An Internal Rearrangement in an Arabidopsis Inverted Repeat Locus Impairs DNA Methylation Triggered by the Locus

Stacey Melquist[1] and Judith Bender[2]

Department of Biochemistry and Molecular Biology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland 21205

Manuscript received August 6, 2003
Accepted for publication September 24, 2003

ABSTRACT

In plants, transcribed inverted repeats trigger RNA interference (RNAi) and DNA methylation of identical sequences. RNAi is caused by processing of the double-stranded RNA (dsRNA) transcript into small RNAs that promote degradation of complementary RNA sequences. However, the signals for DNA methylation remain to be fully elucidated. The Arabidopsis tryptophan biosynthetic *PAI* genes provide an endogenous inverted repeat that triggers DNA methylation of *PAI*-identical sequences. In the Wassilewskija strain, two *PAI* genes are arranged as a tail-to-tail inverted repeat and transcribed from an unmethylated upstream promoter. This locus directs its own methylation, as well as methylation of two unlinked singlet *PAI* genes. Previously, we showed that the locus is likely to make an RNA signal for methylation because suppressed transcription of the inverted repeat leads to reduced *PAI* methylation. Here we characterize a central rearrangement in the inverted repeat that also confers reduced *PAI* methylation. The rearrangement creates a premature polyadenylation signal and suppresses readthrough transcription into palindromic *PAI* sequences. Thus, a likely explanation for the methylation defect of the mutant locus is a failure to produce readthrough dsRNA methylation triggers.

CYTOSINE methylation plays a critical role in directing patterns of heterochromatin formation in the genomes of mammals and plants, with effects on both gene expression and genome stability. In mammals, methylation is required for essential developmental programs including X chromosome inactivation in females and genomic imprinting (reviewed in Bird 2002). In plants, methylation is required for inactivation of invasive parasitic sequences such as transposable elements (Miura *et al.* 2001; Singer *et al.* 2001; Kato *et al.* 2003).

One mechanism of guiding cytosine methylation to appropriate genomic loci involves an RNA signal with sequence identity to the DNA target. This process of RNA-directed DNA methylation has been well documented in plants (reviewed in Matzke *et al.* 2001), and a potentially related process of RNA-directed heterochromatin formation occurs in the fungus *Schizosaccharomyces pombe* (Volpe *et al.* 2002; Schramke and Allshire 2003). RNA-directed DNA methylation is interrelated with another silencing mechanism, RNA interference (RNAi). During RNAi, accumulation of double-stranded RNA (dsRNA) triggers a two-step RNA degradation process. In the first step, the dsRNA is cleaved by dicer RNAse activities to yield small interfering RNAs (siRNAs) of ~25 nucleotides (nt). In the second step, these siRNAs are incorporated into an RNA-induced silencing complex and used as guides to target degradation of complementary transcripts. The observation made from plant systems that generate high levels of dsRNAs, including infecting RNA viruses and highly transcribed inverted repeat transgenes, is that RNAi triggered by the dsRNA is typically accompanied by dense methylation of DNA sequences with homology to the dsRNA precursor and its siRNA products (Dalmay *et al.* 2000; Mette *et al.* 2000; Sijen *et al.* 2001; Vaistij *et al.* 2002). Furthermore, mutations in factors that control the processing of aberrant transcripts into dsRNA, such as an RNA-dependent RNA polymerase mutation, block both RNAi and DNA methylation (Dalmay *et al.* 2000, 2001; Fagard *et al.* 2000; Mourrain *et al.* 2000; Béclin *et al.* 2002). These findings argue that dsRNA or siRNAs are likely to be the triggers for RNA-directed DNA methylation as well as RNAi. However, the mechanistic relationship between the two silencing pathways remains to be fully elucidated.

The phosphoribosylanthranilate isomerase (PAI) tryptophan biosynthetic genes in Arabidopsis provide a model system to study RNA signals for DNA methylation of relatively low-expression endogenous genes. The Wassilewskija (WS) strain of Arabidopsis carries a *PAI1–PAI4* inverted repeat gene arrangement plus unlinked singlet *PAI2* and *PAI3* genes, and all four genes are densely

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession no. AY357734.

