHall_1.3_4:14032767-14032914	-	. . . . 6	21, 21, 21	G, A, U	NA	NA	NA	miR3447	shared
ddAraHall_1.3_4:19891296-19891581	+	5 5 5 5 5 . . 5 2 6 2 . 2 2 . . . 2 . . 5	24, 22, 22	U, U, U	1,12	0,08	AGO1	miR3439	shared
ddAraHall_1.3_4:1759037-1759321	-	4 5 5 . 2 . 5 . 4 4	22, 21	A, U	NA	NA	NA	new	shared
ddAraHall_1.3_5:21286195-21286504	+	2 . 5 5 . . . 5 2 2 5 6 5 . . 2 2 . 3 . 2	22, 21	U, U	12,09	0,03	AGO1	miR840	shared
ddAraHall_1.3_2:3648379-3648485	-	6 6 . . . 6 . . . . . 6 . . . . . 2 . . 6	21	U	NA	NA	NA	new	shared
ddAraHall_1.3_6:13684571-13684753	+	. . . . 4 . . . . . . 5 6 5 . 6 5 . 5 . .	21	C	35,13	0,00	AGO1	new	shared
ddAraHall_1.3_3:4393513-4393665	-	5 2 . . 4 . 5 . 6 5 6	21, 21	U, U	NA	NA	NA	new	shared
ddAraHall_1.3_4:3981747-3982272	+	. . . .	23, 24, 18, 24	A, C, A, A	0,09	24,55	AGO4	new	shared
ddAraHall_1.3_5:4199479-4199610	-	5 5 . . . . . . 6 6 . . . . 6 2 .	21, 21, 21	A, A, U	3,15	0,00	AGO1	new	shared
ddAraHall_1.3_5:20540076-20540187	-	. 2 . 5 . . . 5 . . . . 6 . 6 5	21	C	0,01	0,00	AGO1	new	shared
ddAraHall_1.3_7:11702654-11702827	+	5 5 . . 5 5 . . 5	21, 20	U, U	35,37	0,10	AGO1	miRN153	shared
ddAraHall_1.3_5:9407103-9407260	+	6 6 . . 6 . . . 6	21, 22	G, U	6,83	0,00	AGO1	miR4222	shared
ddAraHall_1.3_3:7288744-7288890	-	. . . . 5	21, 21	C, C	0,00	0,00	not loaded	new	shared
ddAraHall_1.3_5:9408474-9408690	+	6 6 . . . . . . 6 6	21, 22	A, U	94,52	0,00	AGO1	new	shared
ddAraHall_1.3_3:7289489-7289616	-	. 5 . . 6 6	21, 21	G, U	NA	NA	NA	miR2934	shared
ddAraHall_1.3_6:10174779-10175099	-	4 4 . . 2 . . 2 . 4 4 4 4 4 . 4 2 . 5 . .	21, 21	C, U	NA	NA	NA	miR3434	shared
ddAraHall_1.3_2:21092760-21093058	-	2 . . . . 6 2 . . . . 6 .	21, 21	C, G	NA	NA	NA	new	shared
ddAraHall_1.3_3:11617365-11617749	+	1 1 . . . 5 . . 5 5	22, 21	U, G	13,04	0,00	AGO1	new	shared
ddAraHall_1.3_2:3648215-3648321	-	5 5 . . . 5	21	C	NA	NA	NA	new	shared
ddAraHall_1.3_1:27221440-27221610	+	. . . . 6	21, 21, 21	G, A, A	0,00	0,00	not loaded	miR4225	shared
ddAraHall_1.3_1:25331112-25331257	+	6 . . . . 5	21	U	35,79	0,00	AGO1	miRN419	shared
ddAraHall_1.3_1:27770221-27770301	-	. 4 4 . . . . . 6 . . . . 2 1 2 . 2 2 2 .	21	U	NA	NA	NA	new	shared
ddAraHall_1.3_3:25543142-25543262	+	. . . . S	23	A	0,00	0,00	not loaded	new	shared
ddAraHall_1.3_5:19142974-19143373	-	. 5	24	A	NA	NA	NA	new	shared
ddAraHall_1.3_1:2428896-2429270	+	2 2 . . . . . . 5	21	A	0,01	0,00	AGO1	new	shared
ddAraHall_1.3_4:21172412-21172504	-	. 2 2 2 2 . . 2 2 2 2 . 2 . . 5	22	C	NA	NA	NA	new	shared
ddAraHall_1.3_6:5935179-5935557	-	. . . . 5 . . .	22, 22	U, U	NA	NA	NA	new	spe
ddAraHall_1.3_3:7131758-7131914	+	. 5 . . 5 5 . . . . 5 5 6 . . 5 5 4 5 5 5	22	U	7,36	0,00	AGO1	new	spe
ddAraHall_1.3_5:8031750-8031935	+	. 6 . . 6 . . . . 6 . 2 . 2	21, 23, 24	U, A, A	55,93	0,48	AGO1	new	spe
ddAraHall_1.3_6:1581376-1581679	+	6 6 2 . . . . . . 6 5 2 5 5 4 5 6 5 6 . .	21, 21	G, A	0,80	1,57	AGO4	new	spe
ddAraHall_1.3_7:696100-696320	-	. 6 . . . 6	21, 21	U, A	NA	NA	NA	new	spe
ddAraHall_1.3_8:18133488-18133641	+	. 6	21, 21, 21	A, U, A	8,34	0,03	AGO1	new	spe
ddAraHall_1.3_1:26496842-26496995	+	6 6 . . . 6 . . 6 6 . 6 . 2 2 2 . 6 5 . 2	22, 21, 21	A, A, A	0,87	0,00	AGO1	new	spe
ddAraHall_1.3_5:5187768-5187868	+	. 6 . . 5 . . . . . 5 . 5 . 5 . 5 . 6 . .	21, 21, 22, 21	U, A, U, G	5,39	0,05	AGO1	new	spe
ddAraHall_1.3_3:1181997-1182232	+	. . . . S S . . S S . . S . S . S	21	A	0,24	0,01	AGO1	new	spe
ddAraHall_1.3_6:1429141-1429476	+	. 5 2 . 2 . . 2 . . . . 2 5 2 2 2 . 5 . .	21, 21, 21, 21, 21	U, A, C, A, A	59,32	0,14	AGO1	new	spe
ddAraHall_1.3_8:16420568-16420731	+	2 2 . . . 2 . 6 2 2 . 2 . 6 . 5 5 6 5 . 5	21, 21	U, C	13,11	0,00	AGO1	new	spe
ddAraHall_1.3_2:8882974-8883276	-	. 5 . . . . 5 . 5 . 5 . . . 4 .	24	G	0,00	5,08	AGO4	new	spe
ddAraHall_1.3_7:17264581-17264667	-	. 2 . . 4 . . . . . 4 6 5 5 . 4 6 . . . 6	21, 24	C, C	NA	NA	NA	new	spe
ddAraHall_1.3_3:7289322-7289456	-	. 6 . . 6 . . . . . . 6 . . . . 5 .	22, 21	U, U	NA	NA	NA	new	spe
ddAraHall_1.3_2:8724691-8725066	+	. . . . 5	24	G	0,00	5,08	AGO4	new	spe

ddAraHall_1.3_2:22168204-22168422	-	.	.	.	.	2	2	.	.	2	5	5	5	.	.	.	.	.	.	.	.	.	21, 22, 22, 22	U, C, C, U	NA	NA	NA	new	spe
ddAraHall_1.3_3:9066414-9066813	-	2	2	5	2	5	2	5	2	2	2	2	2	.	2	.	2	2	2	2	5	2	21, 21	U, A	9,69	0,00	AGO1	new	spe
ddAraHall_1.3_4:2951034-2951433	-	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	22, 24, 20	C, A, A	0,94	0,00	AGO1	new	spe
ddAraHall_1.3_6:3856282-3856674	+	5	2	.	.	5	.	.	.	.	5	2	2	2	2	.	5	.	.	.	.	.	21, 21, 21	C, A, A	0,10	1,59	AGO4	new	spe
ddAraHall_1.3_7:8846455-8846651	+	6	6	.	.	.	.	.	.	5	5	.	.	.	.	.	.	.	.	.	.	.	21, 21	U, U	11,18	0,08	AGO1	new	spe
ddAraHall_1.3_7:21661551-21661696	-	6	.	.	.	6	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	U, U	17,92	0,11	AGO1	new	spe
ddAraHall_1.3_7:21684820-21684965	-	6	6	.	.	6	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	U, U	17,92	0,11	AGO1	new	spe
ddAraHall_1.3_1:12944744-12945222	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21, 24	A, A	7,02	11,46	AGO4	new	spe
ddAraHall_1.3_1:7150694-7150861	+	.	2	2	2	.	.	.	2	.	.	1	.	.	.	.	.	.	.	.	.	5	22, 20	C, G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_1:16043382-16043504	-	.	.	.	.	.	.	.	.	6	6	.	.	.	.	.	.	.	.	.	.	.	22	A	NA	NA	NA	new	spe
ddAraHall_1.3_2:22914344-22914527	+	.	.	.	.	.	.	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	22	U	22,88	0,00	AGO1	new	spe
ddAraHall_1.3_3:7285651-7285798	-	.	.	.	.	.	.	.	.	.	.	.	.	6	.	.	.	.	.	.	.	.	21	C	NA	NA	NA	new	spe
ddAraHall_1.3_4:12455573-12455769	+	.	6	.	.	.	.	.	.	6	6	.	.	.	.	.	.	.	.	.	.	.	21, 21	U, A	31,68	0,26	AGO1	new	spe
ddAraHall_1.3_4:19972603-19973002	+	5	.	2	2	2	.	.	2	5	5	2	.	2	2	2	2	2	2	2	.	2	24, 24	C, C	0,04	0,45	AGO4	new	spe
ddAraHall_1.3_4:21169354-21169457	-	5	.	.	.	.	.	.	.	.	.	4	.	4	.	.	2	4	5	4	5	.	21, 21	A, A	NA	NA	NA	new	spe
ddAraHall_1.3_4:21249542-21249798	-	6	6	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_5:8782553-8782730	+	.	.	2	6	.	.	.	5	2	2	2	.	.	5	.	2	2	.	2	.	.	21, 24	U, C	10,03	0,09	AGO1	new	spe
ddAraHall_1.3_6:6879442-6879630	+	5	5	.	.	5	.	.	.	4	.	.	.	.	.	.	.	.	.	.	.	.	22	U	47,60	0,09	AGO1	new	spe
ddAraHall_1.3_6:24140858-24141007	-	5	5	4	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	U	NA	NA	NA	new	spe
ddAraHall_1.3_7:10362118-10362514	-	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24, 24	U, C	NA	NA	NA	new	spe
ddAraHall_1.3_7:10446838-10447234	-	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24, 24	U, C	NA	NA	NA	new	spe
ddAraHall_1.3_8:21793206-21793533	-	6	.	.	.	.	.	.	.	6	6	.	.	.	.	.	.	.	.	.	.	.	21	A	NA	NA	NA	new	spe
ddAraHall_1.3_1:9764331-9764433	+	2	6	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	U, U	2,47	0,00	AGO1	new	spe
ddAraHall_1.3_7:16276716-16276897	+	.	6	.	.	.	.	.	.	.	.	.	.	.	.	5	.	.	.	2	.	.	24, 21	G, U	20,00	13,68	AGO1	new	spe
ddAraHall_1.3_1:597043-597210	-	6	6	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_6:19777638-19777707	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	U	0,90	1,53	AGO4	new	spe
ddAraHall_1.3_8:1308929-1309274	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	U	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_1:12476759-12476945	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21	U	AGO1	NA	AGO1	new	spe
ddAraHall_1.3_1:12617460-12617856	-	.	.	.	.	.	2	.	.	.	.	.	.	.	.	5	5	.	.	.	2	.	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_1:12695938-12696235	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	22	C	0,14	0,00	AGO1	new	spe
ddAraHall_1.3_1:16108090-16108201	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	C	0,00	0,14	AGO4	new	spe
ddAraHall_1.3_1:17213304-17213403	+	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	21	U	AGO1	NA	AGO1	new	spe
ddAraHall_1.3_1:27770076-27770170	-	2	2	.	.	.	.	5	.	5	.	.	.	.	.	.	2	.	.	.	.	.	21	A	NA	NA	NA	new	spe
ddAraHall_1.3_2:1040654-1040838	+	2	6	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	U, U	24,46	0,00	AGO1	new	spe
ddAraHall_1.3_2:1079998-1080127	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	C	NA	NA	NA	new	spe
ddAraHall_1.3_2:8104715-8104794	+	5	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24, 24	A, A	NA	NA	NA	new	spe
ddAraHall_1.3_2:17356391-17356790	-	.	.	.	.	2	2	.	.	.	5	2	5	.	.	.	4	4	4	.	.	4	24, 24	A, G	NA	NA	NA	new	spe
ddAraHall_1.3_2:21045619-21045873	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	U	10,45	2,64	AGO1	new	spe
ddAraHall_1.3_2:21366920-21367085	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	3,38	AGO4	new	spe
ddAraHall_1.3_3:489526-489599	+	6	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	9,12	AGO4	new	spe
ddAraHall_1.3_3:3245548-3245945	-	.	.	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	21, 22	A, G	0,00	2,82	AGO4	new	spe
ddAraHall_1.3_3:13770887-13770961	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	U	0,57	13,53	AGO4	new	spe
ddAraHall_1.3_3:21433207-21433591	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	U	0,00	0,43	AGO4	new	spe
ddAraHall_1.3_4:3115375-3115469	-	.	.	.	4	5	.	5	.	.	.	.	.	.	.	.	3	.	.	.	.	.	19, 21	U, A	3,85	0,37	AGO1	new	spe
ddAraHall_1.3_4:3584586-3584733	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	G	0,00	0,02	AGO4	new	spe
ddAraHall_1.3_4:15489381-15489638	+	6	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	22	C	0,14	AGO4	AGO1 and AGO4	new	spe
ddAraHall_1.3_5:7503658-7504019	+	2	5	.	.	.	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	A, G	8,77	0,00	AGO1	new	spe
ddAraHall_1.3_5:8032006-8032190	+	.	2	.	.	2	.	.	.	.	.	.	.	2	.	5	.	6	2	.	.	.	21, 24	U, A	AGO1	AGO4	AGO1 and AGO4	new	spe
ddAraHall_1.3_5:17658696-17658794	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	26,81	AGO4	new	spe
ddAraHall_1.3_5:17696230-17696328	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	26,81	AGO4	new	spe
ddAraHall_1.3_5:17731821-17731912	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_5:18377809-18378006	-	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	A, A	0,38	0,60	AGO4	new	spe
ddAraHall_1.3_5:19794624-19794853	+	.	5	.	.	.	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	U, U	9,85	0,00	AGO1	new	spe
ddAraHall_1.3_6:212049-212147	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	26,81	AGO4	new	spe
ddAraHall_1.3_6:4988724-4988808	+	.	.	.	.	.	.	.	.	6	6	.	.	.	.	.	.	.	.	.	.	.	24	C	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_6:5089955-5090080	-	5	2	2	2	2	.	.	2	1	.	.	2	.	2	.	4	6	1	2	2	1	22	G	NA	NA	NA	new	spe
ddAraHall_1.3_6:5692368-5692661	+	2	.	5	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	11,67	AGO4	new	spe
ddAraHall_1.3_6:13609057-13609365	+	2	5	.	.	.	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	A, G	8,77	0,00	AGO1	new	spe
ddAraHall_1.3_6:23366990-23367141	-	2	2	.	.	6	6	.	.	.	.	.	.	.	.	.	4	.	.	.	.	.	24	G	NA	NA	NA	new	spe
ddAraHall_1.3_6:26579278-26579412	-	4	5	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_7:1134534-1134622	+	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_7:11066865-11067259	-	2	5	.	.	.	.	.	.	.	3	2	.	2	5	.	2	2	.	2	.	2	19, 21	A, U	NA	NA	NA	new	spe
ddAraHall_1.3_7:14409493-14409710	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	A	0,00	6,63	AGO4	new	spe
ddAraHall_1.3_7:15969166-15969565	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	A	NA	NA	NA	new	spe
ddAraHall_1.3_7:16888813-16888989	+	.	.	.	.	.	.	.	.	.	.	5	.	5	2	2	2	2	.	2	2	2	20	U	38,31	0,00	AGO1	new	spe
ddAraHall_1.3_7:21581567-21581647	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	10,33	AGO4	new	spe
ddAraHall_1.3_7:21594891-21595058	+	.	.	.	.	.	.	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	21	A	3,32	0,00	AGO1	new	spe
ddAraHall_1.3_7:23323053-23323272	-	5	5	.	.	.	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	21	U	17,92	0,11	AGO1	new	spe
ddAraHall_1.3_7:25915414-25915560	-	.	.	.	.	.	2	.	.	5	.	.	5	.	.	.	.	.	.	.	.	.	21	U	3,92	0,00	AGO1	new	spe
ddAraHall_1.3_8:3464701-3464771	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_8:5804212-5804280	-	.	.	.	.	5	5	.	.	.	.	3	4	.	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_8:11214846-11215102	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	A	NA	NA	NA	new	spe
ddAraHall_1.3_8:16109880-16110253	+	6	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,15	20,63	AGO4	new	spe
ddAraHall_1.3_8:17090518-17090803	+	2	.	2	2	.	2	.	2	.	2	2	2	.	5	.	2	2	.	.	.	2	21, 21	A, U	17,89	0,12	AGO1	new	spe
ddAraHall_1.3_8:17799486-17799817	-	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21, 21	C, U	10,31	0,00	AGO1	new	spe
ddAraHall_1.3_8:18031457-18031627	-	5	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_7:10346384-10346463	+	5	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	58,52	AGO4	new	spe
ddAraHall_1.3_1:14212519-14212831	+	2	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	3,79	AGO4	new	spe
ddAraHall_1.3_2:12442239-12442523	+	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	6,89	AGO4	new	spe
ddAraHall_1.3_5:4743481-4743756	-	2	2	.	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,04	3,18	AGO4	new	spe
ddAraHall_1.3_5:10212744-10212846	+	.	5	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	25	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_5:11286743-11287141	+	2	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	21	U	1,59	AGO4	AGO1 and AGO4	new	spe
ddAraHall_1.3_7:10468211-10468609	-	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_8:4744148-4744414	+	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	24	A	0,00	7,87	AGO4	new	spe
ddAraHall_1.3_4:2533-2620	+	.	.	.	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	19	G	0,00	0,02	AGO4	new	spe
ddAraHall_1.3_2:12282574-12282660	-	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	22	A	0,78	0,82	AGO1 AGO4	new	spe

ddAraHall_1.3_2:12387224-12387308	+	. . . . . . S	22	A	0,78	0,82	AGO1 AGO4	new	spe
ddAraHall_1.3_8:8695615-8695796	+	. . . .	21	U	13,23	0,00	AGO1	new	spe
ddAraHall_1.3_2:1657040-1657266	-	. S	21	U	14,63	0,03	AGO1	new	spe
ddAraHall_1.3_2:3242554-3242810	-	. 5	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_5:4082060-4082306	-	5	24	C	NA	AGO4	AGO4	new	spe
ddAraHall_1.3_8:2871678-2871800	+	5 2	24	U	0,90	1,53	AGO4	new	spe
ddAraHall_1.3_1:7097-7324	-	6	24	A	0,00	2,03	AGO4	new	spe
ddAraHall_1.3_1:1110805-1110918	-	2 5	24	A	0,00	18,23	AGO4	new	spe
ddAraHall_1.3_1:4182510-4182588	+	. 1 . . . 1 . . . . . . 1 . . 4 4 . . 5 4	24	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_1:5970293-5970366	+	. 5	24	A	0,00	0,88	AGO4	new	spe
ddAraHall_1.3_1:6550675-6551066	-	5	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_1:6796987-6797284	-	4 4 . 5	24	G	0,02	4,62	AGO4	new	spe
ddAraHall_1.3_1:6926780-6926909	+	. 5	23	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_1:7994029-7994165	+	2 5	24	A	0,00	18,91	AGO4	new	spe
ddAraHall_1.3_1:9114798-9115187	-	. 5 . . . 2 . . . . 2 . . . 2	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_1:12478798-12478950	-	. 2 . 6 . 4 4 . . . . 2 . 2	21	A	NA	NA	NA	new	spe
ddAraHall_1.3_1:12970700-12970966	+	. 5	24	A	0,00	21,46	AGO4	new	spe
ddAraHall_1.3_1:13884475-13884654	+	6	22	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_1:15130841-15130959	-	2 . . . . 5	24	U	4,71	5,84	AGO4	new	spe
ddAraHall_1.3_1:15958417-15958786	+	. 5 . . . . 1	24	A	0,00	5,30	AGO4	new	spe
ddAraHall_1.3_1:17280778-17280979	-	5 2	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_1:18940212-18940293	-	5	24	C	0,00	28,24	AGO4	new	spe
ddAraHall_1.3_1:18949129-18949201	+	. 5	24	A	0,03	7,86	AGO4	new	spe
ddAraHall_1.3_1:18965051-18965123	+	. 5	24	A	0,03	7,86	AGO4	new	spe
ddAraHall_1.3_1:21082592-21082878	+	. 6	24	G	NA	NA	NA	new	spe
ddAraHall_1.3_1:24193763-24193839	+	2 . . 5 . 1 1	23	C	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_1:25661705-25661859	+	. 4 . . . . 5	23	A	0,00	10,18	AGO4	new	spe
ddAraHall_1.3_1:26743281-26743351	-	5	24	C	NA	NA	NA	new	spe
ddAraHall_1.3_1:28527238-28527331	+	3 5	24	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_1:29121439-29121514	-	5	23	A	NA	NA	NA	new	spe
ddAraHall_1.3_1:30231157-30231299	+	5 2	24	A	0,00	11,54	AGO4	new	spe
ddAraHall_1.3_1:30923864-30924036	+	. 5	21	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_2:654836-654972	-	2 5	24	A	0,00	18,91	AGO4	new	spe
ddAraHall_1.3_2:1165514-1165830	-	. 5	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_2:1353780-1353919	+	. . . . 5 2	21	U	AGO1	NA	AGO1	new	spe
ddAraHall_1.3_2:1400850-1401248	-	5 2	24	A	0,00	17,00	AGO4	new	spe
ddAraHall_1.3_2:2407586-2407833	-	. 6	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_2:2574865-2575068	-	. . . . 5	21	A	NA	NA	NA	new	spe
ddAraHall_1.3_2:2941148-2941306	-	. 5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_2:3829873-3829970	+	. 5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_2:4234679-4234933	-	. 5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_2:4540022-4540180	-	. 5	24	A	NA	AGO4	AGO4	new	spe
ddAraHall_1.3_2:7018253-7018341	-	. 6	23	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_2:8489772-8489964	+	. . . . S	21	U	4,24	0,00	AGO1	new	spe
ddAraHall_1.3_2:12297552-12297638	+	. . . . . 5	21	U	16,22	0,00	AGO1	new	spe
ddAraHall_1.3_2:12305160-12305231	-	5 . . . . . 4	21	C	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_2:12521689-12521770	+	. . . . 6	22	C	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_2:12728661-12728791	-	2 . . . . 5	21	C	NA	NA	NA	new	spe
ddAraHall_1.3_2:13979848-13980108	-	 5	22	U	NA	NA	NA	new	spe
ddAraHall_1.3_2:14131833-14132128	-	2 5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_2:14643263-14643372	+	. . . . 1 5	19	U	AGO1	NA	AGO1	new	spe
ddAraHall_1.3_2:14761433-14761596	+	2 2 . . 5	25	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_2:16239944-16240336	+	2 . . . . 5	24	C	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_2:16897508-16897631	+	. 5	21	A	0,46	0,00	AGO1	new	spe
ddAraHall_1.3_2:17223468-17223620	-	2 5	23	G	0,00	0,10	AGO4	new	spe
ddAraHall_1.3_2:17773794-17773864	-	5 . . . . 1	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_2:17819501-17819577	+	. 5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_2:18985108-18985352	+	2 2 . . 2 . . . 6	21	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_2:19336888-19337286	-	4 5	20	U	NA	NA	NA	new	spe
ddAraHall_1.3_2:19769962-19770098	+	2 5	24	A	0,00	18,91	AGO4	new	spe
ddAraHall_1.3_2:20183160-20183232	+	. 5	24	A	0,23	0,77	AGO4	new	spe
ddAraHall_1.3_2:20577686-20578077	-	. . . 2 2 . 5 2 2 2 4 4 . 4 4 4 . 4 4 4 4	21	A	0,94	0,06	AGO1	new	spe
ddAraHall_1.3_2:22274527-22274687	+	5 2	24	A	0,00	51,99	AGO4	new	spe
ddAraHall_1.3_2:22344000-22344399	-	2 2 . 2 5 2	24	A	0,00	50,75	AGO4	new	spe
ddAraHall_1.3_2:24134806-24134876	-	1 1 . . . 5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_3:50804-50900	+	. 2 . . . . 5 .	24	C	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_3:380856-381100	-	. 2 . . 5	21	A	0,28	AGO4	AGO1 and AGO4	new	spe
ddAraHall_1.3_3:1519720-1520111	-	2 5	21	U	AGO1	NA	AGO1	new	spe
ddAraHall_1.3_3:2522442-2522528	+	2 2 2 2 2 5 2	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_3:4038590-4038701	+	. 2 . . . 5 . . . . . 3	24	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_3:4742349-4742429	-	. 5	24	A	0,01	44,10	AGO4	new	spe
ddAraHall_1.3_3:5327411-5327644	+	. . . . 5	24	A	0,03	11,44	AGO4	new	spe
ddAraHall_1.3_3:6876400-6876756	+	. 2 . . . 2 5 . 2 . 2 2 . . 2 2 . . 2 . .	21	U	19,28	0,11	AGO1	new	spe
ddAraHall_1.3_3:8354471-8354564	+	5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_3:10162772-10162901	-	. 5	24	A	0,00	0,05	AGO4	new	spe
ddAraHall_1.3_3:10898566-10898829	-	5	23	A	NA	NA	NA	new	spe
ddAraHall_1.3_3:14399138-14399208	-	2 2 . 2 . 5 . . . . 2 . 2	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_3:14523035-14523106	-	. 2 . . 5	23	C	NA	NA	NA	new	spe
ddAraHall_1.3_3:14877193-14877356	-	. 5	23	A	NA	NA	NA	new	spe
ddAraHall_1.3_3:15036899-15037002	+	2 5	24	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_3:16074864-16074958	+	. 5	24	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_3:21006817-21006966	+	. . . . 1 . . . . 5 . . .	24	A	0,00	6,64	AGO4	new	spe
ddAraHall_1.3_3:24698510-24698595	+	5	24	A	0,00	6,85	AGO4	new	spe
ddAraHall_1.3_3:26426776-26426927	+	. 6	21	U	32,16	0,00	AGO1	new	spe
ddAraHall_1.3_4:16548-16634	+	. . . S	18	C	0,00	0,00	AGO1	new	spe
ddAraHall_1.3_4:1692116-1692201	+	1 5	24	G	0,00	AGO4	AGO4	new	spe

