DNA-dependent RNA polymerases are essential enzymes that read the information that is stored in DNA and convert it into RNA. Unlike bacteria and archaea, the genomes of which encode only one multisubunit RNA polymerase, eukaryotes use three essential RNA polymerases to decode their nuclear genomes. RNA polymerase I (Pol I) is the most specialized of the three, transcribing hundreds of copies of genes that are essentially identical and whose primary transcripts are processed into the 18S, 5.8S and 28S (25S in yeast) catalytic RNAs of ribosomes. Pol II is the most versatile of the RNA polymerases, transcribing thousands of protein-coding genes, long non-coding RNAs and small RNAs that modify or process messenger RNAs or ribosomal RNAs. Pol III is also versatile, as it transcribes several classes of smaller RNAs (typically under 500 nucleotides), such as 5S ribosomal RNA, transfer RNAs, small regulatory RNAs and short interspersed nuclear elements (SINEs).

The specialized functions of Pol I, Pol II and Pol III must have been established in the last common ancestor of the Eukarya, given their ubiquity and conservation. How they arose and diversified from a single archaean progenitor enzyme is an intriguing — but intractable — question owing to a lack of alternative evolutionary outcomes, such as groups of eukaryotes that are missing one or more of the polymerases. Against this intellectual backdrop, considerable excitement was generated by the first hints, a decade ago, of the existence of novel plant-specific RNA polymerases, which have since been named Pol IV and Pol V.

Discovery of Pol IV and Pol V

Given the expectation that there would be three, and only three, nuclear multisubunit RNA polymerases in plants, as in other eukaryotes, the Arabidopsis thaliana genome sequence contained a surprise. During the annotation process, C. S. P. noted that the A. thaliana genome possessed sequences for the largest and second-largest subunits of two potential RNA polymerases that were distinct from those of Pol I, Pol II and Pol III. As these two subunits form the active site of RNA polymerases, this observation suggested the existence of atypical RNA polymerases. Additional phylogenetic analyses confirmed that the atypical RNA polymerase subunit genes were present only in plants, suggesting that the atypical RNA polymerases might carry out plant-specific functions.

Recent articles have discussed the evolution of Pol IV and Pol V subunits and the divergence of their catalytic centres compared with those of other RNA polymerases. In this Review, we focus on the biological roles of Pol IV and Pol V in gene silencing, their differences in subunit composition compared with Pol II and the questions that remain concerning the templates and transcripts of Pol IV and Pol V.

Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing

Jeremy R. Haag and Craig S. Pikaard

Abstract | In all eukaryotes, nuclear DNA-dependent RNA polymerases I, II and III synthesize the myriad RNAs that are essential for life. Remarkably, plants have evolved two additional multisubunit RNA polymerases, RNA polymerases IV and V, which orchestrate non-coding RNA-mediated gene silencing processes affecting development, transposon taming, antiviral defence and allelic crosstalk. Biochemical details concerning the templates and products of RNA polymerases IV and V are lacking. However, their subunit compositions reveal that they evolved as specialized forms of RNA polymerase II, which provides the unique opportunity to study the functional diversification of a eukaryotic RNA polymerase family.
Pol IV and Pol V\textsuperscript{14-16} to reflect their distinct functions and subunit compositions (BOX 1).

**RNA-directed DNA methylation**

Pol IV and Pol V are best understood with regard to their roles in RNA-directed DNA methylation (RdDM), a process by which 24-nucleotide small-interfering RNAs (siRNAs) direct the de novo cytosine methylation of complementary DNA sequences (FIG. 2). Thousands of retrotransposons and endogenous repeats are silenced and controlled by RdDM, as are invading DNA viruses, transgenes and some protein-coding genes\textsuperscript{17-20}. Pol IV and Pol V have distinct roles at the beginning and ends of the RdDM pathway, respectively, with Pol IV being required for siRNA biogenesis and Pol V transcripts being required for siRNA targeting of RdDM-affected loci.

**Initiating RdDM: Pol IV, RDR2 and siRNA biogenesis.**

Although Pol IV primary transcripts have proven elusive and remain undefined, Pol IV is thought to initiate the RdDM pathway (FIG. 2, step 1) because it physically colocalizes with loci that undergo RdDM\textsuperscript{21} and it is essential for 24-nucleotide siRNA biogenesis from the vast majority of loci\textsuperscript{18}. Moreover, in Pol IV-null mutants, other proteins of the RdDM pathway are mislocalized, suggesting that they all act downstream of Pol IV\textsuperscript{21}. This observation suggests that proteins of the RdDM pathway require the expression of Pol IV, or the products of Pol IV transcription, for stability.