[1]*Present address:* Mayo Clinic, 4500 San Pablo Rd., Jacksonville, FL 32224.

[2]*Corresponding author:* Department of Biochemistry and Molecular Biology, Johns Hopkins University Bloomberg School of Public Health, 615 N. Wolfe St., Baltimore, MD 21205.
E-mail: jbender@mail.jhmi.edu

methylated over their regions of sequence identity at both CG and non-CG cytosines (Bender and Fink 1995; Luff *et al.* 1999). Of the four genes, only *PAI1* and *PAI2* encode functional PAI enzyme, and only *PAI1* is significantly expressed due to a novel unmethylated promoter that lies upstream of the *PAI1* gene (Melquist *et al.* 1999; Melquist and Bender 2003). The *PAI2* gene is transcriptionally silenced by methylation of its more proximal promoter sequences. Mutations in WS *PAI1*, including missense mutations in the *PAI1* coding sequence (Bartee and Bender 2001) and complete deletion of the *PAI1–PAI4* inverted repeat genes (Bender and Fink 1995), display a PAI-deficient blue fluorescent phenotype due to accumulation of an early intermediate in the tryptophan pathway. Transgene-induced silencing of the upstream promoter that drives *PAI1* expression also confers a blue fluorescent phenotype (Melquist and Bender 2003).

In previous work, we showed that the *PAI1–PAI4* inverted repeat triggers *de novo* methylation of unmethylated *PAI* sequences (Luff *et al.* 1999). Furthermore, we found that the inverted repeat locus is required for maintenance of dense methylation on the unlinked *PAI2* and *PAI3* genes; when the *PAI1–PAI4* locus is deleted or segregated away through genetic crosses, the *PAI2* and *PAI3* genes lose most of their non-CG methylation (Bender and Fink 1995; Jeddeloh *et al.* 1998; Luff *et al.* 1999). The residual CG methylation on these loci is maintained as a relic of the previous dense methylation imprint and can be lost through subsequent rounds of DNA replication (Bender and Fink 1995). The inverted repeat locus is likely to produce an RNA signal for *PAI* DNA methylation, because when transcription of this locus is suppressed, methylation on the *PAI2* and *PAI3* genes is reduced to primarily CG maintenance methylation (Melquist and Bender 2003). However, full-length *PAI* transcripts are not efficiently degraded by RNAi. Furthermore, only a minority of accumulated *PAI* transcripts extend beyond a major polyadenylation site at the end of *PAI1* to include palindromic *PAI4* sequences, and although this *PAI* dsRNA is presumably a substrate for dicer cleavage, *PAI* siRNAs are not detectable by gel blot. Together, these observations lead to the view that RNA-directed DNA methylation can be promoted by much lower levels of trigger RNA species than RNAi can, but do not discriminate whether *PAI* dsRNA, *PAI* siRNA, or some other *PAI*-derived aberrant RNA species guides DNA methylation.

Here we describe a novel rearrangement mutation in the *PAI1–PAI4* locus that creates a new premature polyadenylation site, suppresses readthrough transcripts from *PAI1* into *PAI4*, and impairs both maintenance and *de novo* methylation of *PAI* sequences. This rearrangement mutant thus provides evidence supporting dsRNA or a processed product of dsRNA as the *PAI* methylation trigger.

MATERIALS AND METHODS

Isolation and cloning of the WS *invpai1-Δpai4* mutant: Ethyl methanesulfonate (EMS)-mutagenized WS M2 seed pools were purchased from Lehle Seed, although as noted in the discussion, the nature of the rearrangement mutation recovered is inconsistent with a direct effect of EMS mutagenesis. Seeds were surface sterilized, plated on plant nutrient plus 0.5% sucrose medium (Haughn and Somerville 1986) with 0.75% agar, and screened at 2 weeks postgermination with a hand-held short wave ultraviolet (UV) light source for blue fluorescent mutants. Complementation crosses of the *invpai1-Δpai4* mutant isolate with a blue fluorescent Δ*pai1-pai4* deletion mutant (Bender and Fink 1995) indicated that the phenotype was caused by a *pai1* lesion. The mutant *invpai1-Δpai4* locus was cloned by making a λDASH (Stratagene, La Jolla, CA) library from *Bam*HI-cleaved genomic DNA and screening plaques by hybridization with a probe to direct repeat sequences that flank the inverted repeat *PAI* locus. The entire locus was recovered on a 17-kb *Bam*HI fragment. Restriction mapping and sequencing of the central 6.3 kb of this fragment (GenBank no. AY357734) indicated that the mutant differed from parental WS only in the central region between *PAI1* and *PAI4*, as described in detail in results.