ddAraHall_1.3_4:2193128-2193206	+	24	G	NA	NA	NA	new	spe
ddAraHall_1.3_4:2237258-2237333	+	24	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_4:2694679-2694776	-	23	C	NA	NA	NA	new	spe
ddAraHall_1.3_4:3363683-3364082	-	24	A	0,00	12,73	AGO4	new	spe
ddAraHall_1.3_4:3467940-3468052	+	21	U	17,47	0,00	AGO1	new	spe
ddAraHall_1.3_4:3873421-3873551	+	24	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_4:4223984-4224073	-	23	A	NA	NA	NA	new	spe
ddAraHall_1.3_4:4969248-4969316	-	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_4:6772792-6772950	-	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_4:12505209-12505461	+	21	U	11,19	0,00	AGO1	new	spe
ddAraHall_1.3_4:17255266-17255628	+	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_4:17391815-17391955	-	23	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_4:18812318-18812402	-	20	C	NA	NA	NA	new	spe
ddAraHall_1.3_4:19314274-19314366	+	24	A	0,00	150,98	AGO4	new	spe
ddAraHall_1.3_4:19820795-19821194	-	22	G	3,48	0,00	AGO1	new	spe
ddAraHall_1.3_4:20889608-20889681	+	24	A	NA	AGO4	AGO4	new	spe
ddAraHall_1.3_4:21628295-21628686	-	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_4:22217543-22217924	+	24	A	0,00	2,03	AGO4	new	spe
ddAraHall_1.3_4:25587214-25587350	+	23	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_4:25672130-25672298	-	21	C	NA	NA	NA	new	spe
ddAraHall_1.3_5:799702-799812	+	21	U	14,03	0,00	AGO1	new	spe
ddAraHall_1.3_5:4033392-4033464	-	24	A	0,00	0,11	AGO4	new	spe
ddAraHall_1.3_5:5100924-5101225	-	24	U	NA	NA	NA	new	spe
ddAraHall_1.3_5:5595979-5596075	+	24	G	0,00	2,15	AGO4	new	spe
ddAraHall_1.3_5:6706449-6706540	-	23	U	0,10	0,19	AGO4	new	spe
ddAraHall_1.3_5:6888153-6888471	-	23	U	AGO1	4,27	AGO1 and AGO4	new	spe
ddAraHall_1.3_5:7407494-7407615	-	21	C	NA	NA	NA	new	spe
ddAraHall_1.3_5:7728499-7728592	+	24	G	0,00	11,32	AGO4	new	spe
ddAraHall_1.3_5:8867567-8867760	+	23	A	0,00	2,08	AGO4	new	spe
ddAraHall_1.3_5:9431151-9431336	+	24	A	AGO1	0,00	AGO1	new	spe
ddAraHall_1.3_5:11765922-11766032	+	24	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_5:14836566-14836635	-	24	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_5:15128633-15128961	-	21	G	NA	NA	NA	new	spe
ddAraHall_1.3_5:16669714-16670057	-	24	A	0,00	133,56	AGO4	new	spe
ddAraHall_1.3_5:16670340-16670411	+	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_5:17098697-17098767	+	24	A	NA	AGO4	AGO4	new	spe
ddAraHall_1.3_5:17104660-17104730	+	24	A	NA	AGO4	AGO4	new	spe
ddAraHall_1.3_5:17548540-17548912	-	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_5:18668527-18668787	-	21	C	AGO1	NA	AGO1	new	spe
ddAraHall_1.3_5:18743482-18743610	-	23	C	NA	NA	NA	new	spe
ddAraHall_1.3_5:20201081-20201341	-	24	G	NA	NA	NA	new	spe
ddAraHall_1.3_5:21454319-21454411	+	23	U	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_5:22173663-22173782	+	24	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_6:1183017-1183416	+	22	U	25,89	0,00	AGO1	new	spe
ddAraHall_1.3_6:1824783-1824862	-	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_6:3717935-3718007	-	24	G	0,00	2,15	AGO4	new	spe
ddAraHall_1.3_6:7540165-7540337	-	21	C	NA	NA	NA	new	spe
ddAraHall_1.3_6:10789907-10790097	+	24	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_6:11072188-11072587	+	24	A	0,00	20,45	AGO4	new	spe
ddAraHall_1.3_6:11183398-11183475	-	24	A	0,00	10,54	AGO4	new	spe
ddAraHall_1.3_6:11517954-11518353	-	22	U	25,89	0,00	AGO1	new	spe
ddAraHall_1.3_6:12139312-12139463	-	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_6:13275826-13275899	+	20	U	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_6:16512027-16512115	-	23	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_6:19932709-19932843	+	25	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_6:20058647-20058741	-	24	U	NA	NA	NA	new	spe
ddAraHall_1.3_6:20635169-20635565	-	21	U	4,57	0,00	AGO1	new	spe
ddAraHall_1.3_6:20636173-20636298	-	21	C	NA	NA	NA	new	spe
ddAraHall_1.3_6:21108871-21109269	+	21	A	1,28	0,01	AGO1	new	spe
ddAraHall_1.3_6:22098938-22099051	-	23	U	NA	NA	NA	new	spe
ddAraHall_1.3_6:22530527-22530604	+	23	U	NA	NA	NA	new	spe
ddAraHall_1.3_6:23493380-23493500	-	21	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_6:23718664-23718807	+	22	U	7,43	0,04	AGO1	new	spe
ddAraHall_1.3_6:23718994-23719393	+	22	C	0,00	7,98	AGO4	new	spe
ddAraHall_1.3_6:25358692-25358872	-	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_6:25850330-25850401	-	24	G	NA	NA	NA	new	spe
ddAraHall_1.3_6:25899878-25900066	+	24	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_7:1128578-1128666	-	23	G	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_7:1393035-1393252	-	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_7:2629705-2629774	+	24	U	6,27	10,82	AGO4	new	spe
ddAraHall_1.3_7:3288753-3289152	-	21	U	8,45	0,02	AGO1	new	spe
ddAraHall_1.3_7:3576864-3577040	-	22	C	0,00	0,33	AGO4	new	spe
ddAraHall_1.3_7:3643881-3644280	+	23	A	NA	NA	NA	new	spe
ddAraHall_1.3_7:3930157-3930325	-	21	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_7:4445689-4446037	-	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_7:4915341-4915740	+	21	U	6,13	0,00	AGO1	new	spe
ddAraHall_1.3_7:4922175-4922434	-	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_7:5132326-5132441	-	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_7:5351752-5352040	+	24	G	NA	NA	NA	new	spe
ddAraHall_1.3_7:5657426-5657559	-	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_7:5781414-5781561	-	22	U	0,66	0,00	AGO1	new	spe
ddAraHall_1.3_7:5787104-5787503	-	24	U	AGO1	NA	AGO1	new	spe
ddAraHall_1.3_7:6468310-6468415	-	24	C	NA	NA	NA	new	spe
ddAraHall_1.3_7:8917234-8917632	+	24	A	0,11	22,77	AGO4	new	spe
ddAraHall_1.3_7:10187186-10187277	+	24	A	7,85	0,00	AGO1	new	spe
ddAraHall_1.3_7:10492671-10493061	-	24	U	NA	NA	NA	new	spe

ddAraHall_1.3_7:10547665-10547917	-	. 6	22	C	NA	NA	NA	new	spe
ddAraHall_1.3_7:10746955-10747323	+	. 6	24	A	0,00	9,14	AGO4	new	spe
ddAraHall_1.3_7:11583047-11583142	-	. 5	24	A	0,00	150,98	AGO4	new	spe
ddAraHall_1.3_7:11584825-11584945	-	. . . . S	21	U	11,71	0,01	AGO1	new	spe
ddAraHall_1.3_7:12856721-12856970	-	2 5	24	A	0,00	14,12	AGO4	new	spe
ddAraHall_1.3_7:12880993-12881071	-	. 5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_7:13377081-13377370	+	. 5	23	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_7:13680752-13680872	-	. . . . S	21	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_7:15590528-15590679	-	. 5	23	U	NA	NA	NA	new	spe
ddAraHall_1.3_7:17898939-17899325	-	. 5	24	A	0,00	5,36	AGO4	new	spe
ddAraHall_1.3_7:19048671-19048831	-	4 5	23	G	0,00	0,02	AGO4	new	spe
ddAraHall_1.3_7:20352154-20352230	-	. 2 . 1 5 2 2	22	U	0,71	2,46	AGO4	new	spe
ddAraHall_1.3_7:20646337-20646443	+	. 5	21	U	AGO1	NA	AGO1	new	spe
ddAraHall_1.3_7:23491711-23492022	-	2 . . . 2 . . . 6	21	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_8:2258793-2259095	-	. 5	23	U	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_8:3254822-3254904	+	. 5	18	U	NA	NA	NA	new	spe
ddAraHall_1.3_8:4569074-4569144	+	. . . . 2 . . . . . . 5 . 1	24	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_8:5414175-5414485	+	. . . . 5	24	A	0,00	0,94	AGO4	new	spe
ddAraHall_1.3_8:5469657-5469843	-	. . . . 5	21	U	4,24	0,00	AGO1	new	spe
ddAraHall_1.3_8:6971012-6971119	-	. 5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_8:6975673-6975780	-	. 5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_8:7170316-7170715	-	. 2 . . 5 . . . . 2 . . .	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_8:9061231-9061630	+	5	24	U	75,19	5,13	AGO1	new	spe
ddAraHall_1.3_8:9496290-9496457	-	5	24	A	0,00	AGO4	AGO4	new	spe
ddAraHall_1.3_8:9660035-9660128	-	5	24	A	NA	NA	NA	new	spe
ddAraHall_1.3_8:9669316-9669402	-	. 5	24	A	0,00	9,55	AGO4	new	spe
ddAraHall_1.3_8:11440574-11440688	+	. 5	24	A	0,00	73,20	AGO4	new	spe
ddAraHall_1.3_8:11810338-11810417	-	5	24	U	AGO1	NA	AGO1	new	spe
ddAraHall_1.3_8:12771972-12772195	+	. . . . 5	21	G	NA	NA	NA	new	spe
ddAraHall_1.3_8:13768769-13769168	-	2 5	24	A	0,03	11,44	AGO4	new	spe
ddAraHall_1.3_8:14169929-14170322	-	. 5	21	U	NA	NA	NA	new	spe
ddAraHall_1.3_8:16644499-16644606	+	. . . . 5	21	A	4,67	0,00	AGO1	new	spe
ddAraHall_1.3_8:17090921-17091040	-	2 2 . . . 2 . . . 6	23	U	NA	NA	NA	new	spe
ddAraHall_1.3_8:20021935-20022193	-	2 2 . . . 2 . . . . 5 . . . . 2 .	24	A	0,00	0,00	not loaded	new	spe
ddAraHall_1.3_8:20669057-20669358	-	. 2 . . . . 6	24	U	NA	NA	NA	new	spe
ddAraHall_1.3_8:21283278-21283560	-	. 5	23	U	NA	NA	NA	new	spe
ddAraHall_1.3_8:22300087-22300233	+	. 6	24	A	0,00	39,57	AGO4	new	spe

Arabidopsis lyrata predicted miRNA genes

Loci	strand	CP99	CP99inf	MN47	451894	512644	512645	518429	518430	518431	1271754	TOT3	TOT6	TOT9	miRNA lenght	5' nt	AGO1 enrichment	AGO4 enrichment	Main loading	Family	Conservation
scaffold_3:18339161-18339330	-	6	6	6	6	5	6	2	6	6	6	6	6	6	21, 21	U, G	NA	NA	NA	miR396	conserved
scaffold_7:10466365-10466533	-	6	6	6	6	6	5	6	6	6	6	6	6	6	21, 21	U, C	4,65	0,09	AGO1	miR168	conserved
scaffold_7:22618563-22618732	+	6	6	6	6	6	6	6	6	6	6	6	6	6	21, 21	U, G	86,11	0,31	AGO1	miR396	conserved
scaffold_3:10584820-10584966	+	6	6	6	6	6	6	6	6	6	6	6	6	6	21, 21	G, U	9,00	0,11	AGO1	miR167	conserved
scaffold_5:24170297-24170423	+	6	6	6	5	6	5	6	6	6	6	6	6	6	21, 20	U, U	30,74	0,33	AGO1	miR167	conserved
scaffold_1:7732595-7732893	+	.	.	.	.	.	.	6	.	.	.	5	6	6	21, 20	U, G	13,87	0,08	AGO1	miR159	conserved
scaffold_4:13112227-13112362	-	6	5	5	6	6	6	6	6	6	6	5	6	6	21, 19	A, A	21,25	0,05	AGO1	miR172	conserved
scaffold_6:1487638-1488021	-	6	6	6	6	6	6	6	6	6	6	6	6	.	21, 21	A, G	NA	NA	NA	miR172	conserved
scaffold_2:16334323-16334482	-	6	6	5	5	6	5	.	6	6	6	6	6	6	21, 20	U, U	65,28	0,17	AGO1	miR156	conserved
scaffold_8:2352381-2352513	+	6	6	6	6	6	.	6	6	6	6	6	6	6	21, 20	U, U	6,0	0,1	AGO1	miR168	conserved
scaffold_2:20499132-20499356	-	6	5	6	6	6	6	6	6	6	2	6	6	6	21, 21	U, U	NA	NA	NA	miR159	conserved
scaffold_4:24190644-24190762	+	.	6	5	4	6	5	5	6	6	6	6	6	6	21, 21, 21	U, C, U	65,13	0,42	AGO1	miR164	conserved
scaffold_4:23731527-23731763	+	6	6	6	.	6	.	.	6	.	5	6	6	6	21, 22, 21, 21	G, U, C, U	28,94	0,19	AGO1	miR166	conserved
scaffold_2:16362190-16362324	+	5	6	5	5	2	5	.	6	5	5	.	6	5	21, 20	U, U	65,28	0,17	AGO1	miR156	conserved
scaffold_3:8127044-8127270	-	5	5	5	5	5	5	.	5	5	5	.	5	5	21, 20	U, U	65,28	0,17	AGO1	miR156	conserved
scaffold_6:195185-195404	-	.	6	4	4	6	5	6	6	5	6	5	6	5	21, 21	U, U	103,81	0,05	AGO1	miR164	conserved
scaffold_8:20675760-20675919	+	.	6	5	4	6	6	.	6	6	6	6	6	6	21, 21, 20	G, G, G	6,99	0,17	AGO1	miR172	conserved
scaffold_2:21553271-21553590	+	.	.	.	.	.	.	.	.	.	.	.	5	.	20, 21	U, U	9,64	0,04	AGO1	miR394	conserved
scaffold_6:3343764-3343893	-	6	6	6	4	6	4	5	6	6	5	6	6	6	21	U	NA	NA	NA	miR162	conserved
scaffold_6:10004404-10004598	-	6	6	6	4	6	4	5	6	6	6	6	6	6	21	U	NA	NA	NA	miR162	conserved
scaffold_4:24055424-24055560	+	6	6	6	6	6	5	6	6	.	.	6	6	6	21, 22, 21	U, U, U	66,44	0,25	AGO1	miR403	conserved
scaffold_1:463655-463802	+	6	6	6	.	6	.	.	6	6	5	6	6	6	21, 20	U, U	46,83	0,20	AGO1	miR166	conserved
scaffold_5:23568107-23568256	+	6	6	5	.	6	.	.	.	.	.	6	6	6	21, 22, 21, 21	G, U, G, U	27,37	1,14	AGO1	miR166	conserved
scaffold_6:3758521-3758667	+	6	6	6	.	6	.	.	6	.	.	6	6	6	21, 22, 21	G, U, U	36,47	0,25	AGO1	miR166	conserved
scaffold_6:3763710-3763848	+	6	6	6	.	6	.	.	6	.	.	6	6	6	21, 22, 21	G, U, U	36,47	0,25	AGO1	miR166	conserved
scaffold_6:24787408-24787595	+	5	5	5	.	5	.	.	5	5	5	6	5	5	21, 20	U, U	46,83	0,20	AGO1	miR166	conserved
scaffold_7:29120929-29121103	-	6	.	6	.	6	.	.	.	.	.	.	6	6	21, 22, 21, 20	G, U, U, C	62,52	0,23	AGO1	miR166	conserved
scaffold_8:4346047-4346177	-	6	6	5	.	6	.	.	6	6	.	.	6	.	21, 22, 21, 21, 20,	U, U, G, U, C,	52,92	2,32	AGO1	miR166	conserved
scaffold_6:4497774-4497899	-	6	6	6	6	6	6	.	6	6	.	.	6	6	21	U	NA	NA	NA	miR156	conserved
scaffold_1:14562698-14563002	+	6	6	6	5	6	5	5	.	.	6	.	6	6	22, 20, 21	U, U, U	44,05	0,30	AGO1	miR167	conserved
scaffold_2:2496494-2496776	+	.	6	6	6	.	.	6	.	.	.	6	6	6	22, 21, 21, 21	U, U, A, U	68,25	0,02	AGO1	miR171	conserved
scaffold_1:4824962-4825120	-	.	.	5	6	6	5	6	6	6	.	6	6	6	21, 21, 20	U, A, U	36,28	0,02	AGO1	miR171	conserved
scaffold_4:19897332-19897469	+	6	5	5	6	6	6	6	.	.	.	6	6	6	21, 21, 21	G, A, U	5,99	0,01	AGO1	miR160	conserved
scaffold_4:19511600-19511770	+	6	6	4	.	6	6	.	6	6	6	6	6	6	21, 21	A, C	0,71	0,35	AGO1	miR390	conserved
scaffold_8:20138510-20138645	+	6	6	4	.	6	6	.	6	6	6	6	6	6	21, 20	A, C	0,29	0,42	AGO4	miR390	conserved
scaffold_1:8837470-8837647	+	.	5	5	4	6	1	5	6	6	6	.	5	5	20, 21	U, U	9,64	0,04	AGO1	miR394	conserved
scaffold_8:22639288-22639421	+	2	6	5	.	5	.	.	.	.	.	.	.	.	21, 22, 22, 21	C, U, C, U	41,29	0,09	AGO1	miR166	conserved
scaffold_7:11508636-11508758	-	.	6	5	5	6	.	5	.	.	.	6	6	6	21	U	16,19	0,00	AGO1	miR160	conserved
scaffold_8:1035901-1036026	+	5	5	5	6	6	.	6	.	.	.	6	6	6	21, 21	A, U	8,98	0,01	AGO1	miR160	conserved
scaffold_5:20579266-20579398	+	.	6	4	.	6	6	5	.	.	6	6	6	6	21, 20, 21, 21	A, A, G, A	21,50	0,40	AGO1	miR172	conserved
scaffold_6:6030925-6031075	+	6	.	6	.	.	.	6	5	5	6	6	6	.	21, 21, 21	G, U, A	6,41	0,14	AGO1	miR398	conserved
scaffold_8:18643741-18643868	-	.	6	.	.	6	5	.	5	5	.	.	6	6	21, 20	U, U	NA	AGO4	AGO4	miR156	conserved
scaffold_1:28968029-28968183	-	5	6	5	5	6	.	.	.	.	6	6	.	6	21, 21	G, U	8,38	0,00	AGO1	miR169	conserved
scaffold_1:28971058-28971209	+	5	6	5	5	6	.	.	.	.	6	6	.	6	21, 21	G, U	8,38	0,00	AGO1	miR169	conserved
scaffold_4:20820341-20820555	+	.	5	4	.	4	5	5	5	5	.	.	6	5	21, 21, 21	U, U, U	15,78	1,55	AGO1	miR159	conserved
scaffold_3:4762348-4762490	-	.	5	4	.	6	5	2	.	.	6	5	6	5	21, 19, 21	A, A, A	29,41	0,05	AGO1	miR172	conserved
scaffold_7:8124434-8124652	-	.	6	4	.	6	5	5	2	2	5	2	6	6	21, 21	U, A	NA	NA	NA	miR319	conserved
scaffold_1:24965596-24965813	-	.	6	4	.	6	6	.	5	5	.	.	6	.	21	C	NA	NA	NA	miR156	conserved
scaffold_7:4705154-4705322	-	.	.	5	4	6	5	.	.	6	.	5	6	6	23, 22, 21	U, U, C	50,60	0,00	AGO1	miR156	conserved
scaffold_7:9426941-9427220	+	5	.	.	.	.	.	.	.	.	.	.	.	6	21, 21	G, U	8,38	0,00	AGO1	miR169	conserved
scaffold_4:23451033-23451258	+	.	.	5	.	5	5	5	.	.	.	5	5	5	21, 22, 18, 19	U, U, U, U	43,57	0,26	AGO1	miR159	conserved
scaffold_6:11195930-11196095	+	.	.	5	.	6	6	.	.	.	6	6	6	6	21, 21, 22, 22	C, U, G, C	8,14	0,00	AGO1	miR169	conserved
scaffold_5:22563918-22564038	-	6	.	6	.	2	2	.	5	5	5	.	6	6	21	U	NA	NA	NA	miR171	conserved
scaffold_5:18636711-18636908	+	5	5	4	4	5	4	5	.	.	4	.	5	5	21, 21	U, U	11,43	0,03	AGO1	miR171	conserved
scaffold_7:28925443-28925743	-	.	6	1	.	4	4	6	2	5	.	2	6	6	21	U	NA	NA	NA	miR319	conserved
scaffold_2:18527545-18527664	-	.	5	.	.	5	.	.	.	.	5	.	6	6	21	C	NA	NA	NA	miR395	conserved
scaffold_2:18530518-18530632	-	.	6	.	.	6	.	.	.	.	6	.	6	6	21	C	NA	NA	NA	miR395	conserved
scaffold_2:18534981-18535095	-	.	6	.	.	6	.	.	.	.	6	.	6	6	21	C	NA	NA	NA	miR395	conserved
scaffold_1:5103741-5103945	-	5	5	6	5	.	.	2	2	2	4	.	5	6	21, 22, 21	U, U, U	NA	NA	NA	miR472	conserved
scaffold_3:1730732-1730908	-	.	.	5	.	5	4	5	.	.	.	6	.	5	21, 23	U, U	NA	NA	NA	miR167	conserved
scaffold_3:6285695-6285861	-	5	.	5	.	6	4	.	.	.	6	5	.	.	21, 20, 20, 21, 21,	G, A, G, U, U,	7,16	0,00	AGO1	miR169	conserved
scaffold_3:5637983-5638275	-	.	.	5	5	6	4	6	.	.	6	.	.	.	21, 21	C, U	59,64	0,00	AGO1	miR169	conserved
scaffold_6:12927879-12928004	-	6	.	6	5	2	4	.	.	.	.	.	6	6	21, 21, 20	C, U, U	54,91	0,16	AGO1	miR164	conserved
scaffold_5:20691229-20691405	+	6	.	6	.	6	.	.	.	.	.	.	6	6	21	A	0,16	0,1	AGO1	miR393	conserved
scaffold_8:21058395-21058552	+	.	.	.	.	.	.	.	.	.	.	.	5	5	21, 21, 21	C, U, A	26,19	0,43	AGO1	miR391	conserved
scaffold_7:27198144-27198470	-	.	.	.	.	6	6	.	.	.	6	.	6	.	21, 21	C, C	1,82	0,00	AGO1	miR169	conserved
scaffold_4:20308848-20309031	+	6	.	6	.	.	.	.	.	.	.	.	6	6	21, 22	A, U	7,03	0,15	AGO1	miR393	conserved
scaffold_2:1849331-1849475	+	.	.	5	.	6	.	.	.	.	.	.	6	6	21	U	21,95	0,08	AGO1	miR399	conserved
scaffold_5:1500669-1500912	+	6	.	5	5	.	.	.	.	.	.	.	.	6	19, 21, 20	U, G, U	13,92	0,02	AGO1	miR398	conserved
scaffold_8:24423935-24424037	+	5	6	.	.	.	.	.	.	.	.	.	6	5	21	U	14,18	0,04	AGO1	miR170	conserved
scaffold_6:12200152-12200433	+	.	.	.	.	6	6	.	.	.	6	.	.	.	21, 20	G, U	11,27	0,16	AGO1	miR156	conserved
scaffold_6:22918348-22918473	+	.	.	5	.	.	.	.	.	.	.	6	.	5	21	U	17,29	0,00	AGO1	miR397	conserved

scaffold_8:21885625-21885756	+	.	.	4	.	6	.	.	.	.	.	.	6	6	21	U	21,95	0,08	AGO1	miR399	conserved
scaffold_1:11095790-11095900	-	.	.	4	.	6	.	.	.	.	.	.	6	6	21	C	11,61	0,07	AGO1	miR395	conserved
scaffold_2:18532439-18532565	+	.	.	4	.	5	.	.	.	.	.	.	6	6	21	C	11,61	0,07	AGO1	miR395	conserved
scaffold_2:18535778-18535919	+	.	.	4	.	5	.	.	.	.	.	.	6	6	21	C	11,61	0,07	AGO1	miR395	conserved
scaffold_4:3838938-3839208	-	.	.	.	.	.	.	.	.	.	.	.	.	5	22, 22	U, G	AGO1	NA	AGO1	miR156	conserved
scaffold_7:4307377-4307461	+	.	.	.	.	6	.	.	.	.	.	.	5	.	22, 21	U, G	0,64	0,00	AGO1	miR156	conserved
scaffold_6:5017520-5017655	+	.	.	.	.	6	4	.	.	.	.	.	.	.	22	U	AGO1	NA	AGO1	miR156	conserved
scaffold_3:24345954-24346091	-	.	.	.	.	5	4	.	.	.	.	.	.	.	20	C	NA	NA	NA	miR156	conserved
scaffold_7:15302900-15303010	-	.	.	4	.	.	.	.	.	.	.	6	.	2	21	U	17,29	0,00	AGO1	miR397	conserved
scaffold_4:4754938-4755029	+	.	.	.	.	.	.	.	.	.	.	.	.	S	20	A	3,69	1,98	AGO1	miRN34	conserved
scaffold_4:4756451-4756542	+	.	.	.	.	.	.	.	.	.	.	.	.	S	20	A	3,69	1,98	AGO1	miRN34	conserved
scaffold_4:4758976-4759067	+	.	.	.	.	.	.	.	.	.	.	.	.	S	20	A	3,69	1,98	AGO1	miRN34	conserved
scaffold_1:29644768-29644873	+	6	6	6	6	6	6	6	6	6	6	6	6	6	20, 21, 19	U, C, U	9,37	0,05	AGO1	miR158	brass
scaffold_3:4476040-4476146	-	.	.	.	.	.	.	.	.	.	.	.	6	6	20, 21	U, C	0,27	0,02	AGO1	miR158	brass
scaffold_7:8297990-8298193	+	.	6	.	.	6	6	.	5	5	6	.	6	.	21, 18	C, C	6,24	0,00	AGO1	miR845	brass
scaffold_3:934593-934921	-	.	.	1	4	6	5	.	.	.	6	.	.	.	21	A	NA	NA	NA	miR3435	brass
scaffold_7:8301374-8301477	+	.	.	.	.	6	6	.	.	.	.	.	6	.	21	U	10,87	0,01	AGO1	miR845	brass
scaffold_1:16207693-16207801	-	.	.	.	.	.	.	.	.	.	.	.	6	6	21	G	NA	NA	NA	miR5654	brass
scaffold_1:24641408-24641553	+	6	6	5	4	6	6	5	6	5	5	5	5	5	21, 22, 21, 21, 21, 22	C, U, U, U, C, U	6,58	0,05	AGO1	miR161	brass
scaffold_3:10848249-10848365	+	.	6	6	6	6	6	6	6	6	.	6	6	6	22, 21	U, G	4,71	0,74	AGO1	miR173	brass
scaffold_4:6694851-6694978	+	6	6	5	5	5	5	6	6	6	6	2	5	2	21, 22	U, U	10,87	0,04	AGO1	miR825	brass
scaffold_2:3142643-3143033	-	.	5	6	5	.	.	2	.	6	.	2	2	.	21, 21	U, U	NA	NA	NA	miR846	ath
scaffold_5:1066098-1066404	-	2	5	2	6	5	.	2	.	.	.	2	.	5	22, 21, 22	U, U, A	NA	NA	NA	miR840	ath
scaffold_2:3143101-3143324	-	.	.	.	.	2	.	4	.	.	5	.	5	.	21, 21	U, U	NA	NA	NA	miR842	ath
scaffold_6:12131618-12131807	+	.	.	.	.	.	.	.	.	.	.	.	5	.	21, 24	A, G	0,92	0,03	AGO1	new	ath
scaffold_3:5809720-5810027	-	.	.	.	.	.	.	.	.	.	6	.	6	.	21, 22	C, U	NA	NA	NA	miR823	ath
scaffold_6:1149694-1150092	-	.	.	4	.	6	5	.	.	.	.	2	2	.	22	G	NA	NA	NA	miR822	ath
scaffold_3:11022598-11022973	+	2	2	2	.	2	.	.	.	.	.	.	5	5	22	U	24,0	0,05	AGO1	new	ath
scaffold_8:16370165-16370323	+	.	.	4	.	6	5	.	.	.	.	.	.	.	21	A	8,17	0,43	AGO1	new	ath
scaffold_4:20125655-20125936	+	.	.	5	.	.	.	.	.	.	.	.	.	.	25	G	0,16	0,92	AGO4	new	ath
scaffold_1:15898441-15898721	-	.	.	.	.	.	.	.	.	.	.	.	.	6	24, 22, 22	G, U, U	NA	NA	NA	new	ath
scaffold_7:9586588-9586727	-	.	.	.	.	2	2	.	6	6	.	.	6	.	24	C	NA	NA	NA	miR1887	ath
scaffold_2:21934851-21935048	+	.	.	.	.	.	.	.	.	.	.	1	5	.	22	G	0,00	0,00	not loaded	new	ath
scaffold_3:8085025-8085336	+	.	5	6	.	5	.	.	5	.	6	.	6	.	24, 22, 22, 22	A, U, U, C	5,57	0,02	AGO1	miR3440	shared
scaffold_2:16255112-16255511	+	6	6	2	.	.	5	.	6	6	5	.	6	6	23, 21, 21	U, A, U	7,51	0,02	AGO1	miR163	shared
scaffold_5:2368881-2369005	+	6	.	5	.	6	6	.	.	.	6	6	6	6	21, 21	G, U	44,27	9,46	AGO1	miR3443	shared
scaffold_2:22627822-22627962	+	6	6	6	.	6	6	.	.	.	.	.	6	6	21, 21	G, U	45,65	0,10	AGO1	miR3433	shared
scaffold_1:23962046-23962182	-	6	6	6	.	6	.	.	.	.	6	.	6	.	21	C	NA	NA	NA	miR3436	shared
scaffold_2:22092583-22092814	+	5	5	.	.	5	.	4	.	.	.	5	5	5	22	U	18,57	0,11	AGO1	miR402	shared
scaffold_4:18982834-18983208	+	.	.	5	6	6	.	2	.	.	.	5	5	5	22, 21, 21	U, U, A	9,72	0,08	AGO1	miR3439	shared
scaffold_2:3938794-3938951	-	.	.	.	.	6	.	.	.	.	.	.	6	6	21	U	NA	NA	NA	miR859	shared
scaffold_4:13859879-13860078	-	2	.	2	.	.	5	5	.	.	.	.	2	2	21, 21	U, A	NA	NA	NA	miR3447	shared
scaffold_6:9731847-9732213	-	5	.	4	.	5	.	.	.	.	.	.	6	.	21, 21	G, U	NA	NA	NA	miR3434	shared
scaffold_7:4341870-4342043	-	2	.	6	.	.	.	5	.	.	.	.	6	6	22, 22	U, U	NA	NA	NA	miR3448	shared
scaffold_4:1882693-1883016	+	.	6	.	.	6	.	.	.	.	6	.	.	.	21, 21	U, U	29,75	0,00	AGO1	new	shared
scaffold_2:17015175-17015472	-	.	.	.	.	.	.	.	.	.	5	.	.	.	22, 21, 22	U, A, U	NA	NA	NA	miRN127	shared
scaffold_3:7126857-7127007	-	.	.	.	.	.	.	.	6	6	.	.	6	.	21	G	NA	NA	NA	miRN104	shared
scaffold_2:3938587-3938708	-	.	.	.	.	6	.	.	.	.	.	.	6	6	21, 21, 21	U, U, U	21,51	0,06	AGO1	new	shared
scaffold_3:8412402-8412567	+	.	.	.	.	6	.	.	.	.	.	.	5	.	21, 24	U, U	5,6	0,00	AGO1	miR868	shared
scaffold_7:10303197-10303418	+	.	.	.	.	6	6	.	.	.	.	.	.	.	22	A	9,03	0,00	AGO1	miR4240	shared
scaffold_1:25661712-25661855	+	.	.	.	.	6	.	.	.	.	.	.	6	.	21	G	0,00	0,0	not loaded	miR4225	shared
scaffold_3:4512226-4512476	-	.	.	6	.	.	.	.	.	.	.	.	.	.	21	U	NA	NA	NA	miR3438	shared
scaffold_4:1918968-1919247	-	.	.	4	.	.	.	2	5	5	.	.	.	.	21	A	NA	NA	NA	miR844	shared
scaffold_4:3971727-3972063	+	.	.	4	.	2	.	5	.	.	.	.	.	5	21, 23	U, A	28,17	12,56	AGO1	miR862	shared
scaffold_4:20150548-20150640	-	2	4	4	2	2	2	2	6	6	2	2	.	2	21	C	NA	NA	NA	miRN108	shared
scaffold_1:2102646-2102957	+	.	.	3	5	.	4	6	.	.	2	.	.	.	21	U	NA	NA	NA	new	shared
scaffold_3:7126368-7126486	-	.	.	.	.	6	.	.	.	.	.	.	6	.	21	U	NA	NA	NA	new	shared
scaffold_3:23251785-23252052	-	.	.	5	.	.	.	.	.	.	.	.	.	.	21, 21	U, A	NA	NA	NA	new	shared
scaffold_5:1905615-1906014	+	3	2	2	.	2	.	2	2	2	2	5	4	5	21	C	0,06	0,01	AGO1	new	shared
scaffold_6:13034817-13035194	+	.	.	.	.	.	4	.	.	.	5	.	.	.	21	C	54,37	0,54	AGO1	new	shared
scaffold_7:6908738-6909015	-	.	6	.	.	.	.	.	.	.	2	.	2	.	21	U	NA	NA	NA	new	shared
scaffold_5:15195221-15195378	-	.	.	4	.	.	.	.	.	.	.	.	6	.	22	U	NA	NA	NA	miR4222	shared
scaffold_1:23442246-23442445	+	.	.	.	.	.	.	.	.	.	5	.	.	.	21	C	0,00	0,00	not loaded	new	shared
scaffold_1:26243247-26243327	-	6	.	.	.	.	.	2	.	.	2	4	4	2	21	U	NA	NA	NA	new	shared
scaffold_2:3938433-3938539	-	.	.	.	.	.	.	.	.	.	.	.	5	.	21	U	NA	NA	NA	new	shared
scaffold_2:20872866-20873156	-	.	2	.	.	.	.	.	2	2	.	.	6	.	21	C	NA	NA	NA	new	shared
scaffold_5:3381231-3381568	+	.	.	.	.	.	.	.	.	.	.	.	5	.	24	A	0,00	1,13	AGO4	new	shared
scaffold_5:5460226-5460516	+	.	.	.	.	.	.	.	.	.	.	.	5	.	24	A	0,00	AGO4	AGO4	new	shared
scaffold_5:15191322-15191479	-	.	.	.	.	.	.	.	.	.	.	.	6	.	22	U	NA	NA	NA	new	shared
scaffold_5:20311857-20311988	+	.	.	5	.	2	.	.	.	.	.	.	.	.	21	A	0,90	0,00	AGO1	new	shared
scaffold_7:13930246-13930440	-	.	.	5	.	.	.	.	.	.	.	.	.	.	22	U	NA	NA	NA	new	shared
scaffold_1:29839793-29840001	+	6	6	6	.	2	2	.	.	.	6	.	.	.	24, 21, 24	A, U, C	AGO1	NA	AGO1	new	spe
scaffold_4:22761894-22762156	-	2	.	.	.	.	.	.	.	.	.	.	5	.	24, 21, 21	A, A, U	35,55	0,21	AGO1	new	spe
scaffold_7:9600468-9600645	-	.	.	.	.	6	.	.	6	6	.	.	.	.	21	U	NA	NA	NA	new	spe