Genetic evidence suggests that Pol IV transcripts are copied into double-stranded RNAs (dsRNAs) by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), one of six *A. thaliana* RNA-dependent RNA polymerases\textsuperscript{22,23} (FIG. 2, step 2). Pol IV and RDR2 interact in vivo, as shown by mass spectrometry and co-immunoprecipitation analyses (J.R.H., T. Ream and C.S.P., unpublished observation). A putative chromatin remodeler, CLASSY1 (CLSY1; also known as CHR38), which is a SWI2/SNF2-like helicase domain protein, assists in Pol IV and RDR2-dependent siRNA biogenesis\textsuperscript{24}. The nuclear localization pattern of CLSY1 resembles that of RDR2, and in *clsy1* mutants RDR2 is substantially mislocalized. Pol IV localization is also perturbed slightly in *clsy1* mutants, suggesting that CLSY1 acts at the interface between Pol IV and RDR2 (FIG. 2, step 2), perhaps facilitating transcription or the transfer of RNA from Pol IV to RDR2 (REF 24).
Box 1 | RNA polymerase subunit nomenclature

RNA polymerase IV (Pol IV) and Pol V subunit nomenclature stems from naming systems that are used in other eukaryotes, such as *Saccharomyces cerevisiae*, in which Pol I, Pol II, and Pol III are designated RPA, RPB and RPC, respectively. In plants, the prefix N (for nuclear) is added to circumvent nomenclature conflicts with other genes, thus Pol I, Pol II, Pol III, Pol IV and Pol V subunits are denoted by the prefixes NRP, NRPB, NRPC, NRPD and NRPE, respectively. In the nomenclature system we recently proposed, which is based on the complete subunit compositions of Pol II, Pol IV and Pol V and unpublished subunit compositions we have determined for Arabidopsis thaliana Pol I and Pol III (T. Ream and C.S.P., unpublished observations), plant RNA polymerase subunits are numbered based on their similarity to the corresponding subunits of Pol II in yeast. Thus, NRPD1 and NRPE1 are the largest subunits of Pol IV and Pol V, respectively, and NRPE5 is the Pol V-specific homologue of the fifth-largest subunit in yeast (the yeast subunits decrease in size with an increase in number). In cases where there is more than one variant of a particular subunit gene, these variants are designated with small letters; for instance, the two genes encoding potential NRPB9 subunits are designated NRPB9a and NRPB9b.

In the next step of the RdDM pathway (Fig. 2, step 3), dsRNAs are diced into 24-nucleotide duplexes by DICER-LIKE 3 (DCL3), one of four DICER endonucleases in *A. thaliana*. In *dcl3* mutants, some dicing of precursors still occurs, by virtue of DCL2 and DCL4 (REFS 22, 26), but the 21- and 22-nucleotide siRNAs that they produce are relatively ineffective for RdDM.

Following dicing, the RNA methylase HUA-ENHANCER 1 (HEN1) adds a methyl group to the 2′ hydroxyl groups of siRNAs at their 3′ ends, which helps to stabilize these small RNAs (Fig. 2, step 4). Consequently, siRNA levels and RdDM are reduced in *hen1* mutants. Following 3′ end methylation, one strand of the 24-nucleotide siRNA duplex is loaded into ARGONAUTE 4 (AGO4), or its close relatives AGO6 and AGO9, forming an RNA-induced silencing complex (RISC) (Fig. 2, step 5). The strand with the more weakly hydrogen bonded 5′ end is preferentially selected as the siRNA. The size of the diced product, and the nucleotide at its 5′ end, dictate its loading into compatible AGO proteins, depending on their availability in a given cell type or time in development.

Collectively, Pol IV, RDR2, DCL3 and HEN1 are crucial for siRNA biogenesis and stability, arming AGO4 with the small RNA guidance system that is needed to seek out homologous target sites in the genome. In the absence of compatible siRNAs AGO4 is unstable; thus, in Pol IV (*nrpd1*), *rdr2* or *dcl3* mutants, AGO4 levels are substantially reduced. Whether AGO4 is actively targeted for degradation in the absence of an associated siRNA is unknown.

Generating the target for RdDM: the role of Pol V

Unlike Pol IV, Pol V is not required for 24-nucleotide siRNA biogenesis at most loci; however, it is critical for RdDM. So, what does Pol V do? A breakthrough in our understanding came from the identification of Pol V-dependent non-coding RNAs produced at multiple loci that are subject to RdDM (Fig. 2, step 6). These Pol V-dependent RNAs span the promoter regions of silenced loci, are dependent on the Pol V active site and can be chemically crosslinked to Pol V, indicating that they are direct Pol V transcripts. Importantly, Pol V transcripts are generated even if 24-nucleotide siRNA biogenesis is abolished, indicating that Pol IV and Pol V carry out their functions independently, in parallel pathways that converge in order for RdDM to occur.