DNA and RNA analysis: A *PAI1* (At1g07780) cDNA internal 0.7-kb *Pst*I fragment probe that hybridizes to all four WS *PAI* genes was used for DNA and RNA gel-blot analysis of *PAI* sequences (Bender and Fink 1995). This same sequence was used as a template for generating antisense or sense-strand *PAI* RNA probes. A β-tubulin (At5g44340) cDNA probe was used as a gel loading control in RNA blot analysis. A 180-bp centromere repeat probe from plasmid pARR20-1 (gift of E. Richards, Washington University, St. Louis) was used for analysis of centromere methylation status. DNA probes were labeled using the Amersham (Buckinghamshire, UK) MegaPrime kit, and RNA probes were labeled using the Promega (Madison, WI) *in vitro* transcription kit and T3 polymerase. Bisulfite sequencing of methylation patterns on the upstream sequences of *PAI1* and *PAI2* was performed as previously described (Melquist and Bender 2003). Polyadenylated and nonpolyadenylated RNAs were fractionated from total RNA using oligo(dT) magnetic beads (Dynal, Great Neck, NY). Ambion Millenium RNA markers were used to determine transcript length. Rapid amplification of cDNA ends (RACE) analysis was performed as previously described (Melquist and Bender 2003) using the SMART kit (CLONTECH, Palo Alto, CA), except that a primer in the third *PAI* exon, P137, 5′-CACCAACAGGTTTGGCCCCACCTTCCC-3′, was used for 5′ RACE analysis. This primer is completely identical to all four WS *PAI* genes and lies outside the deleted region in WS *invpai1-Δpai4*. For reverse transcriptase-PCR (RT-PCR) analysis of *PAI1* transcript 5′ ends, the primers used were S15a-RTF1, 5′-GAGTACCTTGCCTCTCGAGCTCCC-3′ in the first upstream exon and P129, 5′-CATCATCCTTAGGAGCTA CATTC-3′ in the third *PAI* exon.

Genetic analysis of WS *invpai1-Δpai4*: The presence of the *invpai1-Δpai4* rearrangement was detected in segregating populations using PCR primers that flank the Δ*pai4* deletion: PIF, 5′-CCGCCGCGTCTCTGCTGACCC-3′ and PIR, 5′-GATTG GAAACAATAGGTTGATGC-3′. The primers yield a 1642-bp product from *PAI2* and *PAI3*, a 1633-bp product from *invpai1*, and a 1123-bp product from Δ*pai4*. Plants homozygous for the rearrangement mutation were further identified by their fluorescent phenotype. In crosses between WS and Columbia (Col), segregating *PAI* loci were scored using polymorphisms linked to each locus (Bender and Fink 1995).

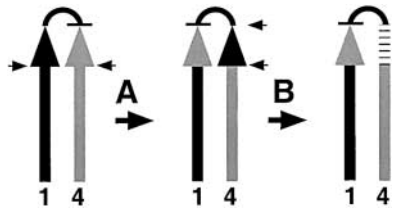

FIGURE 1.—A model for the genesis of the *invpai1-Δpai4* rearrangement. The solid vertical arrow represents *PAI1* and the shaded vertical arrow represents *PAI4*, with the end of the *PAI1* sequences in the central region indicated by a horizontal line. Small arrows indicate points of breakage and rejoining for (A) an inversion between the 3′ ends of *PAI1* and *PAI4* and (B) a deletion of part of the inverted region. The deleted region is indicated by a dashed line. On the basis of the positions of *PAI1 vs. PAI4* polymorphisms in *invpai1*, the inversion breakpoint occurred somewhere in the 421-bp region between the 3′ end of the third exon and the 5′ end of the fifth exon, consistent with the same breakpoint defining the end of the deleted sequences in the *Δpai4* fourth exon.