scaffold_7:24352202-24352382	+	.	.	.	.	6	6	.	5	5	.	.	.	.	21	A	0,62	0,00	AGO1	new	spe
scaffold_1:7844872-7845111	-	2	2	5	6	6	6	4	.	.	.	.	.	.	21, 21	C, G	NA	NA	NA	new	spe
scaffold_8:18712237-18712483	-	.	.	.	.	.	.	.	.	.	.	.	6	6	20, 21, 21	U, U, C	AGO1	NA	AGO1	new	spe
scaffold_2:14156512-14156741	+	.	.	.	.	.	.	.	.	.	.	6	6	6	24	A	0,00	1,76	AGO4	new	spe
scaffold_3:7117029-7117182	-	.	6	.	.	.	.	.	.	.	.	.	6	.	22	U	NA	NA	NA	new	spe
scaffold_4:17164213-17164522	+	.	.	.	.	.	.	.	.	.	.	6	5	5	24	A	0,00	7,39	AGO4	new	spe
scaffold_7:12801900-12802298	+	5	.	2	5	2	5	2	.	.	.	2	2	2	24, 24, 24	G, A, A	0,00	AGO4	AGO4	new	spe
scaffold_5:23841712-23841903	-	.	.	2	.	.	.	.	6	6	.	.	.	.	21	A	NA	NA	NA	new	spe
scaffold_6:5776820-5777064	-	.	.	.	4	6	6	.	.	.	2	.	.	.	22	U	NA	NA	NA	new	spe
scaffold_1:1370974-1371324	-	.	.	.	.	.	.	.	.	.	.	5	5	4	24	A	NA	NA	NA	new	spe
scaffold_1:10217407-10217592	-	5	.	3	.	.	.	.	.	.	.	.	5	.	21	C	NA	NA	NA	new	spe
scaffold_1:15068332-15068648	+	.	.	.	.	5	.	.	.	.	5	.	.	.	24, 24	G, A	NA	NA	NA	new	spe
scaffold_1:22908113-22908512	-	.	5	.	.	.	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_1:25632287-25632685	+	.	.	4	.	2	5	5	2	2	.	.	.	.	21, 21	G, A	NA	NA	NA	new	spe
scaffold_2:1542881-1542980	+	5	.	5	.	.	.	.	.	.	.	.	.	.	22	U	NA	NA	NA	new	spe
scaffold_2:18297311-18297604	-	.	5	4	.	.	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_3:8879569-8879967	-	.	.	2	2	2	2	2	5	5	.	.	.	.	21	A	NA	NA	NA	new	spe
scaffold_3:11218481-11218880	+	.	.	2	2	2	2	2	5	5	.	.	.	.	22	U	AGO1	NA	AGO1	new	spe
scaffold_4:12616676-12617027	+	.	.	.	.	.	6	.	.	.	.	.	.	.	24	A	0,03	8,76	AGO4	new	spe
scaffold_4:19453695-19453765	-	.	.	.	.	5	4	6	2	2	.	.	.	.	22, 22	U, U	2,51	0,0	AGO1	new	spe
scaffold_5:1865409-1865808	+	3	2	2	.	2	.	2	2	2	2	5	4	5	21	C	0,06	0,01	AGO1	new	spe
scaffold_5:1878591-1878990	+	3	2	2	.	2	.	2	2	2	2	5	4	5	21	C	0,06	0,01	AGO1	new	spe
scaffold_5:1891783-1892182	+	3	2	2	.	2	.	2	2	2	2	5	4	5	21	C	0,06	0,01	AGO1	new	spe
scaffold_5:16690551-16690906	-	.	.	.	.	.	.	.	.	.	.	.	6	5	21	U	NA	NA	NA	new	spe
scaffold_6:19200002-19200401	+	.	.	4	.	5	.	5	.	.	.	.	.	.	21	A	NA	NA	NA	new	spe
scaffold_7:1469604-1470003	-	.	.	2	.	2	.	.	5	5	.	.	.	.	21	A	NA	NA	NA	new	spe
scaffold_7:15441479-15441877	-	.	.	.	.	5	5	.	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
scaffold_7:15533002-15533397	-	.	.	.	.	5	6	.	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
scaffold_7:15624551-15624946	-	.	.	.	.	5	6	.	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
scaffold_7:22663748-22663962	+	2	2	5	5	2	2	2	.	2	.	2	2	2	21, 21	U, A	2,18	0,07	AGO1	new	spe
scaffold_1:24401254-24401526	-	.	.	.	.	5	4	.	.	.	.	.	.	.	22	A	NA	NA	NA	new	spe
scaffold_2:17517055-17517123	+	.	.	5	.	.	.	.	.	.	.	.	.	.	24	A	NA	AGO4	AGO4	new	spe
scaffold_3:7123795-7123937	-	.	.	.	.	.	.	.	.	.	6	.	.	.	21	U	NA	NA	NA	new	spe
scaffold_3:23770687-23771070	-	.	.	.	.	.	.	5	.	.	.	.	.	.	21	A	NA	NA	NA	miRN85	spe
scaffold_4:2111969-2112343	+	.	.	.	.	.	.	.	.	.	.	.	.	5	24	A	0,39	0,79	AGO4	new	spe
scaffold_5:20952465-20952612	-	.	.	.	.	.	.	.	2	5	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_7:1633113-1633490	+	.	.	4	.	.	.	6	.	.	.	.	.	.	21	A	NA	NA	NA	miRN85	spe
scaffold_1:104857-105255	-	.	.	2	2	2	.	2	5	2	.	.	.	.	22	C	NA	NA	NA	new	spe
scaffold_1:2050888-2051048	-	.	.	.	.	.	.	.	2	5	.	.	.	.	24	A	0,00	25,11	AGO4	new	spe
scaffold_1:8483709-8483899	+	6	.	.	.	.	.	.	.	.	.	.	.	.	21	U	AGO1	AGO4	AGO1 and AGO4	new	spe
scaffold_1:16092270-16092395	+	.	.	.	.	.	.	.	.	.	.	.	5	.	21	U	61,00	0,19	AGO1	new	spe
scaffold_1:16517926-16518034	-	.	.	5	.	.	.	.	2	2	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_1:17131025-17131423	+	.	.	2	.	2	.	2	6	.	.	.	.	.	22	U	AGO1	NA	AGO1	new	spe
scaffold_1:22028956-22029065	-	.	.	.	.	5	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_1:24848446-24848830	+	.	.	6	.	.	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_1:25642478-25642876	+	.	.	4	.	2	1	5	2	2	.	.	.	.	21	A	NA	NA	NA	new	spe
scaffold_1:26385368-26385733	+	.	.	5	.	.	.	.	.	.	.	.	.	.	22	A	NA	NA	NA	new	spe
scaffold_1:26453528-26453926	-	.	.	5	.	.	.	.	.	.	.	.	.	.	24	G	NA	NA	NA	new	spe
scaffold_1:26456754-26457152	-	.	.	5	.	.	.	.	.	.	.	.	.	.	24	G	NA	NA	NA	new	spe
scaffold_1:28523535-28523934	+	.	.	5	.	.	.	.	.	.	.	.	2	2	23	G	0,00	0,00	not loaded	new	spe
scaffold_1:29081097-29081174	+	.	.	.	.	.	.	.	.	2	.	2	2	6	24	A	0,00	9,44	AGO4	new	spe
scaffold_1:29374658-29375056	-	6	.	.	.	.	.	.	.	.	.	.	.	.	21	A	NA	NA	NA	new	spe
scaffold_2:1479708-1479853	-	.	.	.	.	.	.	.	.	.	.	.	5	.	23	U	NA	NA	NA	new	spe
scaffold_2:1509770-1509915	-	.	.	.	.	.	.	.	.	.	.	.	5	.	23	U	NA	NA	NA	new	spe
scaffold_2:1738771-1739074	+	.	.	5	.	.	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_2:4049697-4050080	+	.	.	5	.	.	.	.	.	.	.	.	.	.	24	U	NA	NA	NA	new	spe
scaffold_2:7874473-7874872	+	.	.	5	2	2	.	2	2	2	.	.	2	2	22	C	NA	NA	NA	new	spe
scaffold_2:12422938-12423108	-	.	.	.	.	.	.	.	.	.	.	.	5	.	22	A	NA	NA	NA	new	spe
scaffold_2:13910279-13910426	-	.	.	.	.	.	.	.	.	.	.	.	.	5	24	A	0,00	3,19	AGO4	new	spe
scaffold_2:16466475-16466860	-	.	.	6	.	.	.	.	.	.	.	.	.	.	24	G	NA	NA	NA	new	spe
scaffold_2:17667946-17668114	-	.	.	5	.	.	.	.	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
scaffold_2:17853754-17853861	+	.	.	.	.	.	.	.	.	.	.	5	2	2	24	C	0,00	0,00	not loaded	new	spe
scaffold_2:21206227-21206626	+	.	.	5	.	.	.	.	.	.	.	.	.	.	23	G	0,00	0,00	not loaded	new	spe
scaffold_3:2731720-2731799	+	.	.	.	.	.	.	.	.	.	.	.	.	6	24	A	0,00	109,06	AGO4	new	spe
scaffold_3:14874510-14874747	-	.	.	.	.	.	.	.	4	5	.	.	2	.	21	A	NA	NA	NA	new	spe
scaffold_3:14915243-14915625	+	.	.	.	.	6	.	.	.	.	.	.	.	.	24	U	10,38	0,60	AGO1	new	spe
scaffold_3:14946335-14946431	-	.	.	.	.	.	.	.	.	.	.	.	6	.	24	A	NA	NA	NA	new	spe
scaffold_3:14950388-14950484	-	.	.	.	.	.	.	.	.	.	.	.	6	.	24	A	NA	NA	NA	new	spe
scaffold_3:15431823-15432003	+	.	.	.	.	.	.	.	.	.	.	.	6	.	24	A	0,00	12,66	AGO4	new	spe
scaffold_3:20471321-20471610	-	.	.	2	.	2	5	2	.	.	.	.	.	.	21	G	NA	NA	NA	new	spe
scaffold_3:22611038-22611113	+	.	.	.	.	5	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_3:24113738-24114064	-	.	.	5	.	.	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_3:24991189-24991322	-	.	.	.	.	5	.	.	.	.	.	.	.	.	23	G	NA	NA	NA	new	spe
scaffold_3:25020616-25020749	-	.	.	.	.	5	.	.	.	.	.	.	.	.	23	G	NA	NA	NA	new	spe

scaffold_3:25378006-25378145	-	.	.	.	.	.	.	.	.	.	.	.	.	5	23	G	NA	NA	NA	new	spe
scaffold_4:1134501-1134653	-	.	.	1	.	.	.	.	.	2	.	.	2	5	24	A	0,00	0,00	not loaded	new	spe
scaffold_4:2244550-2244925	+	.	.	5	.	.	.	.	.	.	.	.	.	.	24	C	NA	NA	NA	new	spe
scaffold_4:4012343-4012741	+	.	.	5	.	.	.	.	.	.	.	.	.	.	24	A	NA	AGO4	AGO4	new	spe
scaffold_4:4243520-4243870	+	.	.	.	.	.	.	.	.	.	5	.	.	.	21	U	AGO1	AGO4	AGO1 and AGO4	new	spe
scaffold_4:5759014-5759139	+	.	.	.	.	.	.	.	.	.	.	.	5	.	21	A	0,00	0,00	not loaded	new	spe
scaffold_4:5926696-5926913	+	.	.	6	.	.	.	.	.	.	.	.	.	.	21	C	NA	NA	NA	new	spe
scaffold_4:6410080-6410476	-	5	.	.	.	.	.	.	.	.	.	.	.	.	24	G	0,00	0,93	AGO4	new	spe
scaffold_4:15065215-15065578	-	.	.	.	.	2	5	.	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
scaffold_4:16498956-16499158	-	.	.	5	.	.	.	.	.	.	.	.	.	.	23	G	0,00	0,50	AGO4	new	spe
scaffold_4:19113094-19113290	+	.	.	1	.	.	.	5	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_4:21583722-21583808	-	.	.	.	.	.	.	.	.	.	.	.	5	.	24	A	NA	NA	NA	new	spe
scaffold_4:22637354-22637479	-	6	.	.	.	.	.	.	.	.	.	.	.	.	24	G	NA	NA	NA	new	spe
scaffold_5:3615724-3615792	-	.	.	.	.	.	.	.	.	.	.	.	5	.	24	A	NA	NA	NA	new	spe
scaffold_5:3627143-3627457	-	.	.	5	.	.	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_5:4644413-4644481	+	.	2	.	.	4	5	.	.	.	2	.	2	.	24	A	NA	AGO4	AGO4	new	spe
scaffold_5:5453168-5453491	+	.	.	.	.	.	.	.	.	.	5	.	.	.	24	G	0,00	0,02	AGO4	new	spe
scaffold_5:5658530-5658927	+	.	.	.	.	.	.	2	.	5	.	.	2	.	24	A	NA	NA	NA	new	spe
scaffold_5:6308589-6308987	+	.	4	.	.	.	.	.	.	.	5	.	2	2	24	A	NA	NA	NA	new	spe
scaffold_5:16412702-16413066	-	.	2	.	.	2	.	.	5	.	2	.	2	.	22	A	NA	NA	NA	new	spe
scaffold_5:16546711-16547110	-	.	.	.	.	.	.	.	.	.	.	.	6	.	24	A	0,00	0,01	AGO4	new	spe
scaffold_5:16803509-16803907	+	.	.	5	.	.	.	.	.	.	.	.	.	.	23	U	NA	NA	NA	new	spe
scaffold_5:18127603-18127778	-	.	.	1	.	6	.	.	.	.	.	.	2	.	22	C	NA	NA	NA	new	spe
scaffold_5:18418097-18418172	-	.	.	.	.	.	.	.	.	.	5	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_5:22957570-22957712	-	.	.	.	.	.	.	.	2	3	.	5	2	2	24	A	0,00	22,18	AGO4	new	spe
scaffold_6:5014118-5014217	-	.	.	.	.	.	.	.	.	.	.	.	5	1	24	A	NA	NA	NA	new	spe
scaffold_6:9834211-9834288	-	.	.	.	.	5	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_6:9881009-9881086	+	.	.	.	.	.	.	.	4	5	.	.	.	.	23	U	NA	NA	NA	new	spe
scaffold_6:10858441-10858663	-	.	.	3	.	.	.	.	.	.	.	.	5	.	24	A	NA	NA	NA	new	spe
scaffold_6:11849171-11849404	+	.	.	.	.	.	.	.	.	.	.	.	.	5	24	A	0,00	0,00	not loaded	new	spe
scaffold_6:12219596-12219990	+	.	.	.	.	5	.	.	.	.	.	.	.	.	24	A	0,00	5,02	AGO4	new	spe
scaffold_6:12681864-12682260	-	.	.	.	.	.	.	5	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
scaffold_6:13937716-13937894	-	.	.	.	.	.	.	.	.	.	.	.	5	2	24	U	NA	NA	NA	new	spe
scaffold_6:20648239-20648638	+	.	.	5	.	.	.	.	.	.	.	.	.	.	24	A	0,00	0,54	AGO4	new	spe
scaffold_6:21418259-21418550	-	5	.	3	.	.	.	.	.	.	.	.	.	.	23	A	NA	NA	NA	new	spe
scaffold_6:23473510-23473637	+	5	.	2	.	.	.	4	2	2	.	2	2	2	21	G	0,00	0,19	AGO4	new	spe
scaffold_7:8423433-8423526	+	.	.	.	.	.	.	.	.	.	.	.	5	.	23	U	0,00	0,00	not loaded	new	spe
scaffold_7:9495429-9495827	-	.	.	2	2	.	2	5	4	1	.	.	.	.	24	U	NA	NA	NA	new	spe
scaffold_7:15416768-15417167	-	.	.	.	.	5	.	.	.	.	.	.	.	.	24	U	10,38	0,60	AGO1	new	spe
scaffold_7:15466182-15466580	-	.	.	.	.	2	5	.	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
scaffold_7:15494080-15494379	-	.	.	.	.	.	.	.	.	.	.	.	6	.	24	C	NA	NA	NA	new	spe
scaffold_7:15502596-15502995	-	.	.	.	.	6	.	.	.	.	.	.	.	.	24	U	10,38	0,60	AGO1	new	spe
scaffold_7:15557736-15558134	-	.	.	.	.	2	5	.	.	.	.	.	.	.	21	U	NA	NA	NA	new	spe
scaffold_7:15585642-15585941	-	.	.	.	.	.	.	.	.	.	.	.	6	.	24	C	NA	NA	NA	new	spe
scaffold_7:15594160-15594559	-	.	.	.	.	6	.	.	.	.	.	.	.	.	24	U	10,38	0,60	AGO1	new	spe
scaffold_7:17598243-17598325	+	.	.	.	.	.	.	.	.	.	5	.	4	.	24	G	NA	NA	NA	new	spe
scaffold_7:25545153-25545551	+	2	4	2	.	2	2	5	2	.	2	2	2	2	22	G	0,00	0,0	not loaded	new	spe
scaffold_7:27664077-27664326	-	.	.	.	.	5	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_7:28814323-28814720	+	.	.	.	.	.	.	.	.	.	5	.	2	.	24	U	4,05	0,00	AGO1	new	spe
scaffold_8:1140302-1140499	+	.	.	6	.	.	.	.	.	.	.	.	4	2	24	A	NA	NA	NA	new	spe
scaffold_8:2813155-2813553	-	.	.	.	.	5	4	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_8:3286664-3286814	-	.	.	.	.	5	.	.	.	.	.	.	.	.	24	A	NA	NA	NA	new	spe
scaffold_8:6396965-6397164	-	.	.	5	.	.	.	.	.	.	.	.	.	.	23	U	NA	NA	NA	new	spe
scaffold_8:18176264-18176489	-	.	.	.	.	.	.	.	.	.	6	.	.	.	24, 24	G, A	0,00	0,00	not loaded	new	spe
scaffold_8:19828479-19828558	+	.	.	5	.	.	.	.	.	.	.	.	.	.	24	C	NA	NA	NA	new	spe
scaffold_8:23443160-23443369	-	5	.	4	.	.	.	.	.	.	2	.	.	.	21	A	NA	NA	NA	new	spe

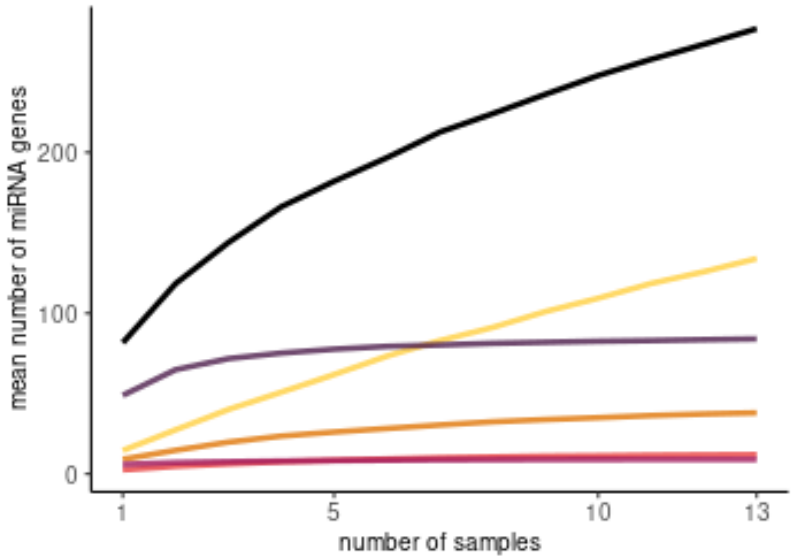


Figure S2. Completeness of the miRNA gene repertories according to the numbers of individuals sampled in *A. lyrata*.

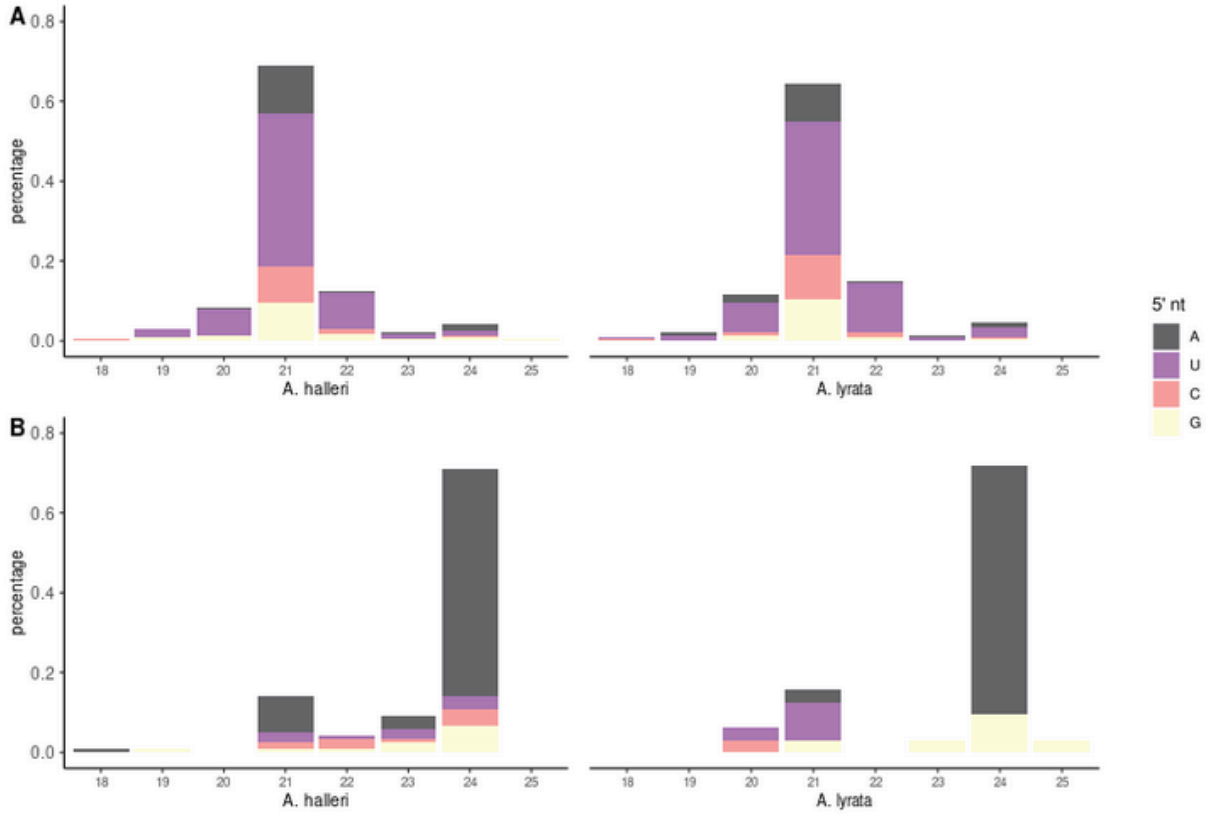


Figure S3. Size distribution and nature of the 5'nt of AGO1(a) **and AGO4** (b) -associated miRNAs in *A. halleri* (left) and *A. lyrata* (right).

Table S6. sRNAseq Datasets for miRNA predictions in the Brassicaceae family.

Species	Tissu	Genome assembly version	Study for sRNAseq libraries	SRA-NCBI	Sequencing technology	Total reads number	Number of miRNA genes	Number of mature miRNA
A. thaliana	Leaves	TAIR10	Wang et al. 2022	SRR17231451	Illumina	9,747,585	127	145
A. thaliana	Roots	TAIR10	Blein et al. 2020	SRR8723396	Proton	15,832,411	160	199
A. thaliana	Seedlings	TAIR10	Choi et al. 2021	SRR15082674	Illumina	87,270,697	122	154
A. thaliana	Seedlings	TAIR10	Choi et al. 2021	SRR15082675	Illumina	98,587,809	138	182
A. thaliana	Seedlings	TAIR10	Choi et al. 2021	SRR15082676	Illumina	67,931,288	106	135
A. thaliana	Buds	TAIR10	submitted soon	SRR27110146	Illumina	181,078,255	130	155
A. thaliana	Buds	TAIR10	submitted soon	NA	Illumina	190,492,470	162	190
A. thaliana	Buds	TAIR10	submitted soon	SRR27110143	Illumina	149,714,781	129	159
Camelina sativa	Leaves	GCF_000633955.1_Cs	Poudel et al. 2015	SRR1736515	Illumina	9,856,027	164	173
Capsella rubella	Leaves	GCF_000375325.1_Caprub1_0	Smith et al. 2014	SRR942635	Illumina	24,194,069	118	126
Raphanus sativus	Leaves	GCF_000801105.1_Rs1.0	Yang et al. 2019	SRR7725716	Illumina	15,239,556	152	182
Brassica oleracea	Leaves	Boleraceacapitata_446_v1.0	Lukasik et al. 2013	SRR799357	Illumina	24,037,208	91	96
Brassica rapa	Leaves	GCF_000309985.2_CAAS_Brap_v3.02	Ahmed et al. 2020	SRR11092574	Illumina	25,331,960	147	166
Brassica napus	Leaves	GCF_000686985.2_Bra_napus_v2.0	Regmi et al. 2021	SRR13071038	Illumina	22,377,118	164	180
Brassica juncea	Leaves	GCA_018703725.1_ASM1870372v1	Cao et al. 2016	SRR3441529	Illumina	12,365,840	156	184
Brassica nigra	Leaves	GCA_016432835.1_Bnig_sang_1.1	Ghani et al. 2014	SRR1592476	Illumina	10,284,599	133	142
Eutrema salsugineum	Leaves	GCF_000478725.1_Eutsalg1_0	Niederhuth et al. 2016	SRR3286330	Illumina	9,602,054	78	85

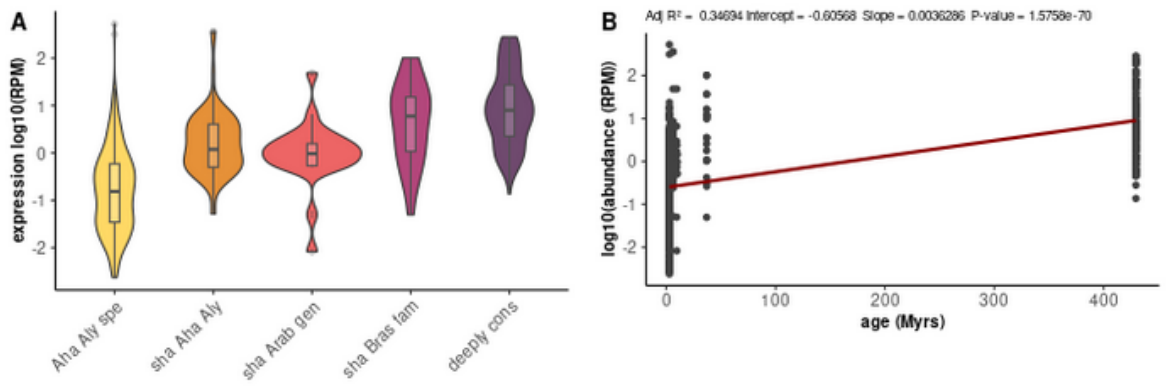


Figure S4. Mature miRNA expression according to their conservation. (a) mature miRNA abundance (RPM) according to age of the gene. (b) Linear regression of the mature miRNA gene expression according to the age.