Several proteins that were identified in genetic screens facilitate Pol V transcription (Fig. 2, step 6): DRD1 is a SWI2/SNF2-related putative chromatin remodeler; DMS3 shares similarity with the hinge domains (dimerization regions) of structural maintenance of chromosomes (SMC) proteins, such as cohesins and condensins; and DMD1 (REQUIRED FOR DNA METHYLATION 1) is a single-stranded DNA binding protein with a strong preference for methylated DNA. Initial studies showed a loss of Pol V transcripts and a loss of Pol V association at target loci in *drd1* and *dms3* mutants. DRD1, DMS3 and RDM1 were subsequently shown to form a multifunctional complex, known as the DDR complex. The DDR complex elutes with Pol V during gel filtration chromatography, and immunoprecipitated DDR fractions that were analyzed by mass spectrometry include subunits of Pol V. Collective evidence suggests that the DDR complex might recruit Pol V to target loci and/or facilitate Pol V transcription through chromatin remodelling.

DMS4 (also known as RDM4), which is a homologue of the yeast Pol II-associated protein Iwr1, is a potential regulator of polymerase abundance or activity (Fig. 2, step 6). Iwr1 binds Pol II in the cytoplasm and directs Pol II import into the nucleus, possibly acting as a sensor for the fully assembled polymerase. Consistent with the yeast data, DMS4 interacts with both Pol II and Pol V in vivo. Yeast two-hybrid analyses indicate that DMS4 can interact with the Pol IV and Pol V largest subunits’ carboxy-terminal domains (CTDs), within their DeCL-like domains (FIG. 1). The DeCL-like domain has homology to the DEFECTIVE CHLOROPLAST AND LEAVES protein, which is needed for chloroplast rRNA biogenesis. However, NRPB1, which forms the largest subunit of Pol II, lacks this domain (Fig. 1), indicating that the DeCL-like domain cannot be the sole determinant of DMS4–polymerase interactions. Indeed, recent evidence shows that Iwr1 binds to Pol II in the active site cleft that is formed by the two largest subunits. Based on the yeast studies, it is likely that DMS4 function in plants has evolved with the Pol II, Pol IV and Pol V family in order to facilitate the nuclear import of all three enzymes.

**RISC recruitment to Pol V transcribed target sites.** AGO4 can be crosslinked to Pol V transcripts, presumably because of base-pairing of siRNAs with the Pol V transcripts (Fig. 2, step 7). AGO4 also interacts...
REVIEWS

Figure 2 | Model for the RNA-directed DNA methylation pathway in Arabidopsis thaliana. RNA polymerase IV (Pol IV) initiates the RNA-directed DNA methylation (RdDM) pathway (step 1), generating transcripts that are then copied into double-stranded RNA (dsRNA) by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) (step 2). The putative chromatin remodeler and/or helicase CLASSY 1 (CLSY1) assists in one or more of these steps. DICER-LIKE 3 (DCL3) cleaves the dsRNA into 24-nucleotide small interfering RNA (siRNA) duplexes (step 3) that are methylated at their 3’ ends by HUA-ENHANCER 1 (HEN1) (step 4). A single strand of the siRNA duplex associates with ARGONAUTE 4 (AGO4) to form an RNA-induced silencing complex (RISC)–AGO4 complex (step 5). Independently of siRNA biogenesis, Pol V transcription is assisted by the DDR complex (DRD1 (DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1), DMS3 (DEFECTIVE IN MERistem SILENCing 3) and RDM1 (REQUIRED FOR DNA METHYLATION 1)) and DMS4 (step 6). AGO4 binds Pol V transcripts through base-pairing with the siRNA and is stabilized by AGO4 interaction with the NRPE1 (the largest subunit of Pol V) carboxy-terminal domain (CTD) and KTF1 (KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1), which also binds RNA (step 7). IDN2 may stabilize Pol V transcript–siRNA pairing (step 8). The RDM1 protein of the DDR complex binds AGO4 and the de novo cytosine methyltransferase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), bringing them to Pol V transcribed regions and resulting in DNA methylation (step 9). Histone modifications resulting from the RdDM pathway include the removal of activating acetylation and methylation marks (the deacetylation of multiple Lys of several core histone proteins and the methylation (step 9). Histone modifications resulting from the RdDM pathway include the removal of activating acetylation and methylation marks (the deacetylation of multiple Lys of several core histone proteins and the methylation of histone H3 Lys 4 (H3K4)) and the establishment of alternative, repressive histone methylation marks (such as the methylation of H3K9 and H3K27), thereby facilitating transcriptional silencing (step 10). The dashed arrow indicates that Pol IV and RDR2 physically interact in vivo, as determined by mass spectrometry, co-immunoprecipitation and enzymatic assays (T. Ream, J.R.H. and C.S.P., unpublished observations). ssRNA, single-stranded RNA.

with a WG/GW-rich domain that is present in the CTD-terminal domain of the largest subunit of Pol V34,45 (FIGS 1, 3), indicating that a physical interaction between Pol V and/or its transcripts with AGO4 can occur. AGO4 also associates with chromatin at Pol V transcribed loci and this association is lost if Pol V transcription is abolished through mutation of its active site39. Collectively, these observations suggest that AGO4–RISC complexes are recruited to target loci through binding to Pol V transcripts and the Pol V largest subunit, thereby bringing the RISC complex into the proximity of the chromatin to be modified39 (FIG. 2, step 7).