RESULTS

An internal inversion/deletion mutation in the *PAI–PAI4* inverted repeat locus confers PAI-deficient phenotypes: In the course of screening for blue fluorescent mutants in the WS strain, we isolated an unusual mutant with an internal rearrangement in the *PAI1–PAI4* inverted repeat locus. The fluorescent mutant was initially identified as having a *PAI1* defect by its failure to complement the fluorescent phenotype of the WS *Δpai1-pai4* strain. Southern blot analysis of mutant genomic DNA showed that there was a partial deletion in the *PAI1–PAI4* genes (see below). To understand the nature of the rearrangement in detail, we cloned and sequenced the *PAI1–PAI4* locus from the mutant. This analysis revealed that the rearrangement consisted of an inversion of the central sequences in the locus together with a 519-bp deletion extending from the noncoding sequences between the two *PAI* genes into the middle of the *PAI4* fourth exon (Figure 1). Although the overall structure of the *PAI1* coding region was intact in this rearrangement, the central inversion introduced a fifth exon 9-bp deletion normally found in the *PAI4* gene into the *PAI1* gene. This small deletion, which removes three amino acids from the coding sequence, was previously shown to abrogate PAI enzyme function (MELQUIST *et al.* 1999). Therefore, the inversion event introduced a loss-of-function mutation into the *PAI1* coding sequence, accounting for the fluorescent PAI-deficient phenotype of the mutant.

The inversion event also changed the sequences immediately downstream of *PAI1*. In WS, the *PAI1–PAI4* inverted repeat is centrally asymmetric, with the 3′ untranslated sequences downstream of *PAI1* extending for 263 bp before colliding with the 3′ untranslated sequences downstream of *PAI4*, which extend for only 20 bp (MELQUIST *et al.* 1999). This central asymmetry was inverted in the rearrangement mutant so that *PAI1* now had only 20 bp of *PAI* 3′ untranslated sequences before colliding with *PAI4* 3′ untranslated sequences, which include 246 bp of this region before the deletion junction. Overall, the rearrangement altered the palindromic structure of the locus so that the palindromic arms were shortened by ~500 bp and the central nonpalindromic sequences were increased by ~500 bp relative to the WS structure. We subsequently refer to the inversion/deletion rearrangement allele as WS *invpai1-Δpai4*.

The WS *invpai1-Δpai4* rearrangement mutant displays reduced *PAI* DNA methylation and silencing: WS *invpai1-Δpai4* mutant genomic DNA was tested for *PAI* methylation changes by both Southern blot and genomic sequencing assays. Southern blot analysis with the methylation-sensitive isoschizomers *Hpa*II and *Msp*I revealed that the mutant DNA had partially reduced methylation at all three *PAI* loci (Figure 2). *Hpa*II and *Msp*I both cleave the sequence 5′-CCGG-3′, but *Hpa*II is inhibited by methylation of either the inner (CG) or the outer (CCG) cytosine whereas *Msp*I is inhibited only by methylation of the outer cytosine. Each *PAI* locus contains a single *Hpa*II/*Msp*I site in the second intron, with flanking sites in unmethylated sequences at different distances from the central site for each locus (BENDER and FINK 1995; LUFF *et al.* 1999). In WS, the three *PAI* loci are highly refractory to cleavage by either *Hpa*II or *Msp*I, diagnostic of dense CG and CCG methylation of the recognition site. In the WS *invpai1-Δpai4* mutant, all three loci displayed increased cleavage by *Hpa*II and *Msp*I. In contrast, there was no difference between wild type and mutant in *Hpa*II/*Msp*I cleavage patterns at methylated centromere repeat sequences, indicating that the methylation changes are specific to *PAI* sequences. The previously characterized WS *Δpai1-pai4* mutant (BENDER and FINK 1995), with a complete deletion of the inverted repeat *PAI* genes, displayed a similar pattern of increased *Hpa*II and *Msp*I cleavage specifically at *PAI* loci.