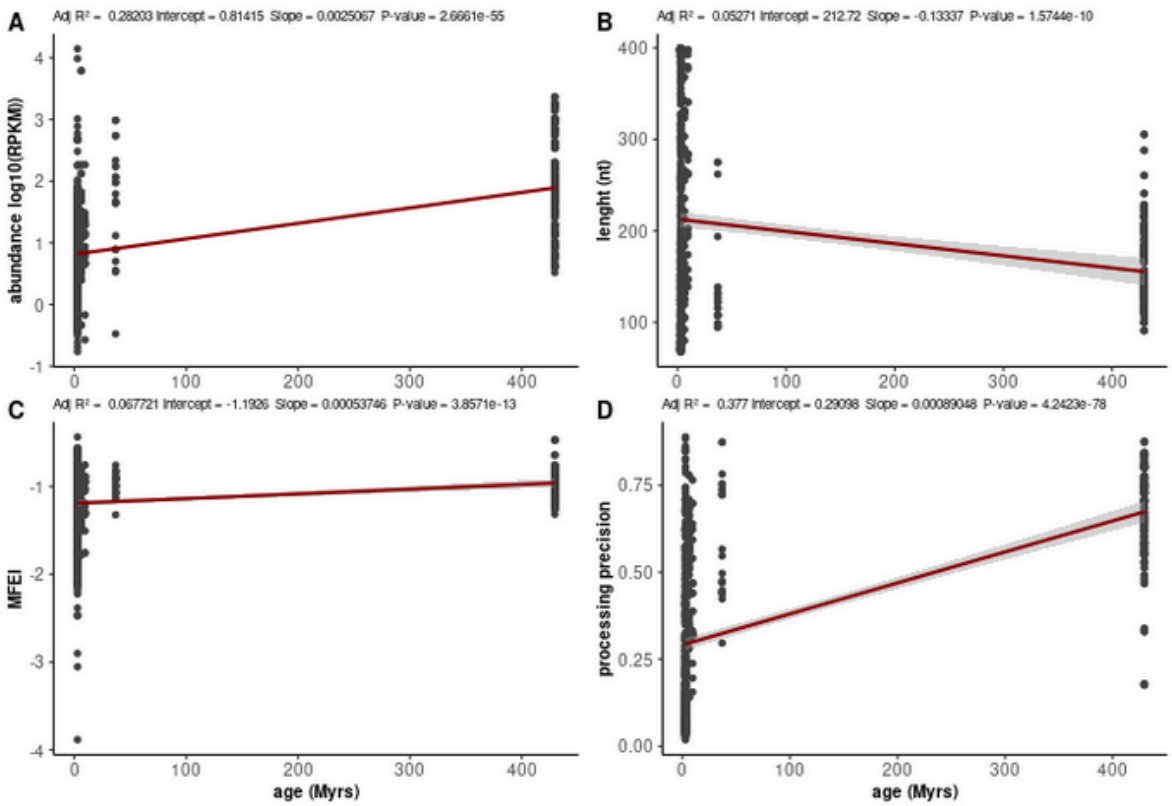


Figure S5: Linear regression of the miRNA genes characteristics according to their age. (a) precursor abundance (RPKM) according to age. (b) predicted hairpin length (c) hairpin stability (MFEI) (d) processing precision.

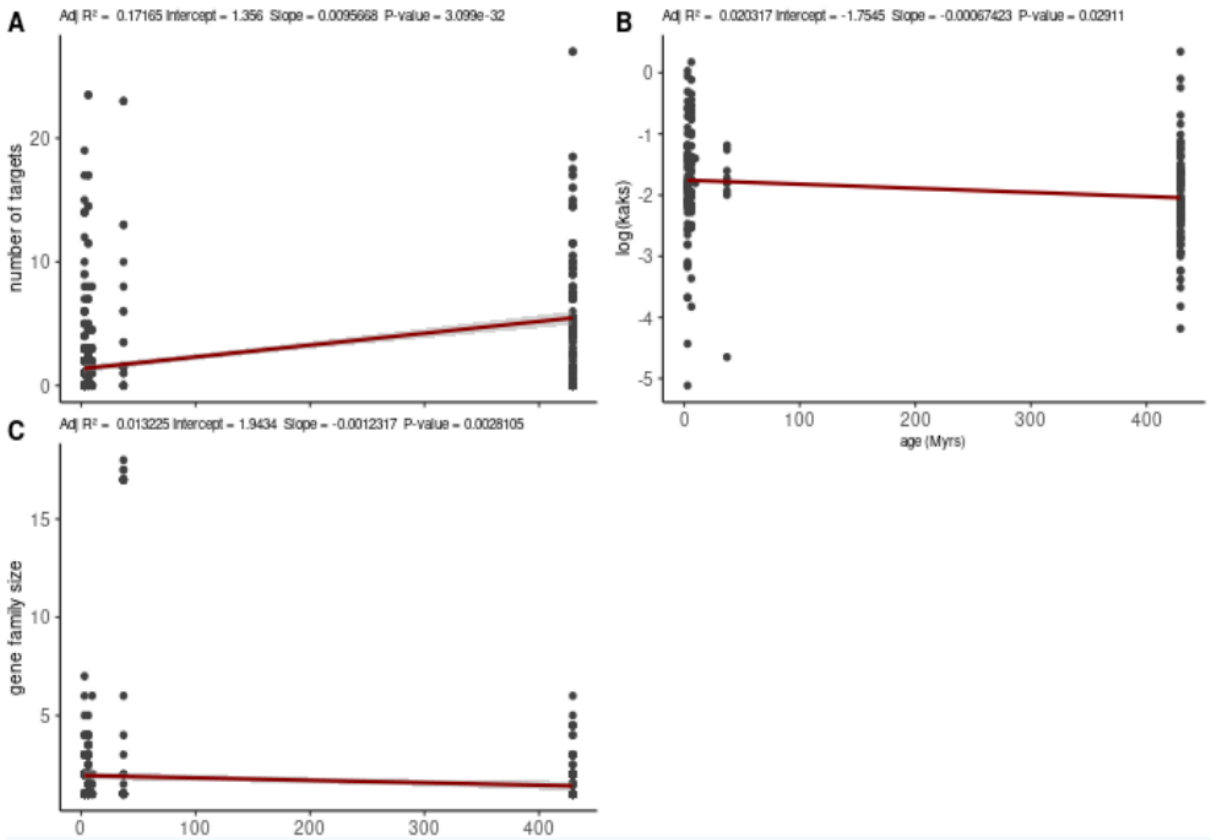


Figure S6. Linear regression of the miRNA genes target characteristics according to their age. (a) Number of targets (b) log of ka/ks ratios (c) log of the family size.

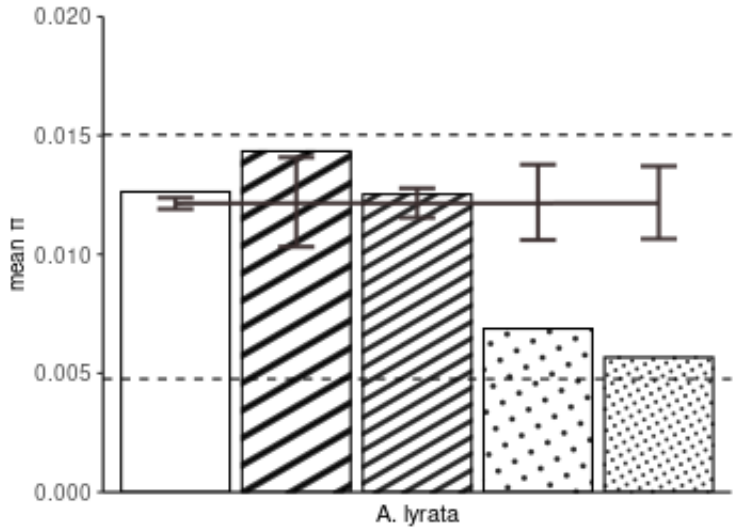


Figure S7. The miRNA/miRNA* duplex is strongly constrained by natural selection. Average nucleotide diversity for the different parts of the miRNA hairpins and upstream and downstream flanking regions (200 bp each) in *A. lyrata.* The dashed lines represent the mean π value for the 0 fold (lower bar) and 4 fold (upper bar) degenerate positions of all genes. The bars represent the 95% confidence interval obtained by random permutation of nucleotides for 1,000 random permutations under the hypothesis of a uniform distribution of polymorphisms along the sequence.

Table S7. GPS coordinates of the plant material collected for this study.

Species	Accession name	GPS Coordinates
A. halleri	Auby (France)	50.40400191 3.091988208
A. halleri	PL22 (Poland)	50.282800, 19.478717
A. halleri	I9 (Italy)	46.73141, 11.43292
A. halleri	I30 (Italy)	45.991119, 10.272050
A. lyrata	Plech (Germany)	49.627550, 11.511536
A. lyrata	CP99	
A. lyrata	LPT (USA)	-80.3875, 42.5797222
A. lyrata	TC (USA)	-81.51750000000001, 45.2416667
A. lyrata	TSS (USA)	-81.58388889999999, 45.1925
A. lyrata	PIN (USA)	-81.8313889, 43.26888890000001
A. lyrata	RON (USA)	-81.8463889, 42.2613889
A. lyrata	IND (USA)	-87.0422175, 41.6689766

Table S8. Comparison of assemblies statistics.

	Necat	SmartDenovo			Flye		
Readset	Full	Full	Filtlong	Longest	Full	Filtlong	Longest
Total length (Mb)	323	327	224	224	280	303	295
Number of contigs	509	1209	707	707	4 205	2 956	2 740
N50 (Kb)	1 597	871	770	770	204	231	254
L50	57	86	64	64	347	332	292
Average contig size (Kb)	634	270	317	317	67	103	107
Merqury score	24.4648	23.0902	21.4396	19.8713	24.727	23.2474	23.208
Complete universal single-copy orthologs	C:98.4% S:59.4% D:39.0%	C:98.5% S:76.9% D:21.6%	C:95.1% S:87.3% D:7.8%	C:93.0% S:87.0% D:6.0%	C:99.1% S:77.7% D:21.4%	C:99.0% S:76.8% D:22.2%	C:98.6% S:73.8% D:24.8%
Fragmented universal single-copy orthologs	0.6%	0.6%	1.3%	1.6%,	0.5%	0.5%	0.6%
Missing universal single-copy orthologs	1.0%	0.9%	3.6%	5.4%	0.4%	0.5%	0.5%

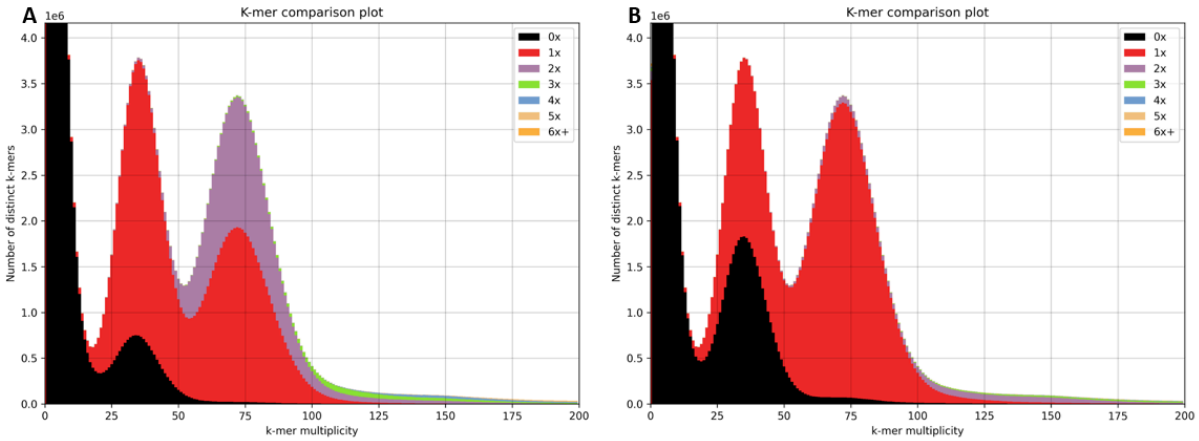


Figure S9. KAT plot. (a) Pre-Haplomerger2. (b) Post-Haplomerger2

CHAPTER II

Scenarios for the emergence of new miRNA genes in *A. halleri*.

Authors

Flavia Pavan[1], Eléanore Lacoste[2], Jean-Marc Aury[2], Vincent Castric[1], Sylvain Legrand[1].

[1] Univ. Lille, CNRS, UMR 8198 - Evo-Eco-Paleo, F-59000 Lille, France

[2] Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France

Author for correspondence : sylvain.legrand@univ-lille.fr

Abstract

MicroRNAs (miRNAs) are among the main players in regulation of gene expression. However, the relative contribution of the different sources of origin are little studied. Here we analyzed the process of emergence of 310 *A. halleri*-specific miRNA genes. Our homology search indicates that the miRNA genes emerge from diverse sources of origin including protein-coding genes, transposable elements and preexisting miRNA genes. Interestingly, only a few could have emerged from protein-coding genes, while almost half of the miRNA genes could have emerged from transposable elements. Particularly MITE, Mariner and Harbinger, TE superfamilies seem to be important contributors to new miRNA gene emergence. We further documented the recent expansion of a miRNA family which is supposed to be derived from MuDR elements and the duplication of miRNA genes formed by two hAT transposons.

Key words: miRNAs, sources of origin, duplication, transposable elements, Arabidopsis.

1. Introduction

Understanding how new genes are formed is essential to explain the genetic basis of the origin and evolution of new phenotypes. A first possible scenario is that new genes can be formed by the modification of sequences that were already genes, and implies various mechanisms such as gene duplication followed by divergence, gene fusion/fission (after a deletion or an insertion of a genomic region), horizontal gene transfer, *i.e.* vertical transmission of gene between two individuals, or reverse transcription of mature messenger RNAs and integration into the genome giving rise to a copy of the parental gene devoid of introns (retroposed gene) (Van Oss and Carvunis, 2019). The formation of new genes by gene duplication is a predominant mechanism in plants with 65% of annotated genes that have a duplicate copy. Of these, most derive from whole genome duplications and/or polyploidization events, *i.e.* the multiplication of a complete chromosome set of a certain species, which occurred multiple times in plant evolution (Panchy et al., 2016). Yet, it is now becoming clear that novel genes can also be formed *de novo* from previously non-genic regions that have gained the ability to be transcribed. The mechanisms of *de novo* gene birth are less understood. *De novo* genes can emerge from non-genic region that gain an open reading frame (ORF) and transcription, the creation of a new ORF from a previous one, but in a different frame or exonization, *i.e.* the creation of a new exon by alternative splicing of an intronic region. Although examples of *de novo* gene birth have been found in various lineages, the extent to which they arise is still debated due to the fact that *de novo* genes are likely to emerge and be lost more frequently than genes emerging from duplication of an ancestral gene (Van Oss and Carvunis, 2019). Transposons represent an important source of genomic novelty and can play an important role in *de novo* gene evolution. Indeed, their insertion in a new genomic region can create new promoters for existing ORFs, form new exons in pre-existing genes, form a new host gene if they are inserted close to a promoter (Etchegaray et al., 2021).

MicroRNAs (miRNAs) are a class of small non-coding RNAs that have emerged as key regulators of gene expression in eukaryotes. miRNAs genes are transcribed by RNA polymerase II into precursors presenting a hairpin-like structure (Bartel et al., 2004; Voinnet, 2009; Rogers and Chen, 2013). In plants, miRNAs are produced through the processing of the hairpin-like precursor by DICER-LIKE (DCL) proteins. Mature miRNAs are 20-24 nucleotides long and can downregulate gene expression by binding to ARGONAUTE proteins resulting in either mRNA degradation or translation inhibition (Zhan and Meyers, 2023). In plants, while some miRNA genes are deeply conserved, the majority of miRNAs

are lineage-specific and thus are evolutionarily young (Fahlgren et al., 2010; Cuperus et al., 2011; Chávez Montes et al., 2014; Guo et al., 2022a; Chapter 1).

Similar to coding genes, new miRNA genes can be formed from ancestral genes or have a *de novo* origin. The duplication of an existing miRNA gene can expand the family from which it originated. These processes can involve whole-genome duplication, segmental duplication and tandem duplication (Sun et al., 2012). Subsequent processes of genomic diversification may follow, resulting in sub-functionalization or neo-functionalization. For example, the miR166 family in Arabidopsis expanded from whole-genome duplication, segmental duplication and tandem duplication, followed by tissue-specific subfunctionalization (Maher et al., 2006), as miR169, miR395 and miR845 families in Brassicaceae that are tandemly organized and supposed to originate from tandem duplications (Rathore et al., 2016). Two key steps are required for the *de novo* origination of miRNA genes: the creation of a hairpin-like structure and the acquisition of a promoter that makes the transcription of the proto-miRNA gene possible. Three hypotheses that could lead to this dual acquisition have been proposed (Nozawa et al., 2012; Cui et al., 2017; Baldrich et al., 2018). The observation in *Arabidopsis thaliana* of an extended similarity between the precursor sequences of some young miRNA genes and their corresponding target gene transcripts led to the hypothesis that they directly originate from duplications of their target genes (Allen et al., 2004). This hypothesis was later supported by specific examples in Arabidopsis (Fahlgren et al., 2010; He et al., 2014; Zhang et al., 2016), *Fragaria vesca* (Xia et al., 2015), Solanaceae (de Vries et al., 2015), Antirrhinum (Bradley et al., 2017) and Vitis (Lu et al., 2019). Finally, a recent spectacular example of horizontal gene transfer leading to miRNA gene formation has been reported in the parasitic plant *Cuscuta campestris* (Yang et al., 2019). The cellular communication between *C. campestris* and its host allows the host DNA to be incorporated into the parasitic plant genome and later give rise to hairpin-like structure by duplication (Johnson et al., 2019). In the simple model of miRNA origination proposed by Allen et al., (2004), a nearly perfect hairpin emerges after the inverted duplication of a portion of a coding gene. This precursor initially exhibits near perfect complementarity, and can produce small interfering RNAs (siRNAs). Over evolutionary time, the accumulation of mutations in the hairpin structure disturbes its complementarity, facilitating recognition by the canonical DCL1 and production of miRNAs (Allen et al., 2004; Voinnet et al., 2009; Baldrich et al., 2018). While this model is particularly elegant, as it explains how new miRNA genes can gain targeting capacity right upon their inception, the proportion of new miRNA genes actually arising through this mechanism remains generally unknown.

A second source of new miRNA genes involves stem-loop sequences derived from transposable elements (TEs). This scenario of origination was supported by observations in

various species including *Oryza sativa, A. thaliana* (Piriyapongsa and Jordan, 2008; Li et al., 2011; Sun et al., 2012), *Populus trichocarpa*, *Sorghum bicolor* (Sun et al., 2012), wheat (Poretti et al., 2019; Crescente et al., 2022) and more widely in Angiospermes (Guo et al., 2022b). TEs can be classified into class I retrotransposons, which use RNA intermediates to replicate in genomes, and class II DNA transposons, which use a “cut and paste” mechanism. Class I retrotransposons are further subdivided into long terminal repeat (LTR) retroelements such as Copia and Gypsy, and non-LTR retroelements such as LINE and SINE (Mhiri et al., 2022). Under this model, the formation of hairpin precursors can occur through the juxtaposition of two inverted LTR copies of cognate TEs. The insertion of these TEs into protein coding genes allows the transcription of hairpin precursors potentially leading to the production of siRNAs, some of which may subsequently evolve into *bona fide* miRNAs (Li et al., 2011). Class II DNA transposons include various families such as Mariner, Harbinger and MuDR harboring terminal inverted repeats (TIRs) (Mhiri et al., 2022). Miniature inverted-repeat TEs (MITEs) are a special type of Class II non-autonomous element appearing as a privileged family of transposons capable of generating new miRNA genes in some plant species. Indeed, MITEs, being truncated derivatives of autonomous Class II DNA transposons, have a short length (generally 50–800 bp long) and present terminal inverted repeats (TIRs), making them ideal for hairpin precursor production. In addition, the adjacent position to a protein-coding gene of MITEs, provides to MITE-derived miRNA genes the transcriptional activity required for their expression (Guo et al., 2022b; Pegler et al., 2023). In *Oryza sativa,* up to 80% of TE-derived miRNA genes are believed to derive from MITEs (Li et al., 2011). More recently, a study analyzed representatives from 21 species spanning from green algae to angiosperm and suggested that 16.2% of the miRNAs loci studied may be derived from MITEs (Guo et al., 2022b). The processing of these MITE-derived precursors by DCL proteins generates 21- or 24-nt miRNAs, which are loaded into AGO proteins and possibly modulate gene expression at either the transcriptional, *i.e.* AGO4-loaded 24-nt miRNAs, or post-transcriptional level, *i.e.* AGO1-loaded 21-nt miRNAs (Pegler et al., 2023).

A third hypothesis regarding the *de novo* origin of miRNA genes implies the random formation of a stem-loop structure, and the subsequent acquisition of the ability to be transcribed. Felippes et al., (2008) detected young miRNA genes in *A. thaliana* with no sequence similarity to any other region of the genome. Among them, miR823 showed homology to its orthologous region in *A. lyrata*, but this region contained two insertions, leading to a modification of the predicted secondary structure, which could explain that the homologous region in *A. lyrata* is not processed as a miRNA. However, this study did not examine the presence of a promoter in the homologous region of *A. lyrata*, so the

transcriptional status of this region in *A. lyrata* is undetermined. Shanfa Lu (2019) analyzed species of the genus Vitis and showed that miR1444 and miR12112 originated from a common ancestral POLYPHENOL OXIDASE (PPO) gene targeted by these same miRNAs. The author then proposed that the promoter sequences enabling the transcriptional activity of miR1444 and miR12112 originated from a MITE superfamily transposable element inserted upstream of the original gene.

The Arabidopsis genus represents a unique opportunity to identify the origin of miRNA genes, by allowing the comparison of closely related genomes with a well-defined recent divergence history and solid genomic resources. The possibility to identify young miRNA genes is crucial, since they are more likely than the more anciently emerged elements to still have retained the “smoking gun” of the mechanisms by which they were formed. Fahlgren et al., (2010) compared the *A. thaliana* and *A. lyrata* genomes, and identified the duplication of portions of protein-coding genes as the dominant source of new miRNA genes. However, a limitation is that the two species they studied diverged about 5 million years ago, which may be considered a quite long time relative to the rapid dynamic of emergence of new miRNA genes. In addition, the annotation of TEs in these species has since been completed (Legrand et al., 2019), with a substantially higher TE content than they used, possibly leading to an underestimation of TEs as loci of origin. Finally, the miRNA annotation they used was based on a single sequencing experiment from a single accession, limiting the detection power, especially for the evolutionarily youngest miRNA genes (see chapter I). To study more directly the process by which new miRNA genes emerge, it is thus necessary to focus on even more closely related species. In chapter I, we identified a large number (n = 310) of very recent miRNA genes, specifically unique to *A. halleri* or *A. lyrata,* which diverged less than one million years ago (Roux et al., 2011). In this study, we took advantage of this set of very young miRNA genes to explore the mutational process(es) by which they emerged. Comparison of the sequences of the *A. halleri*-specific miRNA precursors with the different types of possible progenitor loci (*i.e.* protein-coding gene, TEs and other miRNA genes) revealed that transposons actually represent the most important source of origin of new miRNA genes in *A. halleri*. In particular, certain TE superfamilies, such as MITE, Mariner and Harbinger, appear to contribute preferentially. We illustrate these mechanisms of formation by documenting in detail the recent expansion of a newly created miRNA family derived from a MuDR transposon, and the formation of a new miRNA by a duplication of a tandemly duplicated hAT transposon.

2. Results

2.1 The miRNA genes in A. halleri have a diversity of possible origins

As miRNA genes can originate from the duplication of an existing miRNA gene, the inverted duplication of a coding gene, or a stem-loop structure derived from a transposable element (Cui et al., 2017), a crucial prerequisite for a comprehensive understanding of whether a genomic source is favored was the complete annotation of the *A. halleri* reference genome in terms of genes, miRNAs (refer to Chapter I), and TEs (this study). The 463 miRNA precursor sequences annotated in the *A. halleri* reference genome (Auby-1) represent a mere 0.04% of the total genome space, while the coding and non-coding fractions of the 34,721 protein-coding genes represented 16.7% and 20.1%, respectively. Additionally, the 104,224 TEs sequences we annotated accounted for 31.6% of the assembly (Figure 1a). These results are in agreement with Legrand et al., (2019), who estimated (based on a different genome assembly) the TEs composition of the *A. halleri halleri* genome at 32.7%.

To determine the origin of the *A. halleri* miRNA genes, we aligned the 463 miRNA precursor sequences to the *A. halleri* reference genome. Subsequently, we cross-referenced the positions of the obtained alignments with those of miRNAs, protein-coding genes, and transposable elements annotated in the assembly. If a significant alignment (false discovery rate < 0.01), coincided with the position of any of these genetic elements, we considered that this element was the locus of origin of the miRNA precursor. The analysis resulted in significant alignments for 33 of the 92 deeply conserved miRNA precursors, 4 of the 13 miRNA precursors shared within the Brassicaceae family, 3 of the 11 miRNA precursors shared between *A. thaliana*, *A. lyrata* and *A. halleri*, and 10 of the 37 miRNA precursors shared between *A. lyrata* and *A. halleri.* Hence, only a limited fraction of the miRNA precursors in these relatively conserved miRNA precursors had significant similarity to other genetic elements throughout the genome. In contrast, we detected significant alignments for the majority (230) of the 310 *A. halleri*-specific miRNA precursors, indicating that they are indeed recent enough to have preserved a trace of their origin (Figure 1b). Among them, 12 showed similarities with CDS, 35 with intronic sequences and/or untranslated region (UTR) sequences, 12 with other miRNA genes, 147 with TEs and 24 with unannotated regions of the assembly (Figure 1c, Supplemental Table S1). Hence, we found a clear tendency for these extremely recent miRNA genes to have similarity with TE sequences, suggesting that TEs represent a major source of miRNA progenitors.