Once recruited to sites of Pol V transcription, AGO4 may have a role in siRNA amplification as well as in the recruitment of chromatin modifiers. AGO proteins are modular, including MID and PAZ domains that are involved in small RNA binding and PIWI domains that account for the ‘slicing’ of long RNAs that are base-paired with AGO-associated siRNA46,47. An Asp-Asp-His triad is key to slicer activity and is present in AGO4 (REF. 50). Interestingly, mutagenesis of the AGO4 catalytic triad has differential effects on siRNA accumulation and DNA methylation at different loci50. One possibility is that AGO4-mediated slicing of Pol V transcripts generates new substrates for RDR2, facilitating dsRNA production and secondary siRNA biogenesis; this might explain the reduced levels of siRNAs that are detected at many loci when Pol V is mutated. It is possible that RNA cleavage and siRNA amplification is not needed at the loci that are not affected when AGO4 RNase activity is abolished. Alternatively, AGO6 or AGO9 may substitute for AGO4 at such loci (REFS 29,48).

At least two proteins are thought to assist in RISC recruitment to target sites. One is IDN2 (also known as RDM12), which is a protein that was identified in two independent genetic screens as a component of the RdDM pathway37,45. IDN2 binds dsRNAs that have 5’ overhangs35, but siRNA levels in idn2 mutants are unaffected, suggesting a role for IDN2 downstream of siRNA biogenesis, possibly in the stabilization of siRNAs that are base-paired with Pol V transcripts36 (FIG. 2, step 8).

A second protein that is thought to assist AGO4–RISC targeting is KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1; also known as RDM3 and SPT5-LIKE)51,52. KTF1 is similar to the yeast transcription elongation factor Spt5, and was found by mass spectrometry to be associated with immuno precipitated cauliflower Pol V19 (FIG. 3). However, loss of KTF1 does not cause a decrease in Pol V transcript levels, as would be expected if KTF1 were a positive...
Regulatory DNMT1 (DRM2) (Fig. 2, step 9). A related protein, DRM1, is expressed at very low levels and is thought to have a minor role. It is not yet clear how DRM2 is recruited to its sites of action, but the DDR complex may be key, based on the finding that the DMD1 subunit of the DDR complex interacts with both AGO4 and DRM2 and also binds methylated single-stranded DNA. As the DDR complex also facilitates Pol V transcription, this suggests that DDR might integrate the activities of Pol V, AGO4 and DRM2 in order to bring about de novo cytosine methylation at Pol V transcribed loci.

Chromatin modifications associated with RdDM. In addition to cytosine methylation, several chromatin modifying enzymes are known to have roles in silencing the loci that are subject to RdDM. Among these are histone deacetylases (HDACs), a broad-specificity enzyme that can remove the acetyl groups that are added by at least seven different A. thaliana histone acetyltransferases in vitro. UBIQUITIN PROTEASE 26 (UBP26; also known as SUP32) is required for histone H2B deubiquitylation, which, in turn, is required for the establishment of repressive histone H3 Lys 9 (H3K9) dimethylation. The Jumonji C domain-containing protein JM14 demethylates H3K4, which is a modification that is typical of actively transcribed loci, thereby paving the way for repressive histone modifications to be added. Collectively, the removal of active chromatin marks (histone acetylation and H3K4 trimethylation) and the establishment of repressive chromatin marks (H3K9 methylation and H3K27 methylation) contributes to transcriptional silencing at sites that are subject to Pol IV- and Pol V-mediated RdDM.

A role for Pol II in RdDM. In the fission yeast Schizosaccharomyces pombe, Pol II carries out the functions that are attributed to Pol IV and Pol V, generating both siRNA precursors and scaffold transcripts for siRNA binding at loci that are subject to RNA-directed silencing. In plants, Pol IV and Pol V have apparently taken over most of the RNA silencing duties of Pol II. However, recent findings indicate that Pol II still has a role in RdDM in plants. Specifically, a mutation in the gene encoding NRFB2 (the Pol II second-largest subunit) that causes developmental defects also causes subtle RdDM defects, including decreased siRNA accumulation and decreased silencing at some loci. Pol II requires DRD1 for transcription at some of these loci, interacts with AGO4, DMS4 and RDM1 and influences Pol IV and Pol V recruitment (Fig. 2). (Fig. 3), indicating that Pol II can effect multiple steps of the RdDM process in ways that are not yet understood.