To understand *PAI* methylation patterning in more detail, we performed sodium bisulfite genomic sequencing on mutant DNA in the regions upstream of *PAI1* or *PAI2*, extending from flanking heterologous sequences unique to each gene into *PAI*-identical proximal promoter sequences. Previous sequencing of the same regions in WS showed that the *PAI*-identical regions of both genes are densely methylated at CG and non-CG cytosines with very little spread into the flanking heterologous sequences (LUFF *et al.* 1999). In the WS *invpai1-Δpai4* mutant, we found that methylation in the *PAI*-identical region was reduced for both *PAI1* and *PAI2*, with a strong loss of non-CG methylation (Figure 3). This pattern is similar to that previously observed on the *PAI2* gene when the *PAI1–PAI4* locus was replaced with a singlet *PAI1* gene crossed in from the *PAI*-unmethylated Col strain background (Hyb4 in LUFF *et*

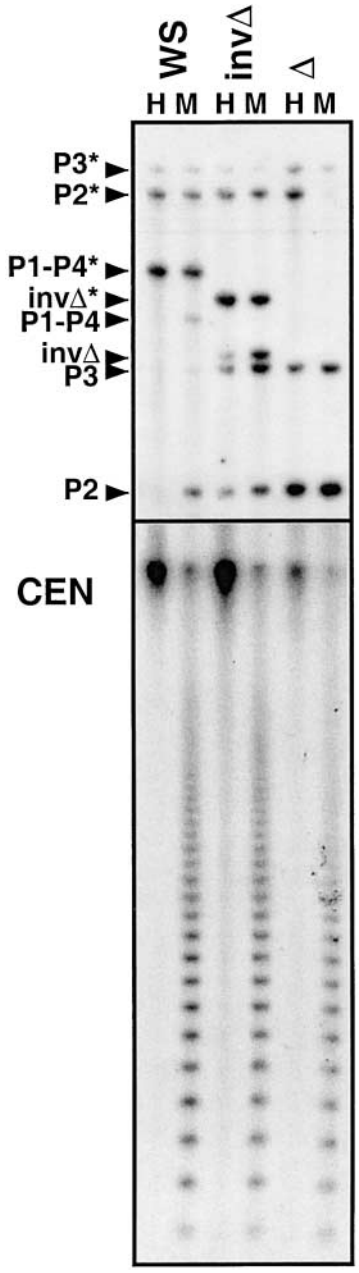

Figure 2.—The *invpai1-Δpai4* rearrangement confers reduced density of *PAI* DNA methylation. Genomic DNA samples prepared from WS, WS *invpai1-Δpai4* (invΔ), or WS *Δpai1-pai4* (Δ) were digested with the methylation-sensitive isoschizomers *Hpa*II (H) and *Msp*I (M). (Top) Samples were probed with a *PAI* internal cDNA fragment. Methylated bands are denoted with asterisks. P1–P4 is *PAI1–PAI4*, invΔ is *invpai1-Δpai4*, P2 is *PAI2*, and P3 is *PAI3*. (Bottom) The same samples were probed with a 180-bp centromere repeat probe (CEN).

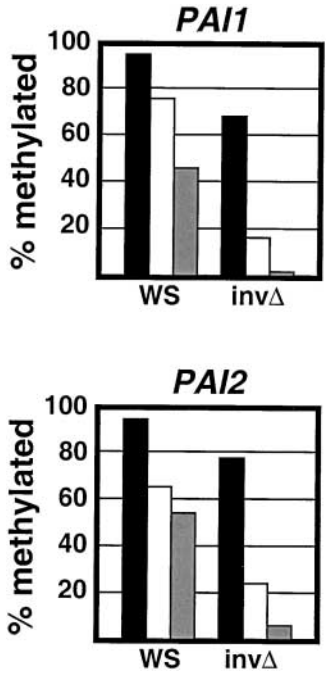

Figure 3.—Genomic bisulfite methylation sequencing data for *PAI1* and *PAI2* proximal promoter regions in *invpai1-Δpai4*. Eight independent top-strand clones were sequenced for *PAI1* or *PAI2*. The percentage of 5-methyl-cytosines out of total cytosines sequenced within the region of *PAI* sequence identity (344 bp for *PAI1* or 338 bp for *PAI2*) is shown, divided into the contexts CG (solid), CNG (open), and other contexts (shaded). For comparison, previously determined wild-type WS *PAI1* and *PAI2* data are shown (Luff *et al.* 1999).