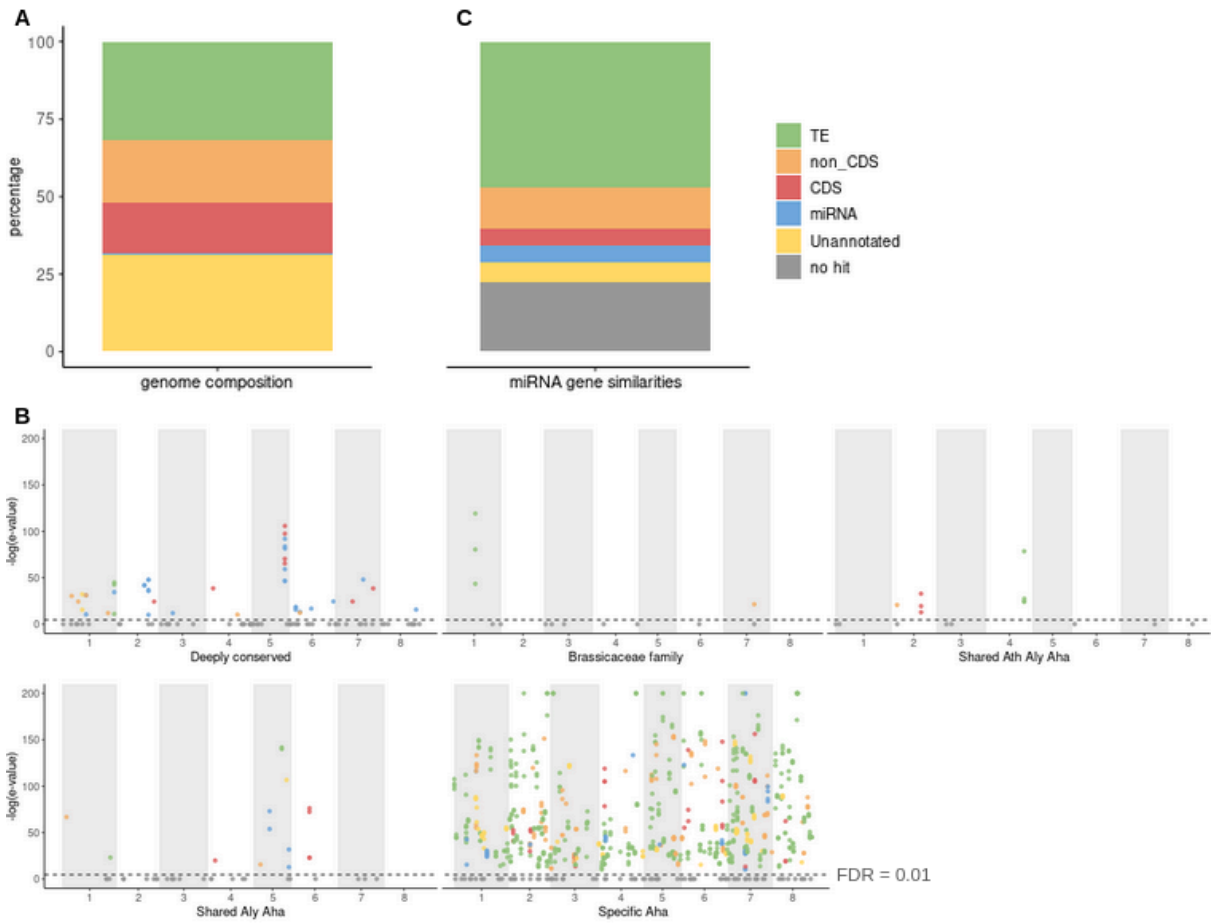


Figure 1 : miRNA genes originate from a diversity of sources. (a) Composition of the *A. halleri* reference genome. Non-CDS annotations include introns and untranslated regions. (b) Detection of miRNA precursors related loci according to their conservation status. Each miRNA precursor sequence was compared to the *A. halleri* reference genome and the BLAST e-values of the top four alignments were plotted. The color of the dots represents the nature of the miRNA-related locus (existing miRNA gene, coding gene, *i.e* CDS and non-CDS, and transposable element). The dotted horizontal line represents the false discovery rate (FDR = 0.01), where points above the line have a FDR < 0.01. (c) Relative contribution of the different sources of miRNA origin in *A. halleri*.

2.2 The MITE, Mariner and Harbinger transposon superfamilies contribute to the birth of new miRNA genes

Among TEs, the MITE superfamily was previously proposed to be favored elements in the generation of miRNA genes, as genomic sequences of this superfamily contain short inverted repeats and therefore resemble miRNA precursors (Piriyapongsa and Jordan, 2008; Li et al., 2011; Guo et al., 2022b; Pegler et al., 2023). To investigate whether this particular TE superfamily or other superfamilies predominantly contribute to the origin of new miRNA genes in *A. halleri*, we compared the proportions of each TE superfamily within the set of TEs potentially associated with new miRNA origins to those in the entire annotated set of

TEs in the genome assembly. In the *A. halleri* reference genome, we observed that the five most represented superfamilies were Gypsy (32.7%), Copia (19.8%), Helitron (10.9%), MuDR (10.9%) and LINE (9.2%) (Figure 2a). This result is consistent with Legrand et al., (2019), who had identified the same five superfamilies as the most represented in the genome of another accession (PL22) of *A. halleri*. Within the set of TEs potentially associated with new miRNA origins the most represented families were the MITE (22.8%), MuDR (18.9%), Harbinger (13.8%) and Gypsy (9.8%) (Figure 2b, Supplemental Table S1). In particular, while some superfamilies were underrepresented among TE-related miRNA genes such as Gypsy, Copia and LINE, others were overrepresented such as MITE (11-fold higher), Harbinger (6-fold higher), Mariner (2.5-fold higher), and MuDR (2-fold higher), suggesting that these superfamilies are favored contributors to miRNA origin.

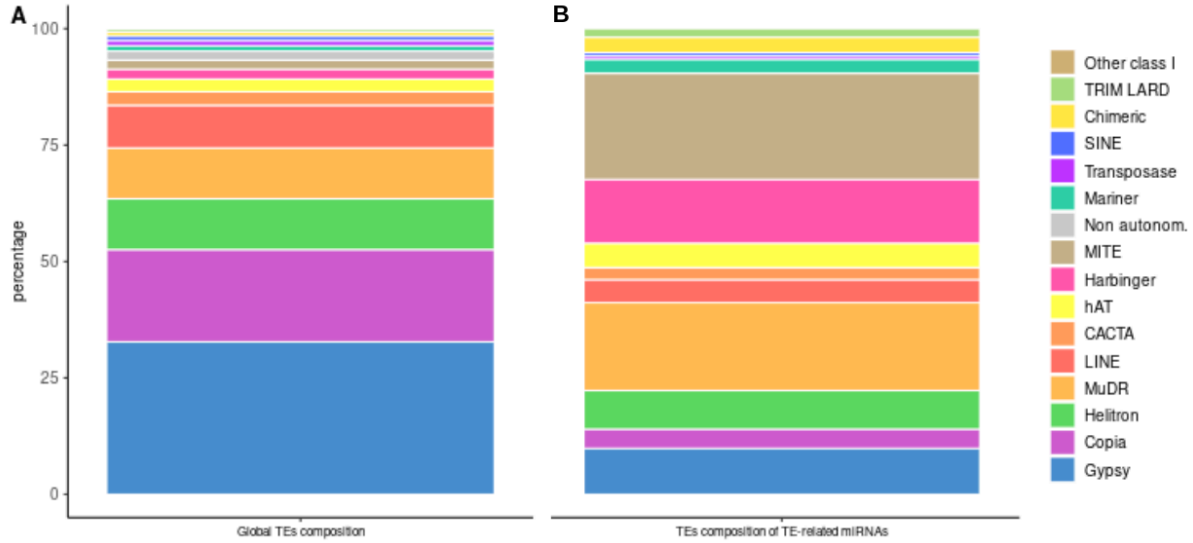


Figure 2 : Main TE families contributing to miRNA gene birth in *A. halleri*. (a) Relative proportions of the different TE superfamilies in the *A. halleri* reference genome. (b) Relative proportions of the different TE superfamilies among the TE-related miRNA genes.

We then searched for specific examples supporting the different mechanisms of emergence listed above. To do this, we first clustered the new miRNA genes into families based on precursor sequence similarity, and explored the distribution of the number of members in each family. The majority of the 310 new miRNA genes (*n* = 232) were classified as singletons (families of just one member). Forty miRNA genes formed clear families with two members, 21 formed families with three members each, and a few families contained four and five members, respectively (Supplemental Figure S1). The largest family was formed by eight miRNA precursors that we named Aha-miR13a to Aha-miR13h.

2.3 A new A. halleri-specific miRNA family derived from a MuDR transposon sequence

We first focused on the large eight-members family of *A. halleri*-specific miRNAs (the Aha-miR13a-h family). Three members of this family (Aha-miR13a, Aha-miR13c and Aha-miR13h) presented substantial similarity with TEs of the MuDR superfamily (Figure 3a), suggesting a recent expansion associated with MuDRs. The cleavage process of miRNA precursors by DCL proteins strongly depends on their secondary structure (Rojas et al., 2020). DCL1 (the canonical DCL involved in miRNA biogenesis) and DCL4 generate duplexes of 21 nt, DCL2 of 22 nt and DCL3 of 24 nt (Roger and Chen, 2013). The multiple sequence alignment revealed that Aha-miR13d, Aha-miR13f and Aha-miR13g precursors presented a deletion of 50 nucleotides, and Aha-miR13c an indel of 22 nt as compared to the other members of the family (Figure 3a). However, we observed that in spite of this relatively large indel, the secondary structure of the precursors were largely similar. Strikingly, the indel spans the miRNA-miRNA* duplex produced from Aha-miR13c, Aha-miR13d, Aha-miR13f and Aha-miR13g (Figure 3a), so the indel does not seem to affect the capacity of the precursor to be cleaved by DCL proteins. Most precursors produced a 21 nt-long mature miRNA sequence, suggesting that their biosynthesis is DCL1-dependent (Figure 3b). Only two precursors (Aha-miR13c and Aha-miR13d) produced mature miRNA sequences with a length of 23 and 24 nt, respectively, indicating that other DCL proteins may be involved in the processing of these precursors (Figure 3b). Interestingly, we observed that the mature miRNAs produced from the three precursors exhibiting the large indel were loaded in AGO4, while those produced by the other precursors were loaded in AGO1 (Chapter I; Figure 3b), suggesting a possible change of the mode of action of the family associated with the mutational event. Indeed, AGO4 is known to direct DNA methylation and thus regulate genes and transposable elements at the transcriptional level (Zhan and Meyers, 2023). Some miRNAs have been shown to be involved in the regulation of TEs. In *A. thaliana*, Borges et al., (2018) showed that one particular miRNA (miR845) is capable of regulating the transposon from which it is derived by cleaving TE transcripts, and initiating the production of epigenetically activated siRNAs (easiRNA). We thus asked if the miRNAs loaded in AGO4 could eventually target the transposon from which they originated. However, we did not predict any target on the homologous MuDR sequence for these 24-nt miRNAs. Instead, some mature miRNAs of the family are loaded in AGO1, suggesting that they could regulate mRNA targets at the post-transcriptional level. Drawing on the results of the miRNA target prediction we previously conducted (chapter I), we observed that Ah-miR13a was predicted to be able to target the coding sequence of two protein-coding genes (Ah6g725047, related to *A. thaliana* OTU1, a deubiquitinase involved in the endoplasmic

reticulum-associated degradation, Zhang et al., 2020) and Ah8g676367, homologous to a *A. lyrata* predicted protein with unknown function). Ah-miR13e was predicted to be able to target the same two genes (Ah6g725047 and Ah8g676367), plus Ah1g889505 (predicted to contain a Ribonuclease H domain related to LTR retroelements, Malik and Eickbush, 2001). Hence, members of this TE-derived miRNA family seem to have lost the capacity to target the TE from which they derive, and are instead predicted to target genes from the host genome.

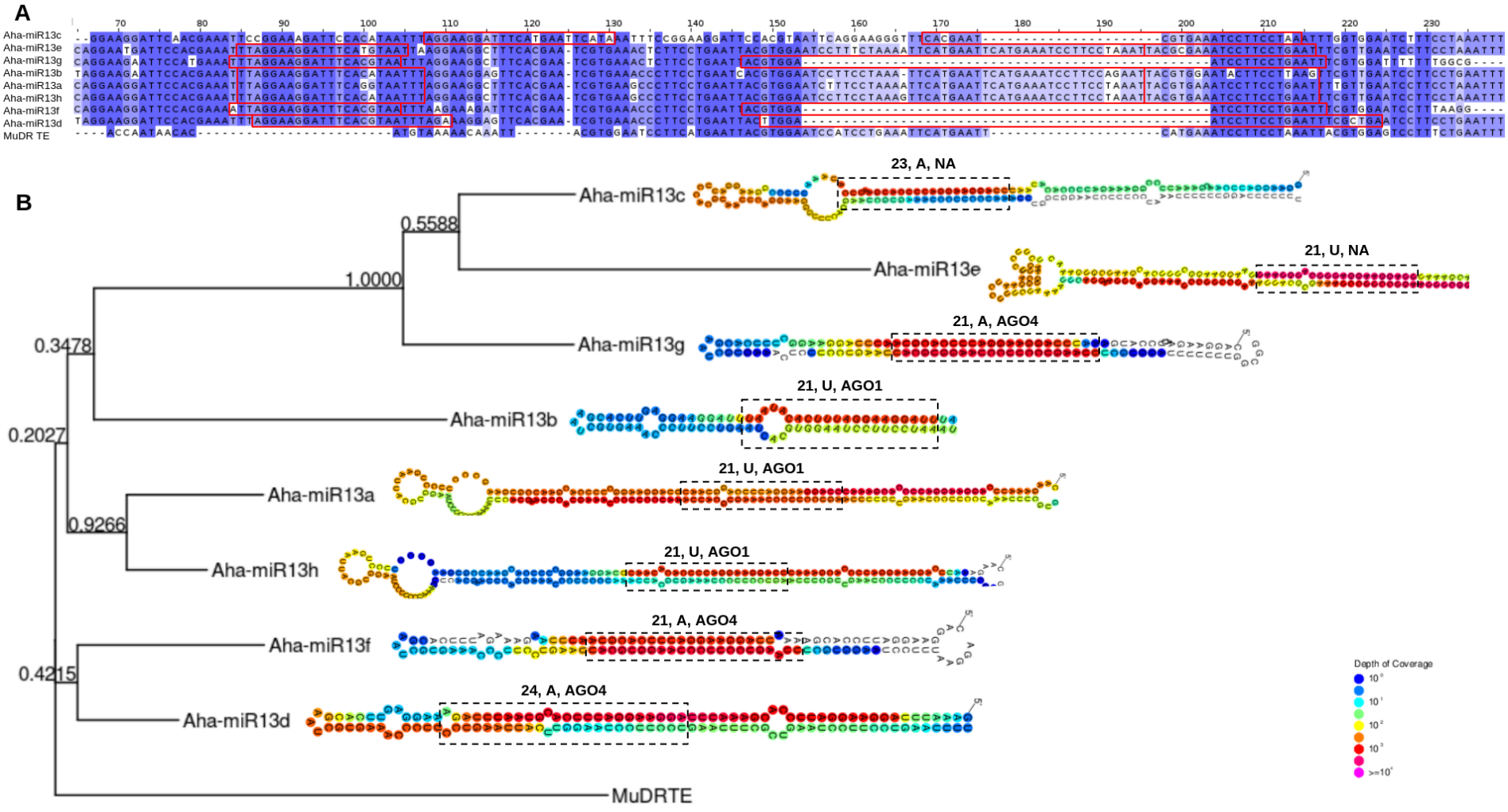

A
Aha-miR13c
Aha-miR13e
Aha-miR13g
Aha-miR13b
Aha-miR13a
Aha-miR13h
Aha-miR13f
Aha-miR13d
MuDR TE
B
23, A, NA
21, U, NA
21, A, AGO4
21, U, AGO1
21, U, AGO1
21, U, AGO1
21, A, AGO4
24, A, AGO4
0.5588
1.0000
0.3478
0.2027
0.9266
0.4215
MuDRTE
Depth of Coverage

Figure 3: Recent expansion of a new miRNA family derived from a MuDR transposon sequence. (a) Sequence alignment of the eight miRNA precursors and their homologous MuDR transposon sequence, aligned with MUSCLE and visualized with Jalview v2.11.3.2. Mature miRNA and miRNA* sequences are indicated by a red rectangle. The color of the nucleotides represents the level of conservation, *i.e*. white for poorly conserved sites and dark blue for highly conserved sites. (b) Phylogenetic relationship of miRNA precursors constructed using the neighbor-joining algorithm implemented in MEGA v11.0.13 (Tamura et al., 2007). Bootstrap values were calculated from 1,000 replicates. Alongside the phylogenetic tree, the secondary structure of miRNA precursors was generated and displayed using structVis v0.4 (github: https://github.com/MikeAxtell/strucVis), with a color scale indicating the read depth. The miRNA-miRNA* duplex predicted by mirkwood is indicated by a black rectangle, labeled with its size, 5' nucleotide and preferential AGO protein loading.

2.4 A new A. halleri-specific miRNA resulting from the tandem duplication of a hAT transposon

We then focused on a family containing three homologous miRNA precursors, Aha-miR53a, Aha-miR53b and Aha-miR53c, whose sequences overlapped with pairs of transposons from the hAT superfamily located head-to-head (Supplemental figure S2). Aha-miR53a was shared with *A. lyrata,* while Aha-miR53b and Aha-miR53c were specific to *A. halleri* (Chapter I), suggesting that they have emerged from a duplication after the divergence of the two species. The secondary structure of the three miRNA precursors, notably Aha-miR53b and Aha-miR53c, appeared fairly conserved, and all share the same mature 24-nt miRNA sequence. In addition, the mature miRNAs were loaded in AGO4 and had a predicted target site in the hAT sequence from which they originated, as well as in other transposons from the hAT superfamily (Chapter I; Figure 4a). To further understand whether the miRNA precursors originated 1) from independent juxtapositions of pairs of unrelated transposons, 2) from independent tandem duplications of initially unrelated isolated transposons or 3) from the subsequent duplication of a single initial pair of transposons, we constructed a phylogenetic tree with the six transposon sequences from which Aha-miR53a, Aha-miR53b and Aha-miR53c could have derived. In scenario 1) we expect no clear phylogenetic structure. In scenario 2) we expect that the pairs of juxtaposed transposons form three independent phylogenetic clusters, while in scenario 3) we expect that members from each of the three pairs group together in the phylogenetic tree, hence forming two separate clusters. Our results showed that the two transposons forming the miRNA precursors clearly fall into separate groups (Figure 4b), suggesting that the three members of this family represent paralogs from one single ancestral copy rather than independent juxtaposition of unrelated TEs or separate duplications of independent copies. Overall, this suggests that the

juxtaposition or tandem duplication of transposons from the hAT family can form a hairpin structure processed by DCL proteins and lead to the production of mature miRNAs. Interestingly, the sRNAs produced by these hairpins are 24nt-long, they are loaded in AGO4 and they have potential targets to regulate the transposon they originated from, as well as other members of the same superfamily (Figure 4c). These observations support the possibility that TE-related miRNA genes can originate from an inverted duplication event of a transposon or emerge from the juxtaposition of two transposons of the same family (Pegler et al., 2023).

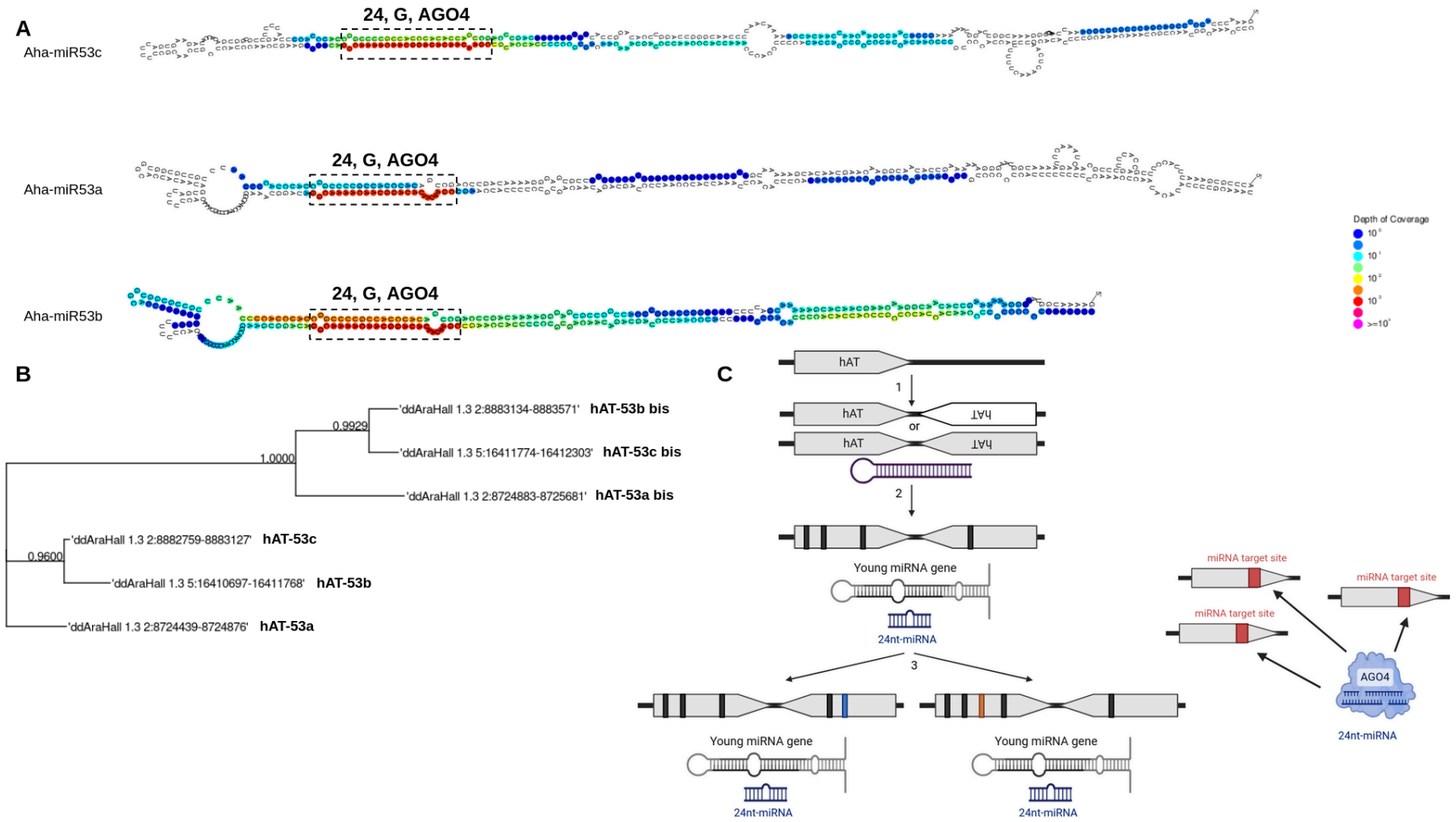

A
Aha-miR53c
24, G, AGO4
Aha-miR53a
24, G, AGO4
Aha-miR53b
24, G, AGO4
Depth of Coverage
B
'ddAraHall 1.3 2:8883134-8883571' hAT-53b bis
0.9929
'ddAraHall 1.3 5:16411774-16412303' hAT-53c bis
1.0000
'ddAraHall 1.3 2:8724883-8725681' hAT-53a bis
'ddAraHall 1.3 2:8882759-8883127' hAT-53c
0.9600
'ddAraHall 1.3 5:16410697-16411768' hAT-53b
'ddAraHall 1.3 2:8724439-8724876' hAT-53a
C
hAT
1
hAT
or
2
Young miRNA gene
24nt-miRNA
3
Young miRNA gene
24nt-miRNA
Young miRNA gene
24nt-miRNA
miRNA target site
miRNA target site
miRNA target site
AGO4
24nt-miRNA

Figure 4: Emergence of new miRNA genes from the juxtaposition of two hAT transposons. (a) The secondary structures of Aha-miR53a, Aha-miR53b and Aha-miR53c precursors were obtained using structVis v0.4 (github: https://github.com/MikeAxtell/strucVis) and the color scale indicates the read depth. The miRNA-miRNA* duplex is indicated by a black rectangle, labeled with its size, 5' nucleotide and AGO protein preferential loading. (b) Phylogenetic relationship of the six hAT transposons from which Aha-miR53a, Aha-miR53b and Aha-miR53c originate. The tree was constructed using MUSCLE and the neighbor-joining algorithm of MEGA v11.0.13, and bootstrap values were calculated from 1,000 replicates. (c) Model of emergence of new miRNA genes from hAT transposons. 1) The inverted duplication or juxtaposition of hAT transposons enables the formation of a hairpin structure. 2) Over the course of evolution, the hairpin is processed by DCL proteins and produces 24 nt-miRNAs. These 24 nt-miRNAs are loaded in AGO4 proteins and have predicted target sites on the DNA sequence of the transposon from which they originate, as well as on those of other transposons from the same family. 3) Duplication of the region to another location in the genome expands the miRNA family.

2.5 New miRNA genes can arise from the inverted duplication of a part of a coding gene

Allen et al., (2004) proposed that coding gene-derived miRNAs can arise from the reverse duplication of part of a coding gene or from a reverse intralocus duplication followed by direct duplication, and several examples of these processes were reported by Fahlgren et al., (2010) in Arabidopsis. We thus looked whether some of the *A. halleri*-specific miRNA genes could have arisen this way. To illustrate this possibility, we focused on sequence similarity between Aha-miR1111 and the Ah7g813868 gene (homologous to a F-box/RNI superfamily protein coding gene in *A. thaliana,* a large gene family with diverse roles in cell cycle transition, transcriptional regulation and signal transduction, Kuroda et al., 2002), which are both located on chromosome 7. A recent duplication of this gene giving rise to miRNA genes has been documented in *Fragaria Vesca* (Xia et al,. 2015). To characterize the duplication event behind the origination of Aha-miR1111, we compared the extended sequence of the miRNA precursor (1000 nucleotides on each side of the loop) with the sequence of the coding gene from which it may have emerged. The precursor sequence was very similar to a 66 bp region in the first exon of Ah7g813868, duplicated directly, and to the same region extended to 263 bp duplicated in reverse orientation (Figure 5). This suggests that Aha-miR1111 is derived from an inverted duplication of part of the Ah7g813868 gene to form a hairpin structure. We observed that the mature miRNA produced by the precursor (21 nt, loaded in AGO1) had a predicted target site in the first exon of the gene from which it was derived (Figure 5), suggesting its potential to bind to the mRNA and to regulate its expression.

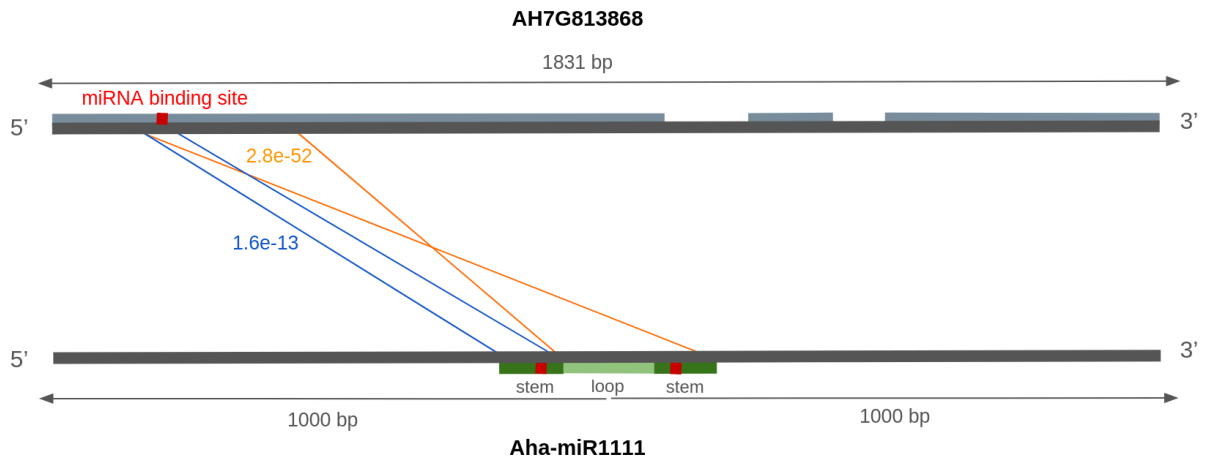


Figure 5: Inverted duplication of a portion of the coding sequence of a gene giving rise to a new miRNA gene. Sequence alignment between a region of 2000 bp centered on the loop of the miRNA precursor Aha-miR1111 and the Ah7g813868 gene using YASS (Noe and kuchero, 2005). The three exons of the gene are represented in grey and the miRNA predicted binding site in red. The stem and the loop of the precursor are indicated in two shades of green and the miRNA and miRNA* in red. The e-value of the direct duplication is indicated in blue and the one of the inverted duplication in orange.

3. Discussion

3.1 The role of transposons in the emergence of miRNA genes

Novel miRNA genes have been proposed to arise from the duplication of protein-coding genes, from transposable elements, from the duplication of preexisting miRNAs or from a region of genome able to form hairpin precursor that gain the ability to be transcribed (Nozawa et al., 2012; Cui et al., 2017; Baldrich et al., 2018). Fahlgren et al., (2010) concluded that, in *A. thaliana* and *A. lyrata,* the majority of new miRNA genes derive from protein-coding genes, with only a few TE-related miRNAs. Here, we identified a much lower fraction of miRNA genes related to protein-coding genes identified here (15.2%) than the one reported in Fahlgren et al., (2010) (85.5%). In contrast, we evaluated that up to 47% of the *A. halleri*-specific miRNA genes have emerged from transposons, which is much higher than the one estimated by Fahlgren et al., (2010) (2.9%). The differences observed may be due to several reasons. First, in this study we were able to track the loci of origin of a large number of very young miRNA genes annotated from several accessions, while Fahlgren et al., (2010) predicted miRNA genes from a single sequencing experiment from a single accession, which is expected to result in a low power to detect the most recent category of miRNA genes (chapter I). Second, the comparison of *A. thaliana* and *A. lyrata* is inherently

limited by the relatively ancient divergence between these two species, as compared to our comparison of a pair of more closely related species, leading to the identification of relatively more ancient miRNA genes which could have lost the trace of their locus of origin. In addition, TE sequences accumulate more mutations than protein-coding genes as they are less constrained by natural selection. Thus, studying more ancient miRNA genes can lead to a bias in the identification of their locus of origin in favor of protein-coding genes. Third, transposons, due to their repetitive nature, are challenging to identify. While Fahlgren et al., (2010) used a library based on *A. thaliana* alone, our TEs annotation in the *A. halleri* Auby1 genome is based on a more comprehensive library of TEs built by combining the repeat contents from *A. thaliana*, *A. lyrata* and *A. halleri* (Legrand et al., 2019). It is thus likely that Fahlgren et al., (2010) missed a number of sequence similarities between miRNAs and TEs. Fourth, the TE content in *A. thaliana* and *A. lyrata* is slightly lower than that of *A. halleri*. Legrand et al., (2019) used an unbiased estimation procedure (remapping of raw illumina reads on a joint TE library combined across the three species) and estimated the TE content in *A. thaliana* and *A. lyrata* at around 19.1% and 25.2% respectively, while in *A. halleri* it was up to 32.7% and 30.2% in the two accessions they sequenced. The estimation in our newer reference genome (31.6%) is consistent with a higher TE content in *A. halleri*, which may also contribute to explain why we identified TEs as such predominant progenitors of miRNA genes. This difference in the level of details of the annotation of TEs could also explain why the proportion of MITE-related miRNA genes is much higher in our study (8% overall) than the one identified by Fahlgren et al., (2010) in *A. thaliana* and *A. lyrata* (2.9%) and in Guo et al., (2020) with 2.5% (7/275) and 4.9% (18/363) in *A. lyrata* and *A. thaliana* respectively, or even by Zhang et al., (2011), who identified no miRNA genes originating from MITE.