Pol IV and RdDM proteins in paramutation. When brought together in the same nucleus, so-called ‘paramutagenic’ silent alleles can transform active, paramutable alleles into the silent, paramutagenic state, which is stable and heritable through meiosis. This remarkable phenomenon of paramutation occurs in...
plants and mammals and has been studied extensively in maize\textsuperscript{4,70}. Genetic screens have shown that the proteins that are required for maize paramutation include orthologues or paralogues of \textit{A. thaliana} proteins that are involved in RdDM. For instance, REQUIRED TO MAINTAIN REPRESSION 6 (RMR6) is the maize orthologue of NRPD1 (the largest subunit of \textit{A. thaliana} Pol IV\textsuperscript{66}). MEDIATOR OF PARAMUTATION 1 (MOP1) is the maize orthologue of \textit{A. thaliana} RDR2 (REF. 67). RMR1 is a maize SWI2/SNF2-like protein that is related to \textit{A. thaliana} CLSY1 and DRD1 (REF. 68), and MOP2 (also known as RMR7) is one of three maize homologues of NRPD2 (also known as NRPE2), which is the second-largest subunit of both \textit{A. thaliana} Pol IV and Pol V\textsuperscript{69,70}. Whether Pol V is also involved in paramutation is not yet clear, but the evidence clearly points to RNA as the conduit for inter-allelic communication and the transcriptional silencing of paramutable alleles.

### Spreading of RNA silencing

RNA silencing signals can spread locally within a tissue or organ as well as move long-distance between organs, presumably as an antiviral defence mechanism. Short-range spreading occurs by signals moving through intercellular connections (plasmodesmata) over a range of 10–15 cells, whereas long-range spreading involves signals that are transported in the phloem of the vascular system\textsuperscript{71}. Genetic screens to identify mobile silencing signalling components have made clever use of transgenes expressing inverted repeats of sequences of a gene whose knockdown produces a visual phenotype. When transcribed, the inverted repeats form dsRNAs that are diced into siRNAs, triggering the silencing of the homologous ‘reporter’ gene that is responsible for the phenotype being monitored. By expressing the inverted repeat in cells of leaf veins (specifically, phloem companion cells), silencing of the endogenous reporter gene in adjacent cells of the leaf can be monitored \textsuperscript{24,27,23}. Both 21- and 24-nucleotide siRNAs are produced during the cell–cell silencing process and Pol IV, RDR2 (REFS 24, 73) and CLSY1 (REF 24) are all involved. However, these three genes are dispensable for the accumulation of the predominant 21-nucleotide siRNAs from the inverted repeat triggers, suggesting that 24-nucleotide siRNAs arise as secondary siRNAs during spreading. Delivery of 21-nucleotide and longer RNA duplexes, but not single-stranded RNA, directly into cells by particle bombardment is sufficient to induce silencing that extends several cells beyond the bombardment site\textsuperscript{4}. Twenty four nucleotide siRNA duplexes are also capable of short-distance spreading and transgene silencing\textsuperscript{74}, suggesting that all siRNAs are mobile to some extent. The silencing that is associated with short-range siRNA mobility is presumably due to a combination of posttranscriptional silencing (mRNA slicing or inhibition of translation) and RdDM.

Long-distance spreading of RNA silencing signals has been studied by grafting roots and shoots of plants from different genetic or transgenic backgrounds and assessing whether silencing signals can traverse the graft junction. Deep sequencing of small RNAs suggests that all siRNA size classes are capable of moving from root to shoot, as well as from shoot to root\textsuperscript{75}. It has been found, through the analysis by high-throughput small RNA sequencing of siRNA spread from shoot to root, that NRPD1 and DICER function (DCL2, DCL3 and DCL4) are required for silencing signal production\textsuperscript{76}. The mobile 24-nucleotide siRNAs requiring NRPD1 and DICER activity can direct long-distance RdDM at transgenic or endogenous targets\textsuperscript{75,76}. These results run counter to an earlier report demonstrating that Pol IV and DCL2, DCL3 and DCL4 were dispensable for generating the silencing signals in the root that ultimately cause green fluorescent protein (GFP) reporter silencing in the shoot. However, Pol IV and DICER activities were required for perception or amplification of the silencing signal or signals in shoots\textsuperscript{77}. Moreover, siRNAs that were detected in shoot tissue, where GFP silencing occurred, nealy corresponded to sequences 3′ to the trigger sequences that were expressed in the root and not to the trigger sequences themselves\textsuperscript{78}. The different conclusions obtained from different studies might be reconciled by the possibility that long RNAs as well mobile siRNAs have roles in long-distance silencing, and that target genes in different contexts might be sensitive to different classes of mobile RNAs.