al. 1999), when the *PAI1–PAI4* locus is deleted (Jeddeloh *et al.* 1998) or when transcription through the *PAI1–PAI4* inverted repeat is suppressed by targeted methylation of the upstream promoter region (Melquist and Bender 2003). Because non-CG methylation is diagnostic of RNA-directed DNA methylation (Pélissier *et al.* 1999), the loss of non-CG methylation in the *invpai1-Δpai4* mutant implies a defect in an RNA signal.

The WS *invpai1-Δpai4* mutant, with a defective *pai1* gene product, is relatively weakly fluorescent with only modest effects on plant morphology and fertility (Figure 4). In contrast, a WS *pai1* missense mutant, with a defective *pai1* gene product and a heavily methylated and silenced *PAI2* gene, is strongly fluorescent in all parts of the plant and has reduced size and fertility relative to the parental WS strain (Bartee and Bender 2001). This comparison suggests that the reduced methylation on *PAI2* in the *invpai1-Δpai4* mutant partially relieves its transcriptional silencing and partially compensates for the *pai1* defect, as we previously showed in other strain backgrounds in which *PAI1* activity is compromised and *PAI2* is demethylated (Bender and Fink 1995; Jeddeloh *et al.* 1998; Bartee and Bender 2001; Bartee *et al.* 2001; Malagnac *et al.* 2002). However, we could not directly detect increased steady-state message levels for the *PAI2* gene in the WS *invpai1-Δpai4* mutant *vs.* parental WS because in both strains *PAI2* expression is masked by stronger expression from the *PAI1* locus (see below).

The weak PAI-deficient phenotypes of the WS *invpai1-Δpai4* mutant were similar to those observed for the WS