3.2 Interplay between transcriptional and post-transcriptional silencing pathways

The regulation of gene expression and gene silencing can occur at two levels. Transcriptional gene silencing (TGS) allows a control of gene expression through directed methylation of the gene, while post-transcriptional gene silencing (PTGS) allows a control of gene expression through directed degradation of the RNA molecules. The siRNAs are main actors of TGS while miRNAs typically function in PTGS. However, crosstalks between these two pathways have been proposed to be common and could provide a fine tuning of gene expression over the short and long terms. In chapter I we showed that “young” miRNA genes tend to be produced from more “perfect” hairpins, to have a length of 24-nt, and be loaded in AGO4. Here, we show that they disproportionately stem from TE-related sequences. These features are reminiscent of the TGS pathway. Over time, the accumulation of mutations in

the precursor sequence lead to the production of 21-nt miRNA loaded in AGO1, suggesting a progressive evolution toward the PTGS pathway. Accordingly, we observed that the Aha-miR53 precursors potentially issued from the duplication of hAT transposons produced 24-nt miRNA that are loaded in AGO4 and target the transposons from which they originated. In contrast, the Aha-miR13 precursors, which probably derived from MuDR transposons, produced mainly 21-nt miRNA loaded in AGO1 and AGO4. These miRNAs do not target the transposon from which they originate, and instead have gained the ability to target sites in the CDS of protein-coding genes in *trans*. These two examples highlight the possible transition from the TGS to the PTGS pathways over the course of evolution : new miRNA loci would predominantly be derived from TE-related sequences, and would initially be neutral or may participate in the regulation of their progenitor TE, as suggested by Borges et al., (2018). The accumulation of mutations along the miRNA sequence may eventually abolish the capacity to target the TE of origin, and confer the capacity to target genes from the host genome. Depending on the functional importance of the newly targeted gene such targeting may be neutral or deleterious or, under rare circumstances it may be beneficial and retained over the long run, eventually shifting to a PTGS regulatory pathway. Transitions between PTGS and TGS have been observed in other contexts. For instance, Mari-Ordonez et al., (2013) proposed that the *de novo* silencing of TEs involves an immediate PTGS response based on 21-22 nt siRNA, which is later replaced after a number of generations (11 generations in their experiment) by a more stable long-term TGS repression involving 24-nt sRNA molecules. Hence, TGS and PTGS mechanisms may act in concert and eventually replace each other over short and long time scales, albeit eventually in opposite orders.

3.3 Evolution of miRNA targeting

The repertoire of targets of a miRNA gene can change over the course of evolution due to the accumulation of mutations in the miRNA sequence or in the mRNA sequence. In chapter I we observed that the number of mRNA targets tends to increase over the course of evolution, with an average of 5.4 targets for the most ancient miRNA genes, versus only 0.9 predicted targets for the *A. halleri*-specific miRNA genes. While TEs and protein-coding genes represent approximately similar overall fractions of the total genome, in this study, we observed that almost half of the very young miRNA genes originate from transposons, with a very low proportion deriving from duplications of protein-coding genes. This helps to understand why the number of targets tends to increase over the course of evolution. Indeed, a miRNA gene that emerges from a protein-coding gene can immediately target the gene from which it originates. In most cases this may have deleterious consequences for the

individual carrying this new miRNA gene, leading to its rapid elimination by natural selection. In contrast, a miRNA gene that emerges from a transposon is able to target the transposon from which it originates. Such a miRNA gene may be either neutral or slightly beneficial, since active transposons can invade the genome with deleterious consequences for the host. Thus, a miRNA emerging and leading to a reduction of the expression of the transposon can be retained by natural selection, at least for some time. This extension of the residence time of newly emerged miRNA genes may provide opportunity for the acquisition of new target sites in protein-coding gene through the apparition of mutations in miRNA sequence or through the insertion of the transposon of origin in a protein-coding gene leading to the creation of a new exon (exonization), a new gene or a new promoter (Etchegaray et al., 2021).

4. Material and methods

TEs annotations in the genome of *A. halleri*

TEs in the *A. halleri* Auby1 genome assembly were annotated using Repeatmasker (v ≥ 4.14, Smit, A.; Hubley, R.; Green, P. RepeatMasker. Available online: http://www.repeatmasker.org) with a bundle library, composed of TEs from *A. halleri*, *A. lyrata* and *A. thaliana*, which was produced by Legrand et al., (2019). Briefly, the library was composed of consensus sequences representative of TEs identified in the three species using the *TEdenovo* pipeline of the REPET package (Quesneville et al., 2005). Each consensus sequence was then classified into TE superfamilies and repeat types using PASTEC (Hoede et al., 2014).

Identification of the progenitor loci

The 463 miRNA precursor sequences identified in *A. halleri* (chapter I) were aligned against the *A. halleri* reference genome assembly (Auby1 v1.3), with BLAST (Camacho et al., 2009). The E-values were converted to *p*-values using the relationship $p = 1\text{-}e^{-E}$, and an FDR cutoff point was determined using the R (v4.1.2; R Core Team 2023) Q-VALUE package (v1.0; Storey, 2002) (as in Fahlgren et al., 2010). Each significant alignment (FDR < 0.01) was intersected with the coding gene, the transposon and the miRNA gene annotations. The relative proportions of the different sources of origin were normalized by base pair.

Identification of miRNA families

We used BLAST to compare the sequences of the miRNA precursors to the complete set of precursors annotated in *A. halleri*. We selected alignment with at least 80% of identity, 80% of covering and E-value < 1e-10. The clusters of duplicated miRNA precursors were constructed using Cytoscape v3.10.1 (Shannon et al., 2003).

Phylogenetic tree

We aligned the precursor sequences with MUSCLE (Edgar, 2004) and constructed phylogenetic trees using the Neighbor Joining algorithms implemented in MEGA4 v11.0.13 (Tamura et al., 2007). Support for the nodes of the tree was evaluated with a bootstrap test (1,000 replicates).

Origin via coding gene duplication

A Region of 2kb centered on the loop of the miRNA precursor sequence was aligned to the DNA sequence of the coding gene from which it may have emerged using YASS (Version, Noe and kucherov, 2005). Inverted duplications of the coding gene were identified with the dotplot provided by YASS.

5. References

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.-H., Spatafora, J.W., and Carrington, J.C. (2004). Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. Nat Genet **36**: 1282–1290.

Baldrich, P., Beric, A., and Meyers, B.C. (2018). Despacito: the slow evolutionary changes in plant microRNAs. Current Opinion in Plant Biology **42**: 16–22.

Bartel DP. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell:**116**:281-97.

Bradley, D. et al. (2017). Evolution of flower color pattern through selection on regulatory small RNAs. Science **358**: 925–928.

Borges, F., Parent, J.-S., Van Ex, F., Wolff, P., Martínez, G., Köhler, C., and Martienssen, R.A. (2018). Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in Arabidopsis. Nat Genet **50**: 186–192.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics **10**: 421.

Chávez Montes, R.A., Rosas-Cárdenas, D.F.F., De Paoli, E., Accerbi, M., Rymarquis, L.A., Mahalingam, G., Marsch-Martínez, N., Meyers, B.C., Green, P.J., and De Folter, S. (2014). Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. Nat Commun **5**: 3722.

Crescente, J.M., Zavallo, D., Del Vas, M., Asurmendi, S., Helguera, M., Fernandez, E., and Vanzetti, L.S. (2022). Genome-wide identification of MITE-derived microRNAs and their targets in bread wheat. BMC Genomics **23**: 154.

Cui, J., You, C., and Chen, X. (2017). The evolution of microRNAs in plants. Current Opinion in Plant Biology **35**: 61–67.

Cuperus, J.T., Fahlgren, N., and Carrington, J.C. (2011). Evolution and Functional Diversification of *MIRNA* Genes. Plant Cell **23**: 431–442.

De Vries, S., Kloesges, T., and Rose, L.E. (2015). Evolutionarily Dynamic, but Robust, Targeting of Resistance Genes by the miR482/2118 Gene Family in the Solanaceae. Genome Biol Evol **7**: 3307–3321.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**: 1792–1797.

Etchegaray, E., Naville, M., Volff, J.-N., and Haftek-Terreau, Z. (2021). Transposable element-derived sequences in vertebrate development. Mobile DNA **12**: 1.

Fahlgren, N., Jogdeo, S., Kasschau, K.D., Sullivan, C.M., Chapman, E.J., Laubinger, S., Smith, L.M., Dasenko, M., Givan, S.A., Weigel, D., and Carrington, J.C. (2010). MicroRNA Gene Evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. The Plant Cell **22**: 1074–1089.

Fenselau De Felippes, F., Schneeberger, K., Dezulian, T., Huson, D.H., and Weigel, D. (2008). Evolution of *Arabidopsis thaliana* microRNAs from random sequences. RNA **14**: 2455–2459.

Gregory, T.R. (2005). Synergy between sequence and size in Large-scale genomics. Nat Rev Genet **6**: 699–708.

Guo, Z., Kuang, Z., Deng, Y., Li, L., and Yang, X. (2022a). Identification of Species-Specific MicroRNAs Provides Insights into Dynamic Evolution of MicroRNAs in Plants. IJMS **23**: 14273.

Guo, Z., Kuang, Z., Tao, Y., Wang, H., Wan, W., Hao, C., Shen, F., Yang, X., Li. **L.** (2022b). Miniature Inverted-repeat Transposable Elements Drive Rapid MicroRNA Diversification in Angiosperms. Molecular Biology and Evolution **39**:msac224.

He, H., Liang, G., Li, Y., Wang, F., and Yu, D. (2014). Two Young MicroRNAs Originating from Target Duplication Mediate Nitrogen Starvation Adaptation via Regulation of Glucosinolate Synthesis in *Arabidopsis thaliana*. Plant Physiol. **164**: 853–865.

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., and Quesneville, H. (2014). PASTEC: An Automatic Transposable Element Classification Tool. PLoS ONE **9**: e91929.

Johnson, N.R. and Axtell, M.J. (2019). Small RNA warfare: exploring origins and function of trans-species microRNAs from the parasitic plant Cuscuta. Current Opinion in Plant Biology **50**: 76–81.

Kuroda, H., Takahashi, N., Shimada, H., Seki, M., Shinozaki, K., and Matsui, M. (2002). Classification and Expression Analysis of Arabidopsis F-Box-Containing Protein Genes. Plant and Cell Physiology **43**: 1073–1085.

Legrand, S. et al. (2019). Differential retention of transposable element-derived sequences in outcrossing Arabidopsis genomes. Mobile DNA **10**: 30.

Li, Y., Li, C., Xia, J., and Jin, Y. (2011). Domestication of Transposable Elements into MicroRNA Genes in Plants. PLoS ONE **6**: e19212.

Lu, S. (2019). *De novo* origination of *MIRNAs* through generation of short inverted repeats in target genes. RNA Biology **16**: 846–859.

Maher, C., Stein, L., and Ware, D. (2006). Evolution of *Arabidopsis* microRNA families through duplication events. Genome Res. **16**: 510–519.

Malik, H.S. and Eickbush, T.H. (2001). Phylogenetic Analysis of Ribonuclease H Domains Suggests a Late, Chimeric Origin of LTR Retrotransposable Elements and Retroviruses. Genome Res. **11**: 1187–1197.

Marí-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O. (2013). Reconstructing de novo silencing of an active plant retrotransposon. Nat Genet **45**: 1029–1039.

Mhiri, C., Borges, F., and Grandbastien, M.-A. (2022). Specificities and Dynamics of Transposable Elements in Land Plants. Biology **11**: 488.

Noe, L. and Kucherov, G. (2005). YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Research **33**: W540–W543.

Nozawa, M., Miura, S., and Nei, M. (2012). Origins and Evolution of MicroRNA Genes in Plant Species. Genome Biology and Evolution **4**: 230–239.

Panchy, N., Lehti-Shiu, M., and Shiu, S.-H. (2016). Evolution of Gene Duplication in Plants. Plant Physiol. **171**: 2294–2316.

Pegler, J.L., Oultram, J.M.J., Mann, C.W.G., Carroll, B.J., Grof, C.P.L., and Eamens, A.L. (2023). Miniature Inverted-Repeat Transposable Elements: Small DNA Transposons That Have Contributed to Plant MICRORNA Gene Evolution. Plants **12**: 1101.

Piriyapongsa, J. and Jordan, I.K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. RNA **14**: 814–821.

Poretti, M., Praz, C.R., Meile, L., Kälin, C., Schaefer, L.K., Schläfli, M., Widrig, V., Sanchez-Vallet, A., Wicker, T., and Bourras, S. (2020). Domestication of High-Copy Transposons Underlays the Wheat Small RNA Response to an Obligate Pathogen. Molecular Biology and Evolution **37**: 839–848.

Quesneville, H., Bergman, C.M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., and Anxolabehere, D. (2005). Combined Evidence Annotation of Transposable Elements in Genome Sequences. PLoS Comp Biol **1**: e22.

Rathore, P., Geeta, R., and Das, S. (2016). Microsynteny and phylogenetic analysis of tandemly organised miRNA families across five members of Brassicaceae reveals complex retention and loss history. Plant Science **247**: 35–48.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., Mesirov, J.P. (2011). Integrative Genomics Viewer. Nature Biotechnology **29**: 24–26.

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

Rojas, A.M.L., Drusin, S.I., Chorostecki, U., Mateos, J.L., Moro, B., Bologna, N.G., Bresso, E.G., Schapire, A., Rasia, R.M., Moreno, D.M., and Palatnik, J.F. (2020). Identification of key sequence features required for microRNA biogenesis in plants. Nat Commun **11**: 5320.

Roux, C., Castric, V., Pauwels, M., Wright, S.I., Saumitou-Laprade, P., and Vekemans, X. (2011). Does Speciation between Arabidopsis halleri and Arabidopsis lyrata Coincide with Major Changes in a Molecular Target of Adaptation? PLoS ONE **6**: e26872.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. **13**: 2498–2504.

Storey, J.D. (2002). A direct approach to false discovery rates. J. R. Stat. Soc., B **64**: 479–498.

Sun, J., Zhou, M., Mao, Z., and Li, C. (2012). Characterization and Evolution of microRNA Genes Derived from Repetitive Elements and Duplication Events in Plants. PLoS ONE **7**: e34092.

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, *24*(8), 1596–1599.

Van Oss, S.B. and Carvunis, A.-R. (2019). De novo gene birth. PLoS Genet **15**: e1008160.

Voinnet, O. (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. Cell **136**: 669–687.

Xia, R., Ye, S., Liu, Z., Meyers, B.C., and Liu, Z. (2015). Novel and Recently Evolved MicroRNA Clusters Regulate Expansive *F-BOX* Gene Networks through Phased Small Interfering RNAs in Wild Diploid Strawberry. Plant Physiol. **169**: 594–610.

Yang, Z. et al. (2019). Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. Nat. Plants **5**: 991–1001.

Zang, Y., Gong, Y., Wang, Q., Guo, H., and Xiao, W. (2020). Arabidopsis OTU 1, a linkage-specific deubiquitinase, is required for ENDOPLASMIC RETICULUM -associated protein degradation. The Plant Journal **101**: 141–155.

Zhan, J. and Meyers, B.C. (2023). Plant Small RNAs: Their Biogenesis, Regulatory Roles, and Functions. Annu. Rev. Plant Biol. **74**: 21–51.

Zhang, Y., Xia, R., Kuang, H., and Meyers, B.C. (2016). The Diversification of Plant *NBS-LRR* Defense Genes Directs the Evolution of MicroRNAs That Target Them. Mol Biol Evol **33**: 2692–2705.

6. Supplementary data

Table S1: Diverse sources of origin of the 310 *A. halleri*-specific miRNA genes.

Loci	Family	Main loading	Homologous miRNA	Homologous TE	Homologous coding gene	code
ddAraHall_1.3_1:7097-7324	miR1a	AGO4	ddAraHall_1.3_4:22217543-22217924	Harbinger	685295	TE
ddAraHall_1.3_1:597043-597210	miR2	NA	.	.	.	no hit
ddAraHall_1.3_1:1110805-1110918	miR3	AGO4	.	Harbinger	.	TE
ddAraHall_1.3_1:4182510-4182588	miR4	not loaded	.	.	.	no hit
ddAraHall_1.3_1:5970293-5970366	miR5	AGO4	.	.	.	no hit
ddAraHall_1.3_1:6550675-6551066	miR6	NA	.	MuDR	.	TE
ddAraHall_1.3_1:6796987-6797284	miR7	AGO4	.	TIR_MITEov10	805159	TE
ddAraHall_1.3_1:6926780-6926909	miR8a	not loaded	ddAraHall_1.3_6:26579278-26579412	MuDR	755439	TE
ddAraHall_1.3_1:7150694-7150861	miR9a	not loaded	ddAraHall_1.3_5:20201081-20201341	.	.	other miR
ddAraHall_1.3_1:7994029-7994165	miR10a	AGO4	ddAraHall_1.3_2:19769962-19770098, ddAraHall_1.3_2:654836-654972	TIR_MITEun10	.	TE
ddAraHall_1.3_1:9114798-9115187	miR11	NA	.	.	.	no hit
ddAraHall_1.3_1:9764331-9764433	miR12	AGO1	.	.	.	no hit
ddAraHall_1.3_1:12476759-12476945	miR13a	AGO1	ddAraHall_1.3_8:5469657-5469843, ddAraHall_1.3_5:17548540-17548912, ddAraHall_1.3_3:14877193-14877356	MuDR	.	TE
ddAraHall_1.3_1:12478798-12478950	miR14	NA	.	.	.	no hit
ddAraHall_1.3_1:12617460-12617856	miR15	NA	.	.	.	unannotated
ddAraHall_1.3_1:12695938-12696235	miR16b	AGO1	ddAraHall_1.3_4:15489381-15489638, ddAraHall_1.3_4:15489381-15489638	.	871602	intron
ddAraHall_1.3_1:12944744-12945222	miR17a	AGO4	.	.	873802	intron
ddAraHall_1.3_1:12970700-12970966	miR18	AGO4	.	.	.	unannotated
ddAraHall_1.3_1:13884475-13884654	miR19	not loaded	.	.	.	unannotated
ddAraHall_1.3_1:14212519-14212831	miR20a	AGO4	ddAraHall_1.3_8:9669316-9669402	Helitron	859016	TE
ddAraHall_1.3_1:15130841-15130959	miR17b	AGO4	ddAraHall_1.3_8:20021935-20022193, ddAraHall_1.3_1:12944744-12945222, ddAraHall_1.3_8:21283278-21283560	Harbinger	.	TE
ddAraHall_1.3_1:15958417-15958786	miR22	AGO4	.	.	.	no hit
ddAraHall_1.3_1:16043382-16043504	miR23	NA	.	.	.	no hit
ddAraHall_1.3_1:16108090-16108201	miR24	AGO4	.	.	.	unannotated
ddAraHall_1.3_1:17213304-17213403	miR25	AGO1	.	.	.	no hit
ddAraHall_1.3_1:17280778-17280979	miR26	NA	.	.	.	unannotated
ddAraHall_1.3_1:18940212-18940293	miR27a	AGO4	.	.	674156	intron
ddAraHall_1.3_1:18949129-18949201	miR27b	AGO4	ddAraHall_1.3_1:18965051-18965123, ddAraHall_1.3_1:18940212-18940293	.	.	other miR
ddAraHall_1.3_1:18965051-18965123	miR27c	AGO4	ddAraHall_1.3_1:18949129-18949201, ddAraHall_1.3_1:18940212-18940293	.	.	other miR
ddAraHall_1.3_1:21082592-21082878	miR28a	NA	ddAraHall_1.3_7:5351752-5352040, ddAraHall_1.3_7:5351752-5352040	TIR_MITEun10	881658	TE
ddAraHall_1.3_1:24193763-24193839	miR29	not loaded	.	SINE	784477	TE
ddAraHall_1.3_1:25661705-25661859	miR30	AGO4	.	.	.	no hit
ddAraHall_1.3_1:26496842-26496995	miR31	AGO1	.	.	.	no hit
ddAraHall_1.3_1:26743281-26743351	miR32	NA	.	Helitron	.	TE
ddAraHall_1.3_1:27770076-27770170	miR33	NA	.	.	.	no hit
ddAraHall_1.3_1:28527238-28527331	miR34	AGO4	.	MuDR	.	TE
ddAraHall_1.3_1:29121439-29121514	miR35	NA	.	hAT	.	TE
ddAraHall_1.3_1:30231157-30231299	miR36	AGO4	.	.	.	unannotated
ddAraHall_1.3_1:30923864-30924036	miR37	not loaded	.	TIR_MITEov10	820024	TE
ddAraHall_1.3_2:654836-654972	miR38a	AGO4	ddAraHall_1.3_2:19769962-19770098, ddAraHall_1.3_1:7994029-7994165	hAT	.	TE
ddAraHall_1.3_2:1040654-1040838	miR38b	AGO1	.	.	.	no hit
ddAraHall_1.3_2:1079998-1080127	miR39	NA	.	TRIM_LARD	.	TE
ddAraHall_1.3_2:1165514-1165830	miR40	NA	.	Helitron	788505	TE
ddAraHall_1.3_2:1353780-1353919	miR41	AGO1	.	.	.	no hit
ddAraHall_1.3_2:1400850-1401248	miR42	AGO4	.	TIR_MITEov10	.	TE
ddAraHall_1.3_2:1657040-1657266	miR43	AGO1	.	Helitron	788816	TE
ddAraHall_1.3_2:2407586-2407833	miR44	NA	.	.	.	no hit
ddAraHall_1.3_2:2574865-2575068	miR45	NA	.	.	677304	cds
ddAraHall_1.3_2:2941148-2941306	miR46	NA	.	Copia	700791	TE
ddAraHall_1.3_2:3242554-3242810	miR47	NA	.	.	.	no hit
ddAraHall_1.3_2:3829873-3829970	miR48	NA	.	.	.	no hit
ddAraHall_1.3_2:4234679-4234933	miR49	NA	.	MuDR	674287	TE
ddAraHall_1.3_2:4540022-4540180	miR50	AGO4	.	Helitron	793216	TE
ddAraHall_1.3_2:7018253-7018341	miR51a	not loaded	ddAraHall_1.3_6:16512027-16512115	Tase	.	TE
ddAraHall_1.3_2:8104715-8104794	miR52	NA	.	.	.	no hit
ddAraHall_1.3_2:8489772-8489964	miR13b	AGO1	ddAraHall_1.3_8:5469657-5469843	MuDR	792015	TE
ddAraHall_1.3_2:8724691-8725066	miR53a	AGO4	ddAraHall_1.3_2:8882974-8883276, ddAraHall_1.3_5:16411558-16411950	hAT	.	TE
ddAraHall_1.3_2:8882974-8883276	miR53b	AGO4	ddAraHall_1.3_2:8724691-8725066, ddAraHall_1.3_5:16411558-16411950	hAT	.	TE
ddAraHall_1.3_2:12282574-12282660	miR54a	AGO1 AGO4	ddAraHall_1.3_2:12387224-12387308	.	.	other miR
ddAraHall_1.3_2:12297552-12297638	miR54b	AGO1	ddAraHall_1.3_2:12282574-12282660	.	796708	intron
ddAraHall_1.3_2:12305160-12305231	miR55	not loaded	.	.	796662	cds
ddAraHall_1.3_2:12387224-12387308	miR54c	AGO1 AGO4	ddAraHall_1.3_2:12282574-12282660	.	.	other miR
ddAraHall_1.3_2:12442239-12442523	miR56	AGO4	.	.	.	no hit
ddAraHall_1.3_2:12521689-12521770	miR57	not loaded	.	Gypsy	.	TE
ddAraHall_1.3_2:12728661-12728791	miR58	NA	.	.	817482	cds
ddAraHall_1.3_2:13979848-13980108	miR59	NA	.	.	797914	intron
ddAraHall_1.3_2:14131833-14132128	miR60a	NA	ddAraHall_1.3_8:11214846-11215102	classII	698843	TE
ddAraHall_1.3_2:14643263-14643372	miR61	AGO1	.	.	710050	intron
ddAraHall_1.3_2:14761433-14761596	miR62	not loaded	.	MuDR	.	TE
ddAraHall_1.3_2:16239944-16240336	miR63	not loaded	.	TIR_MITEov10	.	TE
ddAraHall_1.3_2:16897508-16897631	miR64	AGO1	.	.	.	no hit
ddAraHall_1.3_2:17223468-17223620	miR65	AGO4	.	Harbinger	831121	TE
ddAraHall_1.3_2:17356391-17356790	miR66	NA	.	Harbinger	.	TE
ddAraHall_1.3_2:17773794-17773864	miR67	NA	.	LINE	.	TE
ddAraHall_1.3_2:17819501-17819577	miR68	NA	.	.	.	no hit
ddAraHall_1.3_2:18985108-18985352	miR69a	not loaded	ddAraHall_1.3_7:23491711-23492022	.	865948	intron
ddAraHall_1.3_2:19336888-19337286	miR70	NA	.	TIR_MITEun10	789501	TE
ddAraHall_1.3_2:19769962-19770098	miR39c	AGO4	ddAraHall_1.3_2:654836-654972, ddAraHall_1.3_1:7994029-7994165	TIR_MITEun10	.	TE
ddAraHall_1.3_2:20183160-20183232	miR71	AGO4	.	SINE	.	TE
ddAraHall_1.3_2:20577686-20578077	miR72	AGO1	.	.	806442	intron
ddAraHall_1.3_2:21045619-21045873	miR73	AGO1	.	LINE	855977	TE
ddAraHall_1.3_2:21366920-21367085	miR74	AGO4	.	.	889572	intron
ddAraHall_1.3_2:22168204-22168422	miR75	NA	.	.	.	no hit