### Pol V and heterochromatin organization

Centromeric repeats and other heterochromatic sequences coalesce into chromocenters, which are intensively stained by DNA-binding dyes. An early observation was that chromocenter interactions are perturbed in \textit{nrpd2} mutants\textsuperscript{11}, which lack both Pol IV and Pol V activity. Subsequent studies of \textit{nrpd1} and \textit{nrpe1} mutants show that loss of Pol V, but not Pol IV, causes chromocenter disruption\textsuperscript{79}. Mutations in \textit{DRD1}, which assists Pol V transcription as part of the DDR complex, or in MET1, the major maintenance cytosine methyltransferase, similarly cause chromocenter disruption. This chromocenter disruption is coincident with the derepression of pericentromeric transcription units. By contrast, \textit{rdr2}, \textit{dcl3}, \textit{ago4} or \textit{drm2} mutations have no effect on chromocenter disruption or pericentromeric transcription\textsuperscript{79}. Pol V is also required for the condensation of a 5S rRNA gene locus on chromosome 4, independent of the RdDM pathway\textsuperscript{79}. Collectively, these observations suggest that Pol V transcripts might have roles as structural components of heterochromatin or as components of chromatin modification pathways that are independent of Pol IV or 24-nucleotide siRNAs.

### Pol IV and Pol V in development

The transition from vegetative growth to flowering in plants involves the reprogramming of stem cells in the shoot and is influenced by endogenous and environmental signals, including day length and temperature\textsuperscript{80}. \textit{A. thaliana} RdDM mutants are only slightly delayed in flowering under long-day, short-night conditions, but the delay in flowering is exacerbated when plants are grown under short-day and long-night conditions\textsuperscript{23,81,82}. FLOWERING LOCUS C (FLC), a repressor of flowering, is regulated during development by Pol IV-, RDR2-, DCL3- and AGO4-dependent chromatin
Regulating convergently transcribed genes

The *A. thaliana* genome, like that of other eukaryotes, has hundreds of convergently transcribed gene pairs with the potential to generate dsRNAs and siRNAs. The regulatory potential of this arrangement has been demonstrated for the salt stress response, bacterial defence and seed development. In the cases of the gene pairs that are involved in the salt stress and bacterial defence responses, one gene is constitutively active and the antisense transcript is induced as a result of the stress. The resulting siRNAs post-transcriptionally silence the constitutive locus, facilitating an adaptive response. In the case of the gene pair that affects fertilization and seed development, both transcripts are sperm cell-specific, giving rise to siRNAs that inhibit the expression of an E3 ubiquitin ligase that interferes with proper fertilization and seed development. Genetic evidence indicates that Pol IV is required in each of the three examples above, with Pol V involvement being variable, but precisely how Pol IV (or Pol V) contributes to these phenomena is unclear.

**Pol IV and Pol V: a diversified family**

The subunit compositions for *A. thaliana* Pol II, Pol IV and Pol V, which have been defined by mass spectrometry, reveal their striking similarities, with half of their 12 subunits encoded by the same genes (FIG. 4; BOX 2). The remaining subunits of Pol IV and/or Pol V are encoded by paralogues of Pol II subunit genes that underwent duplication at different times in plant evolution. Multiple subunits of cauliflower Pol V have also been identified by mass spectrometry, the gene encoding the fourth subunit of Pol IV and Pol V was identified in a genetic screen and a reverse genetics approach identified the fifth subunit of Pol V. All of these studies support the conclusion that Pol IV and Pol V evolved as specialized forms of Pol II.

Many of the intron–exon junctions of the Pol IV and Pol V largest subunit genes are identical to those in their Pol II paralogue, but not the corresponding Pol I or Pol III paralogues. Based on phylogenetic analyses of the largest subunits, Luo and Hall proposed that Pol IV evolved first through duplication of the Pol II largest subunit and Pol V evolved later, through duplication of the Pol IV largest subunit. This hypothesis is supported and extended by the recent determinations of the complete subunit compositions of Pol II, Pol IV and Pol V, which show that Pol IV is intermediate in subunit composition between Pol II and Pol V.

The second largest subunits of Pols IV and Pol V are encoded by the same gene, NRPD2 (also known as NRPE2), which is distinct from the Pol II NRBP2 gene. In *A. thaliana*, a gene closely related to NRPD2 was duplicated after the divergence of dicots and monocots, but it is a non-functional pseudogene. Interestingly, as mentioned previously, there are three NRPD2-like genes in maize, underscoring the fact that Pol IV and Pol V are still evolving in different plant lineages. In fact, the Pol IV and Pol V largest subunits, NRPD1 and NRPE1, have estimated amino acid substitution rates that are 20-fold greater than the Pol II largest subunit, and
REVIEWS

Box 2 | Functions of Pol II subunits and implications for Pol IV and Pol V

The two largest subunits, subunits 1 and 2, form the active site, DNA entry channel and RNA exit channel. The RNA polymerase II (Pol II), Pol IV and Pol V largest subunits have distinct carboxy-terminal domains (CTDs) (Fig. 1). The CTD of the Pol II largest subunit mediates its interactions with numerous regulatory proteins, and the long CTD of the Pol V largest subunit is likely to serve as an interaction site for multiple regulatory proteins, including ARCGONAUTA 4 (AGO4). In yeast, the Rpb3, Rpb10, Rpb11 and Rpb12 subunits help to assemble and stabilize Pol II. Arabidopsis thaliana Pol II, Pol IV and Pol V make use of the same genes for these four subunits, but Pol V can use an alternative form of the third subunit (NRPE3b) that is not appreciably used by Pol II or Pol IV. Whether this results in a Pol V isoform with unique properties is unclear.