ddAraHall_1.3_2:22274527-22274687	miR76	AGO4	.	TIR_MITEun10	.	TE
ddAraHall_1.3_2:22344000-22344399	miR77a	AGO4	ddAraHall_1.3_5:11286743-11287141	MuDR	.	TE
ddAraHall_1.3_2:22914344-22914527	miR78	AGO1	.	.	.	no hit
ddAraHall_1.3_2:24134806-24134876	miR79	NA	.	LINE	.	TE
ddAraHall_1.3_3:50804-50900	miR80	not loaded	.	.	.	no hit
ddAraHall_1.3_3:380856-381100	miR81	AGO1 and AGO4	.	.	756269	intron
ddAraHall_1.3_3:489526-489599	miR82	AGO4	.	.	.	no hit
ddAraHall_1.3_3:1181997-1182232	miR83	AGO1	.	.	.	no hit
ddAraHall_1.3_3:1519720-1520111	miR84a	AGO1	ddAraHall_1.3_4:21628295-21628686, ddAraHall_1.3_8:14169929-14170322	TIR_MITEun10	.	TE
ddAraHall_1.3_3:2522442-2522528	miR85	NA	.	TIR_MITEun10	.	TE
ddAraHall_1.3_3:3245548-3245945	miR86	AGO4	.	.	.	no hit
ddAraHall_1.3_3:4038590-4038701	miR87	AGO4	.	.	.	unannotated
ddAraHall_1.3_3:4742349-4742429	miR88	AGO4	.	Mariner	.	TE
ddAraHall_1.3_3:5327411-5327644	miR89a	AGO4	ddAraHall_1.3_8:13768769-13769168	TIR_MITEun10	.	TE
ddAraHall_1.3_3:6876400-6876756	miR90	AGO1	.	.	769079	intron
ddAraHall_1.3_3:7131758-7131914	miR91	AGO1	.	.	.	no hit
ddAraHall_1.3_3:7285651-7285798	miR92	NA	.	.	.	no hit
ddAraHall_1.3_3:7289322-7289456	miR93	NA	.	.	.	no hit
ddAraHall_1.3_3:8354471-8354564	miR94a	NA	ddAraHall_1.3_8:9660035-9660128	Helitron	.	TE
ddAraHall_1.3_3:9066414-9066813	miR95	AGO1	.	.	772753	intron
ddAraHall_1.3_3:10162772-10162901	miR96	AGO4	.	MuDR	.	TE
ddAraHall_1.3_3:10898566-10898829	miR97	NA	.	.	.	unannotated
ddAraHall_1.3_3:13770887-13770961	miR98	AGO4	.	.	753650	intron
ddAraHall_1.3_3:14399138-14399208	miR99	NA	.	.	859091	cds
ddAraHall_1.3_3:14523035-14523106	miR1000	NA	.	Harbinger	.	TE
ddAraHall_1.3_3:14877193-14877356	miR13c	NA	ddAraHall_1.3_8:5469657-5469843, ddAraHall_1.3_1:12476759-12476945, ddAraHall_1.3_5:17548540-17548912	MuDR	752359	TE
ddAraHall_1.3_3:15036899-15037002	miR1001	AGO4	.	.	.	unannotated
ddAraHall_1.3_3:16074864-16074958	miR1002	AGO4	.	MuDR	681064	TE
ddAraHall_1.3_3:21006817-21006966	miR1003	AGO4	.	TIR_MITEun10	.	TE
ddAraHall_1.3_3:21433207-21433591	miR1004	AGO4	.	TRIM_LARD	708138	TE
ddAraHall_1.3_3:24698510-24698595	miR1005	AGO4	.	.	.	no hit
ddAraHall_1.3_3:26426776-26426927	miR1006	AGO1	.	.	.	no hit
ddAraHall_1.3_4:2533-2620	miR1007	AGO4	.	.	.	unannotated
ddAraHall_1.3_4:16548-16634	miR1008	AGO1	.	.	.	unannotated
ddAraHall_1.3_4:1692116-1692201	miR1009	AGO4	.	MuDR	.	TE
ddAraHall_1.3_4:2193128-2193206	miR1010	NA	.	MuDR	.	TE
ddAraHall_1.3_4:2237258-2237333	miR1011	not loaded	.	.	.	unannotated
ddAraHall_1.3_4:2694679-2694776	miR1012	NA	.	hAT	.	TE
ddAraHall_1.3_4:2951034-2951433	miR1013	AGO1	.	Gypsy	.	TE
ddAraHall_1.3_4:3115375-3115469	miR1014	AGO1	.	MuDR	.	TE
ddAraHall_1.3_4:3363683-3364082	miR1015	AGO4	.	.	887927	cds
ddAraHall_1.3_4:3467940-3468052	miR1016	AGO1	.	.	.	no hit
ddAraHall_1.3_4:3584586-3584733	miR1017a	AGO4	ddAraHall_1.3_7:19048671-19048831	TIR_MITEov10	.	TE
ddAraHall_1.3_4:3873421-3873551	miR13d	AGO4	ddAraHall_1.3_6:23493380-23493500, ddAraHall_1.3_7:13680752-13680872, ddAraHall_1.3_8:5469657-5469843, ddAraHall_1.3_1:12476759-12476945, ddAraHall_1.3_5:17548540-17548912	MuDR	.	TE
ddAraHall_1.3_4:4223984-4224073	miR118	NA	.	.	.	no hit
ddAraHall_1.3_4:4969248-4969316	miR119	NA	.	classII	.	TE
ddAraHall_1.3_4:6772792-6772950	miR120	NA	.	Helitron	.	TE
ddAraHall_1.3_4:12455573-12455769	miR121	AGO1	.	Gypsy	.	TE
ddAraHall_1.3_4:12505209-12505461	miR122	AGO1	.	.	.	no hit
ddAraHall_1.3_4:15489381-15489638	miR16b	AGO1 and AGO4	ddAraHall_1.3_1:12695938-12696235, ddAraHall_1.3_1:12695938-12696235	.	871602	intron
ddAraHall_1.3_4:17255266-17255628	miR1023	NA	.	MuDR	879859	TE
ddAraHall_1.3_4:17391815-17391955	miR1024	not loaded	.	classII	.	TE
ddAraHall_1.3_4:18812318-18812402	miR1025	NA	.	.	.	no hit
ddAraHall_1.3_4:19314274-19314366	miR1026a	AGO4	ddAraHall_1.3_7:11583047-11583142	Helitron	.	TE
ddAraHall_1.3_4:19820795-19821194	miR1027	AGO1	.	.	.	unannotated
ddAraHall_1.3_4:19972603-19973002	miR1028	AGO4	.	.	.	no hit
ddAraHall_1.3_4:20889608-20889681	miR1029	AGO4	.	hAT	.	TE
ddAraHall_1.3_4:21169354-21169457	miR1030	NA	.	.	.	no hit
ddAraHall_1.3_4:21249542-21249798	miR1031	NA	.	.	.	no hit
ddAraHall_1.3_4:21628295-21628686	miR84b	NA	ddAraHall_1.3_3:1519720-1520111, ddAraHall_1.3_8:14169929-14170322	TIR_MITEun10	758610	TE
ddAraHall_1.3_4:22217543-22217924	miR1032a	AGO4	ddAraHall_1.3_8:18031457-18031627	Harbinger	876021	TE
ddAraHall_1.3_4:25587214-25587350	miR1033	not loaded	.	.	.	unannotated
ddAraHall_1.3_4:25672130-25672298	miR1034	NA	.	.	.	no hit
ddAraHall_1.3_5:799702-799812	miR1035	AGO1	.	.	.	no hit
ddAraHall_1.3_5:4033392-4033464	miR1036	AGO4	.	.	.	no hit
ddAraHall_1.3_5:4082060-4082306	miR1037	AGO4	.	LINE	.	TE
ddAraHall_1.3_5:4743481-4743756	miR17c	AGO4	ddAraHall_1.3_1:12944744-12945222	.	873802	intron
ddAraHall_1.3_5:5100924-5101225	miR1039	NA	.	MuDR	798834	TE
ddAraHall_1.3_5:5187768-5187868	miR1040	AGO1	.	.	.	no hit
ddAraHall_1.3_5:5595979-5596075	miR1041a	AGO4	ddAraHall_1.3_6:3717935-3718007	Harbinger	730011	TE
ddAraHall_1.3_5:6706449-6706540	miR1042	AGO4	.	Chimeric	.	TE
ddAraHall_1.3_5:6888153-6888471	miR1043	AGO1 and AGO4	.	CACTA	.	TE
ddAraHall_1.3_5:7407494-7407615	miR1044	NA	.	.	.	no hit
ddAraHall_1.3_5:7503658-7504019	miR1045a	AGO1	ddAraHall_1.3_6:13609057-13609365	.	846915	intron
ddAraHall_1.3_5:7728499-7728592	miR1046	AGO4	.	.	.	unannotated
ddAraHall_1.3_5:8031750-8031935	miR1047	AGO1	.	.	.	no hit
ddAraHall_1.3_5:8032006-8032190	miR1048	AGO1 and AGO4	.	.	.	no hit
ddAraHall_1.3_5:8782553-8782730	miR1049	AGO1	.	.	.	no hit
ddAraHall_1.3_5:8867567-8867760	miR1050	AGO4	.	Mariner	.	TE
ddAraHall_1.3_5:9431151-9431336	miR1051	AGO1	.	.	.	unannotated
ddAraHall_1.3_5:10212744-10212846	miR1052	not loaded	.	.	796238	intron
ddAraHall_1.3_5:11286743-11287141	miR77b	AGO1 and AGO4	ddAraHall_1.3_2:22344000-22344399	MuDR	.	TE
ddAraHall_1.3_5:11765922-11766032	miR1053	AGO4	.	Copia	.	TE

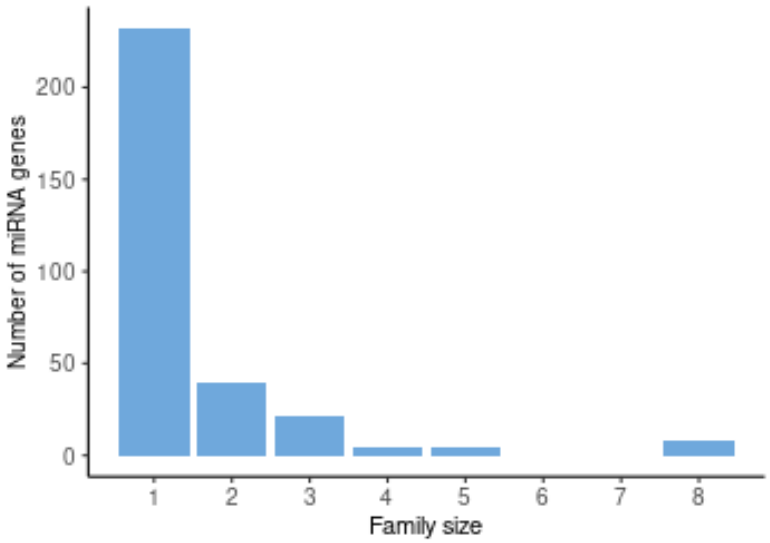
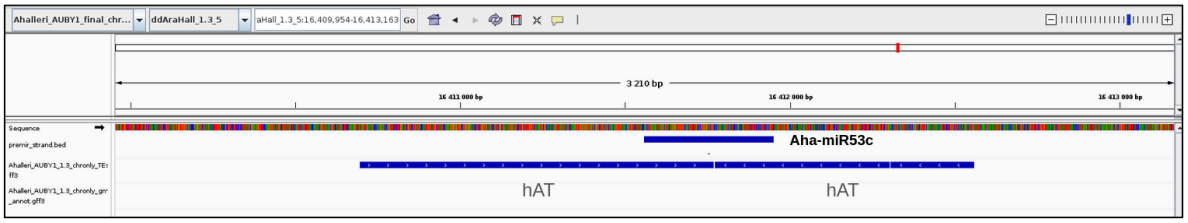
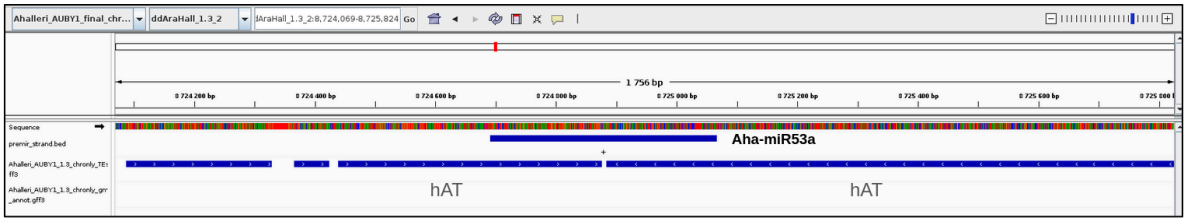


Figure S1: Distribution of family size of specific miRNA genes in *A. halleri*.

ddAraHall_1.3_5



ddAraHall_1.3_2



ddAraHall_1.3_2

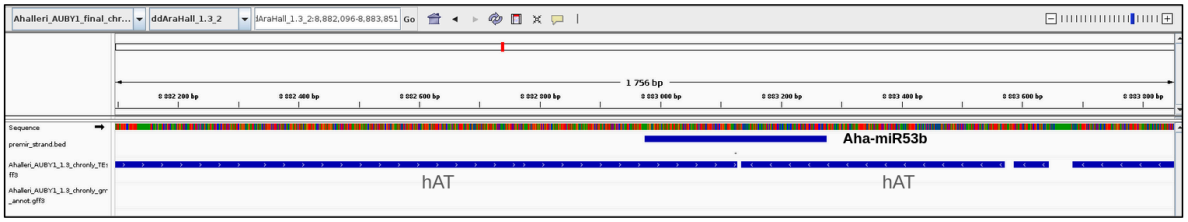


Figure S2: Aha-miR53 precursor sequences overlap hAT transposons. Genomic localizations of Aha-miR53a, Aha-miR53b and Aha-miR53c and hAT transposons are visualized with IGV v.2.13.0 (Robinson et al., 2011).

Conclusions and perspectives

In this thesis, I first performed a deep annotation of miRNA genes in a new reference genome of *A. halleri*, and I evaluated the level of conservation of these miRNA genes at increasing phylogenetic scale. I further investigated the processes by which proto-miRNAs eventually integrate the "canonical" miRNA regulatory network over short and long evolutionary times and characterized how natural selection acts on variation of these miRNA genes and their mRNA targets (Figure 1). Secondly, I took advantage of the identification of a large number of *A. halleri*-specific miRNA genes to identify the loci from which they have emerged. In particular I focused on three main hypotheses for the origin of miRNA genes, including coding genes, transposable elements and preexisting miRNA genes (Figure 1). However, some issues remain and new questions have arisen. I will now describe these new problems and outline potential avenues of research for each of them.

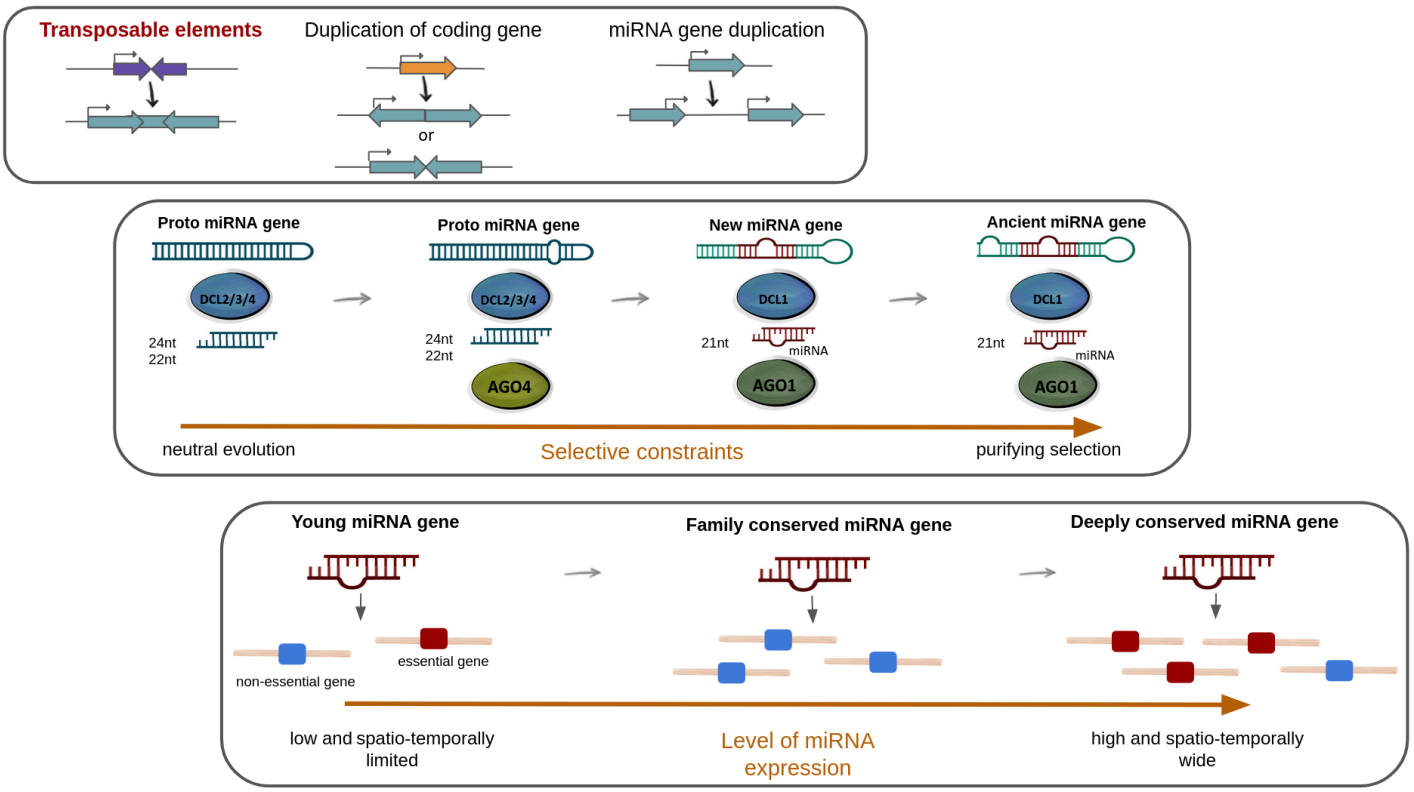


Figure 1: Schematic version of the main results of the thesis.

1. Conclusions

1.1 An “open” pangenome for miRNA genes

The genome of a single individual is generally not sufficient to account for the genetic variation across species. The recent advances in sequencing technologies now allow to sequence a large number of individual genomes at a reasonable cost and in a reasonable time, enabling the construction of pangenomes. Pangenomes represent the full repertoire of genes present in one species and can be divided into the core genome comprising the set of genes that are shared between all the individuals and into the accessory genome including the set of genes that are shared only by few or unique individuals. Pangenomes can be categorized into two groups, “open” pangenomes with a large accessory genome and a small core genome and “closed” pangenomes with a small accessory genome and a large core genome. Open pangenomes can be observed in species that occupy various environments and have large population sizes, such as bacterial species (Brockhurst et al., 2019). Park et al., (2019) compared more than 27,000 genomes belonging to seven prokaryotic species, and they evaluated the saturation curve of core and accessory genomes of the seven species. The core genomes exhibited no fluctuations with a number of genes ranging from about 1,000 to 4,000 genes according to the species, while the accessory genomes curve exhibited no saturation with a number of genes ranging from about 25,000 to 125,000 genes. Due to the high rate of horizontal gene transfer and the large amount of data available, the concept of pangenome in prokaryotic species is well established (Brockhurst et al., 2019). The closed pangenomes are characterized by small accessory genomes and large core genomes, usually in species that colonize relatively stable environments with smaller population sizes, as in plants and animals (Brockhurst et al., 2019). Sherman et al., (2019) compared 910 individual human genomes from African descent to the human reference genome. They revealed that the pangenome constructed contained about 10% more DNA sequence than the reference genome, and included up to 315 genes. Song et al., (2020) constructed the pangenome of eight *Brassica napus* accessions and showed that the pangenome tended to be saturated with six genomes. In addition, 42% of the genes were indispensable, present in all individuals, 56% were core genes present in at least seven individuals, and only 2% were specific to individual genomes. Contrary to protein-coding genes, our study of the repertoire of miRNA genes in *A. halleri* and *A. lyrata* showed that a large number of the species-specific miRNA genes are found in a few accessions only, indicating an open type of “panmiRNAome”. This raises the question of the factors responsible for the structure of this repertoire. A first adaptive

explanation could be that the large number of accessory miRNA genes results from a beneficial process of gains and losses, eventually conferring advantages under local environments. However, a second possibility linked to the “drift-barrier” hypothesis (Lynch, 2007) could be that selection fails to prevent the spread of neutral to mildly deleterious miRNA gene acquisitions due to their ease of emergence, particularly from transposons.

1.2 Evolution of the gene regulatory networks

A major source of phenotypic changes relies on changes of gene regulatory networks (GRN). GRNs are characterized by nodes (regulators and regulated genes) and edges (the regulatory interactions). Gene regulatory networks can evolve through the gain or loss of connections. Such changes can occur as a result of mutations appearing in the sequence of the binding site of the gene targeted, leading to the loss or gain of the binding site usually affecting a single target, while a mutation appearing in the sequence of the regulator can affect many target genes (Jones and Vandepoele, 2020). Wu et al., (2021) constructed the GRN related to salt stress response in *A. thaliana* and *Marchantia polymorpha* based on transcriptome analysis. The GRNs were hierarchical, dominated by transcription factors regulating a large number of genes (more than ten targets). In *A. thaliana,* transcription factors formed various small networks while in *M. polymorpha* they observed a single large network. However in both networks WRKY transcription factors were central nodes, highly connected to other transcription factors. Knockout mutants of these factors confirmed their central role causing the disruption of salt-response GRNs in *M. polymorpha* and *A. thaliana,* while other TFs in peripheral nodes were more divergent. In contrast, the cis-regulatory elements were more divergent suggesting that a mutation arising in the region had lower consequences on the GRN. Conserved miRNA genes have a large number of targets and the mature miRNA produced by these genes is highly constrained by natural selection (chapter I). In contrast, the miRNA binding site in mRNA targets seem to be less constrained (chapter I), suggesting that as transcription factors, a mutation arising in targets has less impact on GRN than a mutation appearing in mature miRNA sequence.

Modification in the regulatory network can also arise after the duplication of a regulator or a target gene leading to redundancy of the function, neofunctionalization, *i.e.* acquisition of a new function by the paralog, subfunctionalization, *i.e.* partition of the ancestral gene functions between the paralogs (Jones and Vandepoele, 2020). Vlad et al., (2014) investigated morphological differences between *A. thaliana*, which has simple leaves, and its relative *Cardamine hirsuta*, which has dissected leaves comprising distinct leaflets. The

REDUCED COMPLEXITY (RCO) homeodomain transcription factor evolved through gene duplication in Brassicaceae and was lost in *A. thaliana,* contributing to the simplification of the leaf structure in this species. In addition, the neo-functionalization of an enhancer element in RCO conferred a novel gene expression pattern in developing leaf leading to the difference pattern observed in relative species of the Brassicaceae family. Doroshkov et al., (2018) analyzed the GRN evolution of trichome formation using phylogenetic analysis in a wide range of plant species. They observed that the appearance of new functions in the GRN of trichome morphogenesis in *A. thaliana* was linked to duplication events in the different plant taxa studied. Transcription factors duplication followed by mutations in their DNA-binding sites is an important contributor to GRN divergence. This process allows the transcription factor to be promiscuous in the sense that each transcription factor recognizes the common motif shared by the duplicates, but also a new motif gained by mutations. This modularity allows the transcription factor to diversify while not affecting the binding core, and thus facilitate the overcoming of the negative effects of pleiotropy (Voordeckers et al., 2015). However, protein-DNA interactions are complex and rely on various factors. Even if it is not impossible, it is difficult to imagine how a unique mutation occurring in a transcription factor can lead to the gain of new binding activity. In contrast a duplication of a miRNA gene followed by mutations can easily slightly disturbed its secondary structure leading to a different cleavage pattern and the production of various isomirs, *i.e.* miRNA variants (chapter I; chapter II).

The evolution of GNR depends also on the level of connectivity of the node, with hotspots, *i.e.* highly connected factors, supposed to have a major impact on the network. Ichihashi et al., (2014) used comparative transcriptomic to construct the GRN involved in leaf development in tomato and two related species with different leaf morphologies. They showed that a variation in BLADE-ON-PETIOLE (BOP) transcription factor expression was responsible for the differences observed in development of the leaf in the pieces studied. This factor was part of the peripheral gene network and controlled the KNOTTED-like HOMEOBOX gene which was part of the core network. This highlights the importance of changes in peripheral gene networks in rewiring the interactions in the whole GRN and possibly contributing to morphological diversity. Plant miRNA genes seem to integrate progressively the regulatory network with a number of essential targets increasing in the course of evolution (chapter I), suggesting that young miRNA genes are part of the peripheral network while more ancient miRNA genes are part of the core network. However, this has not been assessed properly and a transcriptomic study analyzing GRNs including miRNA is lacking. This could help to better understand the impact of the arrival of a new

miRNA gene on the structure of GRNs and how they could be involved in morphological novelties.

1.3 Evolutionary significance of young miRNA genes

New protein-coding genes can arise either from ancestral genes by gene duplication, gene fusion/fission, horizontal gene transfer or retrotransposition, either *de novo* from non-genic regions that gain the ability to be transcribed, by exonization or by creation of new ORF in a different frame (Van Oss and Carvunis, 2019). Gene duplication was thought to be a major process giving rise to new protein-coding genes. However, recent research highlights the importance of *de novo* gene origination (Van Oss and Carvunis, 2019). Li et al. (2016) analyzed a large number of *A. thaliana* accessions using genome, transcriptome, epigenome and translatome data. They identified 782 potential *de novo* protein-coding genes and analyzed the evolutionary forces behind their maintenance in the genome. Most of them were methylated, suggesting a process of “neutralization” of these loci that may at first be deleterious but can be potentially beneficial in some conditions. Then demethylation of the locus can allow the recovery of its transcriptional activity. A large number of new miRNA genes are issued from transposable elements (chapter II). Like *de novo* protein-coding genes, a *de novo* miRNA emergence from transposable elements could maintain the new miRNA while avoiding large deleterious effects. Indeed, we can imagine that a miRNA gene emerging from protein-coding can directly target the gene it originates and have deleterious consequences on the individual leading to its rapid elimination by natural selection. In contrast, miRNA genes emerging from transposons do not target protein-coding genes at first. Thus a miRNA gene emerging from these elements is supposed to be mostly neutral and “hidden” from natural selection. In the course of evolution these miRNA genes can be retained or eliminated by natural selection if they acquire mutations leading to acquisition of a target in protein-coding genes, depending on their beneficial or deleterious effect.

The repertoires of miRNA genes are characterized by their high number of young miRNA genes but which seem to evolve under neutral constraint (Fahlgren et al., 2010; chapter I), raising the question of their evolutionary significance. A recent example suggests that some young miRNA genes can contribute to the evolution of phenotypic diversity. Bradley et al., (2017) analyzed two closely related populations of *Antirrhinum majus* species which showed differences in flower color pattern. They found that these differences were due to a young fold-back hairpin structure arising from the inverted duplication of the gene chalcone 4′-Oglucosyltransferase, which encodes an enzyme involved in the synthesis of yellow pigment. This young hairpin was able to produce small RNAs targeting the chalcone

4′-Oglucosyltransferase gene leading to the yellow pattern observed in flowers of some populations of *A. majus.* Another example are the small RNAs produced by the self-incompatibility locus (S locus) in *A. halleri*. Self-incompatibility forces species to outcross instead of self-fertilize. In Arabidopsis species, it is controlled by a key-lock system composed of two proteins encoded by the S locus, S-locus cysteine-rich (SCR) in pollen and stigma S-locus receptor kinase (SRK) in stigma. If the SRK protein recognizes SCR, the fertilization is abolished, thus the system is expected to show high diversity favoring the reproduction. The system is based on hierarchical dominance-recessivity relations between the alleles at the S-locus. Particularly, sRNAs are produced by the most dominant alleles leading to the transcriptional silencing of the SCR recessive alleles which allow to increase the number of partners (Durand et al., 2014). On the other hand, some young miRNA genes can have deleterious impacts. Berube et al., (2023) described a segregation distortion system, *i.e.* a distortion in normal segregation in favor of a selfish genetic element, involving RNA interference in maize. This system is composed of a toxin that kills pollen and antidots restoring pollen viability. Particularly, the pollen abortion is mediated by 22-nt small RNAs produced by a hairpin encoded by Teosinte Pollen Drive selfish genetic element that target essential genes in pollen grain. The way the pollen survives is mediated by a hypomorphic allele of DCL2, *i.e.* allele with a mutation which alters the gene product. These are punctual examples of possible roles of young miRNA genes, however the collective significance of the regulatory role of these genes has been poorly studied. Distinguishing among the large number of young miRNA genes those that do have functional relevance, either by being clear deleterious elements or by contributing to organismal fitness, from those that can be considered neutral will be an interesting challenge for further research.

2. Perspectives: Testing the regulatory potential of young miRNA genes in *A. halleri*

A main result of the previous chapters is that the *A. halleri* and *A. lyrata* genomes contain a large number of young miRNAs that are loaded in AGO proteins, but collectively seem to evolve under weaker selective constraints than those that are deeply conserved at the scale of Viridiplantae. This raises the question of their functional relevance. One perspective to my work is to determine whether these recently appeared miRNAs already have the capacity to reduce the transcript levels of their predicted target genes.

To address this question, a first possibility could have been to try and generate knock-in and knock-out genetically modified plants to test the impact of each individual young miRNA gene on the transcript level of their predicted targets. While this would allow for a direct test, this approach would represent a huge and actually overwhelming effort. Thus, I have set the stage for an alternative and experimentally tractable and powerful approach that can now be implemented. Briefly, the idea is to take advantage of a backcross population between *A. halleri* and *A. lyrata (A. halleri* x *A. lyrata* F1 plant backcrossed with the *A. lyrata* parent) to generate a series of individuals in which the *A. halleri*-specific miRNA genes segregate (Figure 2a). By generating RNA-seq data from a number of these individuals, we will be able to compare the transcript levels of the predicted target genes between the backcross individuals containing the *A. halleri*-specific miRNAs and those that do not contain them. This will provide a direct test of their regulatory potential (Figure 2b). Indeed, if the miRNAs are able to regulate the expression of their targets, we expect to observe a reduced expression of the targeted genes in the group of individuals containing the *A. halleri*-specific miRNAs compared to the group that lack the *A. halleri*-specific miRNAs. Furthermore, by using the SNPs specific to the *A. halleri* and the *A. lyrata* mRNA sequences, we will be in position to test whether the *A. halleri*-specific miRNAs specifically target the *A. halleri* transcripts, or indiscriminately target both *A. halleri* and *A. lyrata* transcripts (Figure 2b). An asset of the proposed approach is that with a single RNA-seq experiment we will have the potential to reveal at once the effect of all *A. halleri*-specific miRNAs, since a different set of them will segregate in the different backcross individuals.