Yeast Rpb4, Rpb5, Rpb6, Rpb7, Rpb8 and Rpb9 influence the functionality of Pol II. Rpb4 and Rpb7 are positioned near the RNA exit pore and interact to form a dissociable subcomplex in yeast. Rpb7 has an RNA binding domain and mutations in Rpb7 disrupt small interfering RNA (siRNA)-dependent heterochromatin formation in Schizosaccharomyces pombe. A. thaliana Pol II, Pol IV and Pol V each have unique fourth–seventh subunit subcomplexes, which are likely to confer distinctive properties on the three enzymes.

In yeast, Rpb5 and Rpb9 are positioned near where DNA enters the polymerase, with Rpb5 contacting the DNA. Rpb9 is involved in transcription start site selection, processivity and transcript cleavage, which is mediated by the elongation factor TFIIS. The fifth subunits of A. thaliana Pol I, Pol II, Pol III and Pol IV are encoded by the same gene, which is the yeast Rpb5 orthologue. However, Pol V uses a unique fifth subunit, NRPE5, that imparts unique template recognition properties on the enzyme. A. thaliana has two highly similar Rpb9-like genes. It is unclear whether they are redundant with respect to Pol II, Pol IV and Pol V.

The RNA exit pore separates Rpb6 and Rpb8 in yeast Pol II. A. thaliana Pol II, Pol IV and Pol V use the same genes for these subunits.

NRPD2 has a substitution rate that is ten times greater than its Pol II counterpart. A. thaliana has two highly similar NRPB3 genes, one used by Pol II, Pol IV and Pol V and the other used only in a subset of Pol V complexes. Whether the use of this alternative third subunit results in functionally distinct forms of Pol V is unclear.

The fourth subunits of Pol IV and Pol V are encoded by the same gene, NRPD4 (also known as NRPE4), which is distinct from the Pol II NRPB4 gene.

Subunits six, eight, ten and twelve are common to Pol I, Pol II and Pol III, from yeast to humans. These subunits are also common to Pol II, Pol IV and Pol V in A. thaliana, as are subunits nine and eleven. For most of these subunits, there are duplicate genes whose redundancy is assumed but not proven.

The fifth subunits of Pol I, Pol II, Pol III and Pol IV are encoded by the same gene, from yeast to humans. However, Pol V makes use of a distinct fifth subunit, NRPE5 (REFS 15, 16, 52), which is likely to impart different template specificities and/or interactions with regulatory proteins based on the roles of the Pol II form of this subunit. The yeast Rpb4 and Rpb7 subunits of Pol II form a subcomplex that has roles in transcription initiation and RNA 3′ end processing. This Rpb4–Rpb7 subcomplex is able to dissociate from Pol II and chaperone mRNA to the cytoplasm to stimulate mRNA decay. The A. thaliana fourth–seventh subunit subcomplexes of Pol II, Pol IV and Pol V are unique in all three polymerases owing to the use of distinct seventh subunits and the use of a fourth subunit in Pol IV and Pol V that is different from the NRPB4 subunit of Pol II.

Immunolocalization of NRPD4 reveals that the protein does not always colocalize with Pol IV and Pol V catalytic subunits, suggesting that the fourth–seventh subunit subcomplexes of Pol IV and/or Pol V may also be able to dissociate from the polymerase holoenzyme, as in yeast, possibly having roles as a chaperone of Pol IV and/or Pol V transcripts.

The subunits that are unique to Pol II, Pol IV and Pol V must be responsible for the unique functions of the three polymerases. Future studies aimed at uncovering the functions of these subunits are likely to yield important insights.

Pol IV and Pol V as enzymes

Pol IV and Pol V enzymatic activity has not yet been demonstrated in vitro, despite attempts by several groups. Many of the amino acids that are invariant in the catalytic subunits of Pol I, Pol II and Pol III are different in Pol IV and Pol V. Interestingly, the sequences of the Pol IV and Pol V catalytic subunits that have diverged most are in the vicinity of their active sites. Nonetheless, the core sequences of the Metal A and Metal B sites (which coordinate two magnesium ions that are essential for positioning the incoming nucleotide triphosphate relative to the 3′ end of the growing RNA chain within the active site) are retained in Pol IV and Pol V, as in all other multisubunit RNA polymerases. Clustered site-directed mutagenesis of these essential amino acids abolishes all known biological functions of Pol IV and Pol V, as do single mutations. Moreover, Pol V-dependent transcripts that are detected in vivo have either 5′ triphosphates or 7-methylguanosine caps, but not 5′ phosphates or 5′ hydroxyl groups, consistent with being initiated, rather than processed, RNAs. Collectively, the evidence supports the hypothesis that Pol IV and Pol V are DNA-dependent RNA polymerases, despite the lack of biochemical proof in vitro.

Conclusions and perspective

Pol IV and Pol V have diverse roles in non-coding RNA-mediated silencing and heterochromatin formation. Most of what we know about these activities is based on genetic analyses, as development of biochemical assays has been problematic. Numerous questions are in need of answers. What are the templates for Pol IV and Pol V? Are these templates DNA, RNA, chromatin that bears specific modifications, or some combination thereof? What recruits Pol IV or Pol V to their sites of action? Do Pol IV and Pol V recognize conventional promoters, assisted by transcription factors, in a manner similar to Pol II? How do changes in amino acids that are otherwise invariant in other RNA polymerases affect the templates that are recognized by Pol IV and Pol V, or the processivity or fidelity of the enzymes? How do the subunits that differ between Pol II, Pol IV and Pol V account for their different functions? What is the full, functional significance of the unique NRPD1 and NRPE1 CTDs? Are there tissue- or development-specific isoforms of Pol IV and/or Pol V that make use of alternative subunit variants? These and other questions abound, and answers are few, presenting important challenges for studies in the years to come.
siRNA-mediated transcriptional silencing. This paper identified Pol V transcripts and adjacent genes. Pol V mediates transcriptional silencing of overlapping RdDM downstream of siRNA biogenesis, thus showing that Pol V transcription traversing the nuclear enzyme that is required for heterochromatin formation.

Evolutionary history of plant multisubunit RNA polymerases IV and V relative to other RNA polymerases. This minireview presents a mechanistic discussion, and interpretation of ongoing research, highlighting the surprising amino acid divergence in the catalytic centres of Pol IV and Pol V relative to other RNA polymerases.

Using a reverse genetics approach, these authors show that Pol V transcripts and the roles in each enzyme for siRNA biogenesis. This paper presented a genome-wide study to show that Pol IV and Pol V are required for siRNA biogenesis at thousands of loci in the Arabidopsis thaliana genome.

siRNA-dependent and adjacent genes. This paper showed that Pol IV and Pol V have distinct roles in RdDM.

Atypical RNA polymerase subunits are required for RNA-directed DNA methylation. The authors presented a mechanistic discussion, and interpretation of ongoing research, highlighting the surprising amino acid divergence in the catalytic centres of Pol IV and Pol V relative to other RNA polymerases.

RdDM downstream of siRNA biogenesis, thus showing that Pol V transcription traversing the nuclear enzyme that is required for heterochromatin formation.

Evolutionary history of plant multisubunit RNA polymerases IV and V relative to other RNA polymerases. This minireview presents a mechanistic discussion, and interpretation of ongoing research, highlighting the surprising amino acid divergence in the catalytic centres of Pol IV and Pol V relative to other RNA polymerases.

Using a reverse genetics approach, these authors show that Pol V transcripts and the roles in each enzyme for siRNA biogenesis. This paper presented a genome-wide study to show that Pol IV and Pol V are required for siRNA biogenesis at thousands of loci in the Arabidopsis thaliana genome.

siRNA-dependent and adjacent genes. This paper showed that Pol IV and Pol V have distinct roles in RdDM.

Atypical RNA polymerase subunits are required for RNA-directed DNA methylation. The authors presented a mechanistic discussion, and interpretation of ongoing research, highlighting the surprising amino acid divergence in the catalytic centres of Pol IV and Pol V relative to other RNA polymerases.
REVIEWS


This study revealed that the abundant siRNAs that are produced in developing seeds are entirely derived from the maternal chromosomes owing to a previously unknown imprinting mechanism.


This intriguing paper shows that the vegetative cell of pollen derepresses transposons in order to generate siRNAs that move to the adjacent sperm cells to programme gene silencing.


The authors report how siRNA-mediated regulation of convergently transcribed gene pairs can mediate an adaptive response.


This interesting paper details the stepwise evolution of Pol IV and Pol V catalytic subunits from genes encoding Pol II subunits, beginning in green algae.


Acknowledgements
Our studies of RNA polymerases IV and V are supported by grants from the US National Institutes of Health (GM077590) and the Monsanto Company. All opinions expressed are those of the authors and do not necessarily reflect the views of our sponsors. We apologize to our many colleagues whose work was not cited owing to limitations on the numbers of references.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Craig S. Pikaard’s homepage:
http://sites.bio.indiana.edu/~sikaard

ALL LINKS ARE ACTIVE IN THE ONLINE PDF