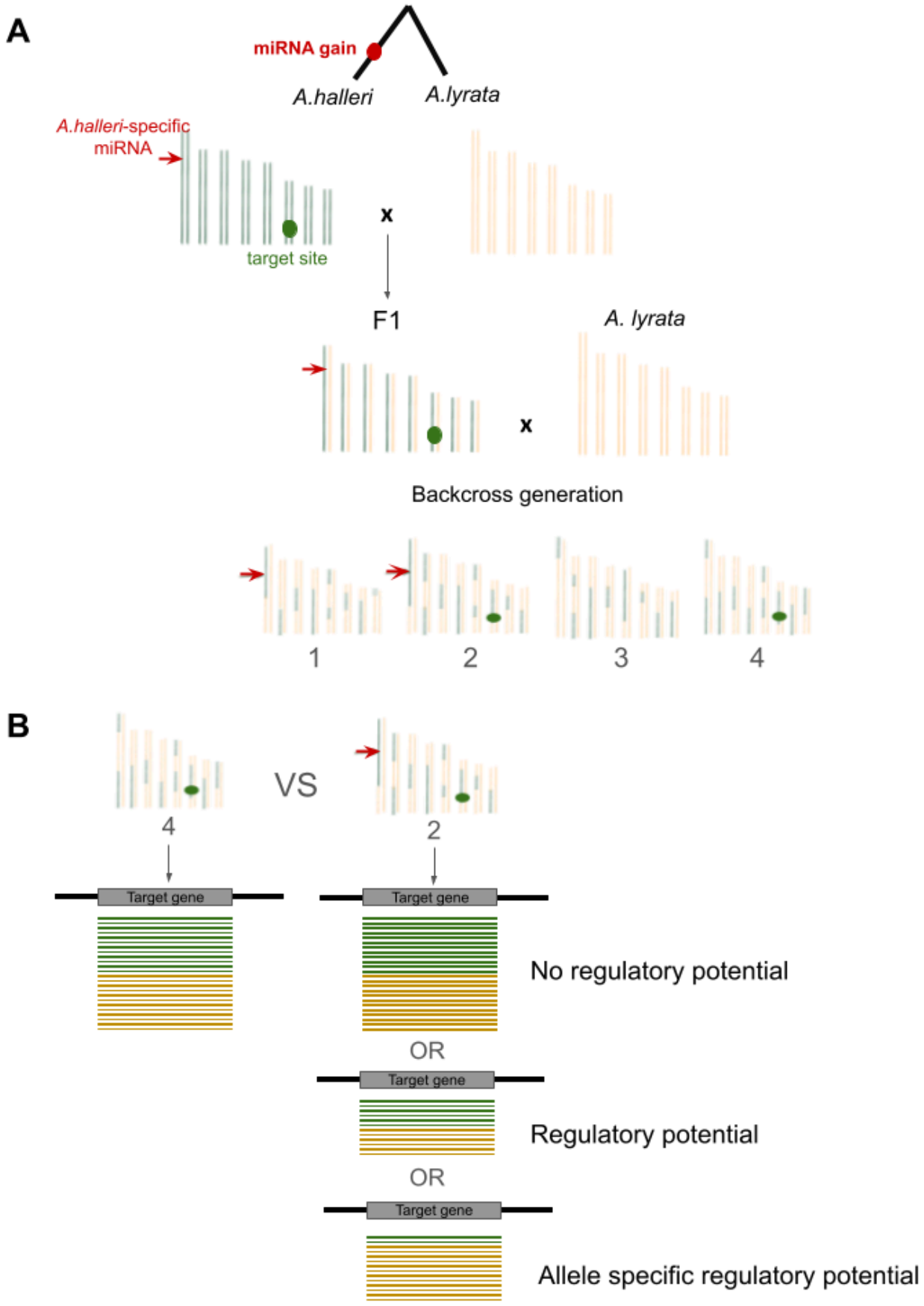

A
miRNA gain
A.halleri
A.lyrata
A.halleri-specific
miRNA
target site
x
F1
A. lyrata
x
Backcross generation
1
2
3
4
B
VS
4
2
Target gene
Target gene
No regulatory potential
OR
Target gene
Regulatory potential
OR
Target gene
Allele specific regulatory potential

Figure 2: An experimental approach to evaluate the regulatory potential of young miRNA genes. (a) *A. halleri* with specific miRNA genes and *A. lyrata* were crossed to obtain a F1 hybrid, which was backcrossed with the *A. lyrata* parent. This allows the segregation of the *A. halleri*-specific miRNA genes in the backcross population, which can be divided in four groups: 1) individuals with the *A. halleri*-specific miRNAs but without the *A. halleri* targets, 2) individuals with the *A. halleri*-specific miRNAs but with the *A. halleri* targets, 3) individuals without the *A. halleri*-specific miRNAs and without the *A. halleri* targets, 4) individuals without the *A. halleri*-specific miRNAs but with the *A. halleri* targets. (b) Comparison of the levels of transcripts of the genes targeted by the *A. halleri*-specific miRNAs of group 2 and 4 of the backcross populations will allow us to assess the regulatory potential of the *A. halleri*-specific miRNAs. The comparison of the level of transcripts specifics to *A. halleri* or *A. lyrata* allele (using specific markers as SNP) will allow us to determine whether the *A. halleri*-specific miRNAs target preferentially the *A. halleri* allele or not. Equal level of transcripts between groups 2 and 4 would mean that the *A. halleri*-specific miRNAs are not able to regulate the expression of their targets. Reduced levels of *A. halleri* or *A. lyrata* allele transcripts would mean that the *A. halleri*-specific miRNAs are able to regulate the expression of both alleles of their targets, while a reduction of the transcripts from *A. halleri* allele only would mean that regulatory effect of *A. halleri*-specific miRNAs is allele specific in favor of *A. halleri* alleles.

The first step will be to identify the repertoires of miRNAs and their target genes in the two parents, the F1 individual and in individuals of the backcross. To do this, a *A. halleri* x *A. lyrata* F1 plant was backcrossed with the *A. lyrata* parent, and a large number of backcross individuals was obtained (by our collaborators Lea Hoerdemann and Juliette DeMeaux, University of Cologne). The resulting backcross population was initially constructed to map QTLs using a high-density genetic map based on GBS markers. A total of 3,451 molecular markers were obtained. This map will allow us to identify the parental origin of each chromosomal segment, and determine whether a given *A. halleri*-specific miRNA gene was transmitted to any particular backcross individual. As we showed in chapter I, a large number of species-specific miRNA genes are actually accession-specific, so for this particularly detailed analysis it was important to have a direct repertoire of miRNA genes to follow them in the parental genomes. The two parental genomes were sequenced and *de-novo* assembled, as well as that of the F1 of the backcross using long reads (PACBIO HiFi), and we used the trio binning approach to obtain high-quality phased chromosome-level assemblies (collaboration with Wiliam Marande, CNRGV Toulouse). The resulting *A. halleri*, *A. lyrata* and F1 assemblies were composed between 15 and 332 contigs and had a cumulative size from 192 to 232 Mbp with an N50 between 12.7 to 22.2 Mbp (Table 1). The completeness of the genomes were assessed using BUSCO and 98.7 to 99.3% complete universal single-copy orthologs, 0.1 to 0.2% fragmented universal single-copy orthologs and 0.6 to 1.1% missing universal single-copy orthologs were found from the Brassicales dataset odb10 (Table 1). The contigs of the *A. halleri* and *A. lyrata*

parent genome assemblies will be scaffolded with RagTag (Alonge et al., 2022) using the *A. halleri* Auby-1 reference genome assembly (chapter I) and the *A. lyrata* MN47 genome assembly (Kolesnikova et al., 2023). We have started to annotate the genes in the three genome assemblies, using Liftoff (Shumate and Salzberg, 2021) to transfer the annotations from the *A. halleri* Auby-1 (chapter I) and *A. lyrata* MN47 (Kolesnikova et al., 2023) references.

For this experiment we have chosen to focus on a single tissue, leaves, and I have flash-frozen leaf material and extracted total RNA from all three parents as well as 100 backcross individuals. At this stage I have produced sRNA sequencing libraries from the three parents, as well as from a subset of ten backcross individuals in order to verify that we are indeed able to track the presence of the *A. halleri*-specific miRNAs from the chromosomal fragments inferred from the GBS markers alone. We obtained a median of 17.9 million sequencing reads per sample (ranging from 1.2 to 45.8 millions).

The second step will be to evaluate the gene expression levels in the different samples to analyze the differential gene expression of the genes targeted by the *A. halleri*-specific miRNAs. We produced classical RNA sequencing data from the three parents (to quantify the baseline gene expression levels) as well as from the same subset of ten backcross individuals. We will now be ready to produce RNA-seq libraries from all 100 individuals of the backcross populations to quantify normalized transcript levels among them.

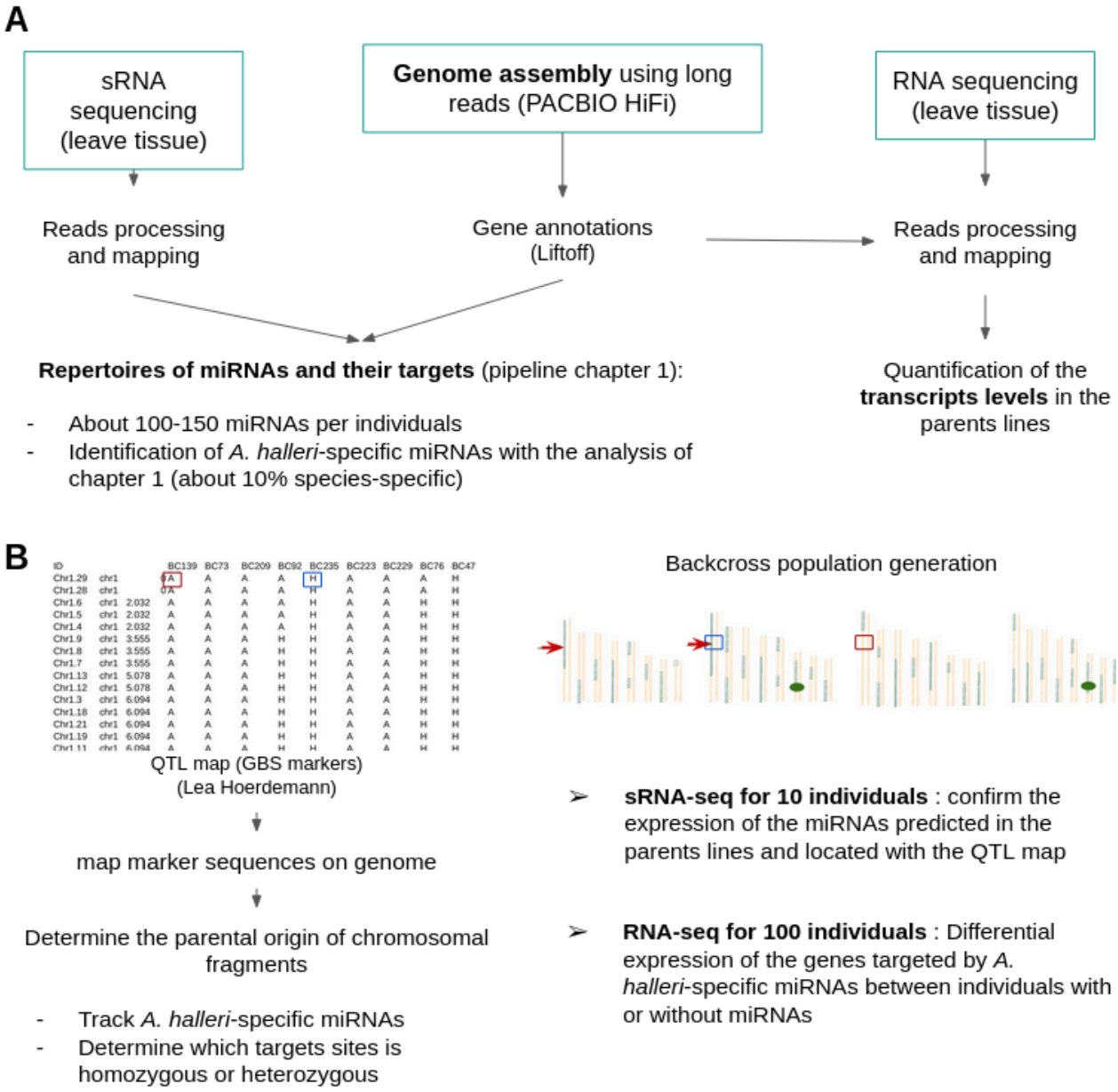


Figure 3 :Pipeline for miRNA genes annotations and the evaluation of transcript levels in parents and backcross population. (a) Analysis of the three parent lines. The production of three genomes assemblies and small RNA sequencing data allow to annotate the miRNA genes and their targets in the three individuals, while the RNA-seq data allow to quantify the baseline levels of gene expression. (b) Analysis of the backcross population. The QTL map with the individuals of the backcross population in rows and the molecular markers in line. The red rectangles represent homozygous sites while blue rectangles represent heterozygous sites.

	A. halleri	*A. lyrata*	F1, hap1	F2, hap2
Contig number	30	15	83	332
Total length (Mbp)	232	192	225	206
N50 (Mbp)	22.2	23.4	12.7	18.3
L50	5	4	6	5
Complete BUSCOs	98.8%	99.3%	98.7%	99.2%
Fragmented BUSCOs	0.2%	0.1%	0.2%	0.1%
Missing BUSCOs	1.0%	0.6%	1.1%	0.7%

Table 1: Resume statistics of the *A. halleri*, *A. lyrata* and F1 genome assemblies.

3. References

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z.B., Schatz, M.C., and Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol **23**: 258.

Berube B, Ernst E, Cahn J, Roche B, de Santis Alves C, Lynn J, Scheben A, Siepel A, Ross-Ibarra J, Kermicle J, Martienssen R. (2023). *Teosinte Pollen Drive* guides maize diversification and domestication by RNAi. bioRxiv [Preprint]. 2023

Bradley, D. et al. (2017). Evolution of flower color pattern through selection on regulatory small RNAs. Science **358**: 925–928.

Brockhurst, M.A., Harrison, E., Hall, J.P.J., Richards, T., McNally, A., and MacLean, C. (2019). The Ecology and Evolution of Pangenomes. Current Biology **29**: R1094–R1103.

Durand, E. et al. (2014). Dominance hierarchy arising from the evolution of a complex small RNA regulatory network. Science **346**: 1200–1205.

Doroshkov, A.V., Konstantinov, D.K., Afonnikov, D.A., and Gunbin, K.V. (2019). The evolution of gene regulatory networks controlling Arabidopsis thaliana L. trichome development. BMC Plant Biol **19**: 53.

Fahlgren, N., Jogdeo, S., Kasschau, K.D., Sullivan, C.M., Chapman, E.J., Laubinger, S., Smith, L.M., Dasenko, M., Givan, S.A., Weigel, D., and Carrington, J.C. (2010). MicroRNA Gene Evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. The Plant Cell **22**: 1074–1089.

Ichihashi, Y., Aguilar-Martínez, J.A., Farhi, M., Chitwood, D.H., Kumar, R., Millon, L.V., Peng, J., Maloof, J.N., and Sinha, N.R. (2014). Evolutionary developmental transcriptomics reveals a gene network module regulating interspecific diversity in plant leaf shape. Proc. Natl. Acad. Sci. U.S.A. **111**.

Jones, D.M. and Vandepoele, K. (2020). Identification and evolution of gene regulatory networks: insights from comparative studies in plants. Current Opinion in Plant Biology **54**: 42–48.

Kolesnikova, U.K., Scott, A.D., Van De Velde, J.D., Burns, R., Tikhomirov, N.P., Pfordt, U., Clarke, A.C., Yant, L., Seregin, A.P., Vekemans, X., Laurent, S., and Novikova, P.Y. (2023). Transition to Self-compatibility Associated With Dominant *S* -allele in a Diploid Siberian Progenitor of Allotetraploid *Arabidopsis kamchatica* Revealed by *Arabidopsis lyrata* Genomes. Molecular Biology and Evolution **40**: msad122.

Li, Z.-W., Chen, X., Wu, Q., Hagmann, J., Han, T.-S., Zou, Y.-P., Ge, S., and Guo, Y.-L. (2016). On the Origin of De Novo Genes in *Arabidopsis thaliana* Populations. Genome Biol Evol **8**: 2190–2202.

Lynch, M. (2007). The evolution of genetic networks by non-adaptive processes. Nat Rev Genet **8**: 803–813.

Park, S.-C., Lee, K., Kim, Y.O., Won, S., and Chun, J. (2019). Large-Scale Genomics Reveals the Genetic Characteristics of Seven Species and Importance of Phylogenetic Distance for Estimating Pan-Genome Size. Front. Microbiol. **10**: 834.

Sherman, R.M. et al. (2019). Assembly of a pan-genome from deep sequencing of 910 humans of African descent. Nat Genet **51**: 30–35.

Shumate, A. and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations. Bioinformatics **37**: 1639–1643.

Song, J.-M. et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. Nat. Plants **6**: 34–45.

Van Oss, S.B. and Carvunis, A.-R. (2019). De novo gene birth. PLoS Genet **15**: e1008160.

Vlad, D., Kierzkowski, D., Rast, M.I., Vuolo, F., Dello Ioio, R., Galinha, C., Gan, X., Hajheidari, M., Hay, A., Smith, R.S., Huijser, P., Bailey, C.D., Tsiantis, M. (2014). Leaf Shape Evolution Through Duplication, Regulatory Diversification, and Loss of a Homeobox Gene. Science **343**: 780–783.

Voordeckers, K., Pougach, K., and Verstrepen, K.J. (2015). How do regulatory networks evolve and expand throughout evolution? Current Opinion in Biotechnology **34**: 180–188.

Wu, T.-Y., Goh, H., Azodi, C.B., Krishnamoorthi, S., Liu, M.-J., and Urano, D. (2021). Evolutionarily conserved hierarchical gene regulatory networks for plant salt stress response. Nat. Plants **7**: 787–799.

Acknowledgements

Je tiens à exprimer ma profonde gratitude à tous ceux qui m'ont accompagné de près ou de loin tout au long de cette aventure doctorale. Votre soutien, vos conseils et vos encouragements ont été essentiels à l'aboutissement de ce travail.

Mon jury de thèse

Je remercie tout d'abord les membres du jury pour avoir accepté de participer à ma soutenance de thèse. En particulier, je suis reconnaissante à Laurent Duret et Karine Alix d'avoir accepté d'être rapporteur/rice de mon manuscrit de thèse et d'avoir évalué mon travail. Je remercie également Clémentine Vitte et Hélène Touzet pour leur participation en tant que membres du jury.

Ceux qui m'ont financé

Ce travail a été rendu possible grâce au soutien financier de la Région Nord Pas de Calais (projet MICRO2), le Conseil Européen de la Recherche (projet NOVEL, subvention #648321) et l'Agence Nationale de la Recherche (projet TE-MoMa, subvention ANR-18-CE02-0020-01). Je remercie particulièrement l'ERC et l'Université de Lille pour le financement de mon projet de thèse et de mes missions. Je remercie également l'école doctorale SMRE pour leur soutien.

Ceux qui m'ont encadré

Je remercie chaleureusement mes encadrants Vincent Castric et Sylvain Legrand pour leur accompagnement durant ces trois ans et demi. Je n'aurais pas pu rêver d'un meilleur encadrement. Votre soutien et vos conseils avisés ont été d'une aide précieuse. J'ai beaucoup appris, tant sur le plan intellectuel qu'humain. Merci également de m'avoir aidé à avoir confiance en mon travail. Vincent merci d'avoir pris du temps tard le soir pour corriger mes travaux. Merci de m'avoir encouragé à passer le permis et m'avoir fait réviser le code sur les routes en direction de Cologne. Sylvain merci d'avoir pris du temps pour m'apprendre les béabas de python et m'y avoir donné goût. Merci de m'avoir accompagné dans mon parcours d'enseignante et merci pour tous les conseils. Merci également d'avoir eu la patience de me laisser blablater chaque fois où je venais te déranger dans ton bureau juste pour '2min'.

Merci à vous deux pour votre soutien et bienveillance !

Ceux qui ont collaboré

Merci à l'équipe de bioinformatique, Sophie Gallina, Clément Mazoyer et Mathieu Genete qui m'ont accompagné dans mon apprentissage de la bioinformatique. En particulier Mathieu qui a eu la patience de m'aider à résoudre les bugs de mes scripts (même si parfois c'était juste moi le bug..) et avec qui j'ai passé de très bon moments (cf Roscoff !).
Merci à l'équipe de biologie moléculaire Christelle Lepers-Blassiau, Laurence Debacker, Cécile Godé et Anne-Catherine Holl pour les bons moments. En particulier merci à Christelle pour m'avoir accompagné dans les expériences de biologie moléculaire. Merci pour ton soutien émotionnel et pour m'avoir appris à respecter mes limites. Tu resteras pour moi ma maman du labo. Mais aussi, merci à Laurence pour l'aide en biologie moléculaire, merci pour tes blagues (même si certaines m'ont plutôt fait peur) et les rires.
Merci à l'équipe des serres, Chloé Ponitzki, Esther Houzé, Eric Schmitt et Nathalie Faure pour votre aide en serre. En particulier, merci à Chloé et Esther d'avoir pris soin de mes plantes.

Merci à Jacinthe Azevedo-Favory et Thierry Lagrange pour m'avoir accueilli au LGDP de Perpignan. En particulier à Jacinthe pour m'avoir guidé dans l'aventure des immunoprécipitations.

Thanks to Juliette De Meaux and Léa Hoerdemann for their warm welcome at the Institute for Plant Sciences in Cologne and for allowing me to work on the backcross population.

Thanks to Blake Meyers for welcoming me into his laboratory and giving me moments of his precious time to comment on the first chapter. Thanks also to Patricia Baldrich, Noah Fahlgren, Kerrigan Gilbert and Keith Slotkin for their comments. Patricia in particular for accompanying me. And thanks to the other people in Meyers' lab, especially Lily and David, for their hospitality.

Merci aux membres de mes CSI, Eléonore Durand, Jacinthe, Anamaria Necsulea et Filipe Borges pour les nouveaux points de vue et les commentaires sur le projet.

A ceux qui s'occupent de l'administration

Merci à Sandrine Belingheri pour m'avoir accompagné dans les méandres de l'administration pour les missions. Merci à Christophe Van Brussel pour celles de l'école doctorale et des inscriptions à chaque année de thèse.

Ceux qui m'ont accompagné au jour le jour au laboratoire

Je remercie aussi les collègues de l'équipe Ecologie Evolution, pour m'avoir fourni un environnement de travail chaleureux. Merci Céline P., Pierre S.L., Xavier V., les réunions de groupe et vos précieux conseils. En particulier, merci Pierre pour être le Macgyver de l'équipe. Merci à Jean-François A., Anne D. et Isabelle D.C. pour les conservations lors des pauses/pots de soutenance. Merci Christelle F. pour tes conseils de détente en fin de thèse. Merci Camille R. pour ton humour et le partage de ton expérience (qui reste peut-être à discuter) en mixologie. Merci Pascal T. d'avoir pris le temps de discuter avec moi lors d'une période difficile. Merci Sylvain B. pour les discussions sur la philosophie et l'art et tes remarques pertinentes sur le projet (même si sur le moment c'était un peu brusque à vrai dire). Merci Véronique D.L. pour nous permettre de travailler dans un environnement sain, mais surtout merci pour les discussions.

Aux thésards et postdocs

Merci aux thésards déjà présents lorsque j'ai commencé ma thèse, Thomas, Estelle, Audrey, Zoé, Chloé et Rita pour votre accueil chaleureux et les beaux moments passés autour d'une bière. En particulier merci Audrey pour toutes ces discussions en pause. Merci d'avoir organisé des petits mystères au sein du labo qui ont réveillé nos instincts de Sherlock Holmes. Merci Zoé pour les discussions sur les droits sociaux et tabous. Thank you Rita for finding the right words to comfort us in times of doubt. Thanks for all the comments on my first chapter.

Merci aux thésards/étudiants de ma promo et des suivantes, Emilie, Agathe, François, Claire, Guillaume, Justine, Camille, Alix, Muskaan, Achille, Matteo, Fabien, Timothee, Arthur et Clément d'avoir fait de ces années merveilleuses. Merci pour tous les bons moments lors des sorties, des bières, des restos et des blagues qui n'en finissent pas. Merci à la team d'origine du bureau 206, alias le côté obscur de la force, Emilie (notre regrettée ex 206 passée du côté clair de la force), François, Claire et Guillaume pour tous les rires et votre bienveillance. Merci à Claire pour les discussions sur les miRNAs, mais surtout pour ton ingéniosité dans la création les blagues et pour avoir été ma partenaire dans l'établissement d'un nouvel Ordre. Merci à Guillaume pour ta gentillesse et pour avoir été notre panda du bureau. Merci à la team du 222, alias le côté clair de la force, Agathe, Justine puis Emilie les ayant rejoint, pour avoir eu la patience de supporter le bureau 206, merci pour votre bonne humeur et votre soutien. Merci Justine pour m'avoir fait découvrir que le bourdon n'est pas le mâle de l'abeille. Merci à Agathe pour toutes tes histoires et pour les devinettes carambar.

Merci surtout pour la découverte de la pêche au chinchilla qui m’a fait pleurer de rire. Merci à Achille, Fabien et Tim, la nouvelle team bloc qui va assurer une relève haut la main.

En particulier énorme merci à François (mon acolyte du 206, acolyte de grimpe et acolyte de rire) et à Emilie (la prankeuse la ‘plus dangereuse’ du bureau : gare à vos crayons !) pour tous les bons moments passés à discuter autour d’un bon repas. Merci pour vos conseils de tempérance (en particulier Emilie de qui je ne me lasserai du ‘non’). Mais également, merci d’avoir été là pour moi dans les moments difficiles. Vous êtes tous deux gravés à jamais en moi !

A ma famille

Je remercie, bien entendu, ma famille et mes amis pour leur soutien et pour s'être intéressé ou tout du moins avoir fait mine de s'intéresser à mes travaux. Merci à mon frère Donovan pour m’avoir aidé à m’améliorer en programmation. Merci à ma sœur Ophélie de m'avoir écouté me plaindre de mon stress. Gracias Emilio por dejar tu país para seguirme en esta maravillosa aventura y por estar a mi lado en los momentos de tensión. Et surtout, merci à mes parents pour leur éducation, pour les sacrifices faits pour leurs enfants et pour leur soutien. En particulier, c’est grâce à vous que j’ai pu en arriver jusqu’ici.

Merci à vous tous pour ces merveilleuses années et je finirai par dire…
Qu’est ce qu’on a bien rigolé !

Abstract (English)

Understanding the origins of genomic novelties is a central question in evolutionary biology. Differences in the regulation of gene expression are an important cause of phenotypic variability, and microRNAs (miRNAs) have emerged as pivotal regulators of gene expression in plant and animal genomes. miRNAs negatively regulate gene expression at the post-transcriptional level by interacting with messenger RNA targets. While some miRNAs are deeply conserved, many appear to be species-specific, raising the question of how they emerge and integrate into cellular regulatory networks. However, we still lack a proper understanding of the evolutionary origins of new miRNA genes and of the processes by which they progressively become functionally specialized. In my PhD project, I focused on two closely related species of the plant genus Arabidopsis, *A. halleri* and *A. lyrata* that diverged about one million years ago. In the first chapter, I used a large set of small RNA sequencing data, to perform a detailed annotation of miRNA genes in the *A. halleri* and *A. lyrata* genomes. I investigated the conservation status of these miRNA genes among eighty five plant species to characterize the process by which newly emerged miRNA genes progressively acquire the properties of “canonical” miRNA genes over the course of evolution (in terms of features of the hairpin precursor, level of polymorphism in natural populations, loading into Argonaute proteins and number of target genes). Overall, my results suggest a rapid birth-and-death process of the miRNA repertoire, whereby “proto” miRNA genes appear steadily with little to no functional constraint, only a small number of which will be maintained over time and eventually integrated into "core" biological processes. In the second chapter, I reasoned that since species-specific miRNAs have emerged recently, they may have retained a record of their mutational origin. To test this idea, I evaluated the relative contribution of several proposed sources of miRNA genes (other miRNA genes, protein-coding genes, transposable elements, non-coding intergenic DNA) by comparing their genomic sequences to databases of these putative evolutionary progenitors. Overall, this thesis provides a detailed picture of the micro- and macro-evolution of miRNA genes in the Arabidopsis genus. The results it contains show that the regulatory network constituted by miRNA genes and their target genes can be either rapidly rewired or remain stable over extended evolutionary times. They provide insight into the evolutionary significance of the fluidity of the repertoire of miRNA genes in plant genomes.

Résumé (Français)
Comprendre les origines des nouveautés génomiques est une question centrale en biologie évolutive. Les différences dans la régulation de l'expression des gènes sont une cause importante de la variabilité phénotypique, et les microARNs (miARNs) sont apparus comme des régulateurs essentiels de l'expression des gènes dans les génomes végétaux et animaux. Les miARNs régulent négativement l'expression des gènes au niveau post-transcriptionnel en interagissant avec les cibles ARN messager. Si certains miARNs sont profondément conservés, beaucoup semblent être spécifiques à une espèce, ce qui soulève la question de savoir comment ils émergent et s'intègrent dans les réseaux de régulation cellulaire. Cependant, nous ne comprenons toujours pas bien les origines évolutives des nouveaux gènes de miARN et les processus par lesquels ils se spécialisent progressivement sur le plan fonctionnel. Dans le cadre de mon projet de doctorat, je me suis concentrée sur deux espèces étroitement apparentées du genre Arabidopsis, A. halleri et A. lyrata, qui ont divergé il y a environ un million d'années. Dans le premier chapitre, j'ai utilisé un grand nombre de données de séquençage de petits ARNs pour réaliser une annotation détaillée des gènes de miARN dans les génomes d'A. halleri et d'A. lyrata. J'ai étudié l'état de conservation de ces gènes de miARN parmi quatre-vingt-cinq espèces de plantes afin de caractériser le processus par lequel les gènes de miARN nouvellement apparus acquièrent progressivement les propriétés des gènes de miARN "canoniques" au cours de l'évolution (en termes de caractéristiques du précurseur en épingle à cheveux, de niveau de polymorphisme dans les populations naturelles, de chargement dans les protéines Argonaute et de nombre de gènes cibles). Dans l'ensemble, mes résultats suggèrent un processus rapide de naissance et de mort du répertoire des miARNs, par lequel des "proto" gènes de miARN apparaissent régulièrement avec peu ou pas de contraintes fonctionnelles, et dont seul un petit nombre sera maintenu au fil du temps et finalement intégré dans des processus biologiques "centraux". Dans le deuxième chapitre, j'ai émis l'hypothèse que les miARNs spécifiques à une espèce étant apparus récemment, ils pourraient avoir conservé une trace de leur origine mutationnelle. Pour tester cette idée, j'ai évalué la contribution relative de plusieurs sources proposées de gènes de miARN (autres gènes de miARN, gènes codant pour des protéines, éléments transposables, ADN intergénique non-codant) en comparant leurs séquences génomiques aux bases de données de ces progéniteurs évolutifs supposés. Dans l'ensemble, cette thèse fournit une image détaillée de la micro- et de la macro-évolution des gènes de miARN dans le genre Arabidopsis. Les résultats qu'elle contient montrent que le réseau de régulation constitué par les gènes de miARN et leurs gènes cibles peut être rapidement remanié ou rester stable sur de longues périodes d'évolution. Ils donnent un aperçu de la signification évolutive de la fluidité du répertoire des gènes de miARN dans les génomes végétaux.