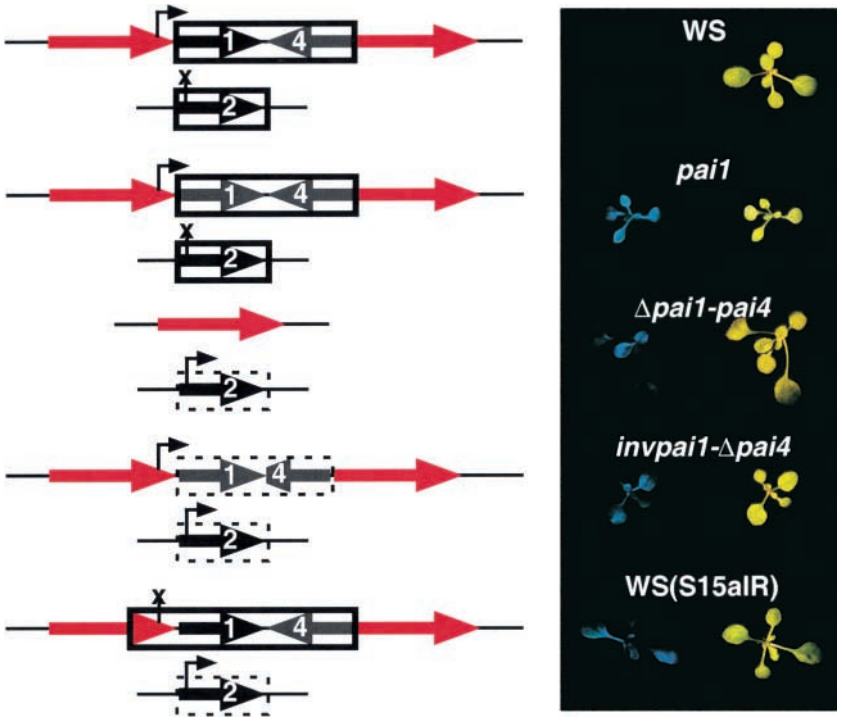

FIGURE 4.—The *invpai1-Δpai4* rearrangement confers a PAI-deficient blue fluorescent phenotype. (Left) Diagrams indicate the functional status, methylation density, and transcriptional activity of the *PAI1* and *PAI2* genes in *invpai1-Δpai4* and other mutants with defects at the *PAI1–PAI4* locus, with wild-type WS for comparison. *PAI* genes that encode functional enzyme are indicated by solid arrows, and *PAI* genes that encode nonfunctional enzymes are indicated by shaded arrows. Direct repeats that flank the *PAI1–PAI4* locus are shown as red arrows. Boxes around the *PAI* genes indicate DNA methylation, with a solid line indicating dense CG and non-CG methylation and a dashed line indicating reduced, mostly CG methylation. Arrowheads indicate transcription, and X's indicate transcriptional silencing. (Right) Visible (left) and UV (right) light photographs of representative 2-week-old seedlings of the indicated genotypes.

Δ*pai1-pai4* mutant, which has a complete deletion of the inverted repeat and reduced *PAI2* methylation (Figure 4; BENDER and FINK 1995). However, the *invpai1-Δpai4* mutant phenotypes were stable, with no nonfluorescent revertants detected out of thousands of plants screened, whereas the Δ*pai1-pai4* mutant phenotypes were only semistable, with 1–5% nonfluorescent and *PAI*-demethylated progeny resulting in each generation of self-pollination (BENDER and FINK 1995). This difference suggests that the partial deletion of *PAI1–PAI4* is able to maintain residual methylation of the *PAI2* gene better than the complete deletion of the locus. The weak PAI-deficient phenotypes of the WS *invpai1-Δpai4* mutant were also similar to those observed for WS carrying a transgene S15aIR that triggers methylation and silencing of the upstream promoter that drives *PAI1* transcription, with a concomitant reduction in *PAI2* methylation (Figure 4; MELQUIST and BENDER 2003). Like the WS *invpai1-Δpai4* mutant, the WS(S15aIR) strain is stably fluorescent, implying that this strain can maintain residual *PAI2* methylation, in contrast to a complete deletion of *PAI1–PAI4*.

The WS *invpai1-Δpai4* rearrangement does not affect transcription initiation or 5′ processing of the *PAI1* transcript: To understand the effects of the WS *invpai1-Δpai4* rearrangement mutation on *PAI* transcripts, we performed both RNA gel-blot analysis and RACE PCR analysis of *PAI* RNAs. Previous analysis of *PAI* transcripts in parental WS showed that *PAI1* is the only significantly expressed gene (MELQUIST *et al.* 1999; MELQUIST and BENDER 2003). Accumulated *PAI1* transcripts include a major correctly spliced and polyadenylated transcript of ~1200 nt, a polyadenylated species of ~1900 nt, and low levels of longer polyadenylated transcripts (MELQUIST and BENDER 2003; Figure 5). RNA gel-blot analysis with a *PAI* cDNA probe showed that the WS *invpai1-Δpai4* mutant accumulated a major broad band of transcripts ranging from ~1000 to 1200 nt (Figure 5A). This major transcript population was ~2.5-fold more abundant than the major 1200 nt transcript population in WS. In addition, the mutant accumulated an ~1900-nt species. As in WS, the mutant *PAI* RNAs were recovered predominantly in the polyadenylated fraction.

For RACE analysis, the 5′ or 3′ ends of *PAI* cDNAs were amplified by RT-PCR from WS *invpai1-Δpai4* RNA using an anchoring primer at the transcript end plus an internal *PAI* gene-specific primer. Individual PCR products were cloned and sequenced. The *PAI* gene-specific primers were designed to avoid regions that contain polymorphisms among the various WS *PAI* genes and also to avoid the deleted region in Δ*pai4*, so that transcripts from any of the *PAI* genes could potentially be amplified. The RACE analysis showed that all the detectable transcripts in the WS *invpai1-Δpai4* mutant corresponded to *PAI1* (Table 1; Table 2). This result indicates that even if other *PAI* genes such as *PAI2* are reactivated by the reduced methylation in the mutant, *PAI1* transcripts still accumulate to the highest steady-state levels.

5′ RACE analysis showed that mutant *PAI1* transcripts initiated in the same region and had the same two types of splice variant structures as previously determined by

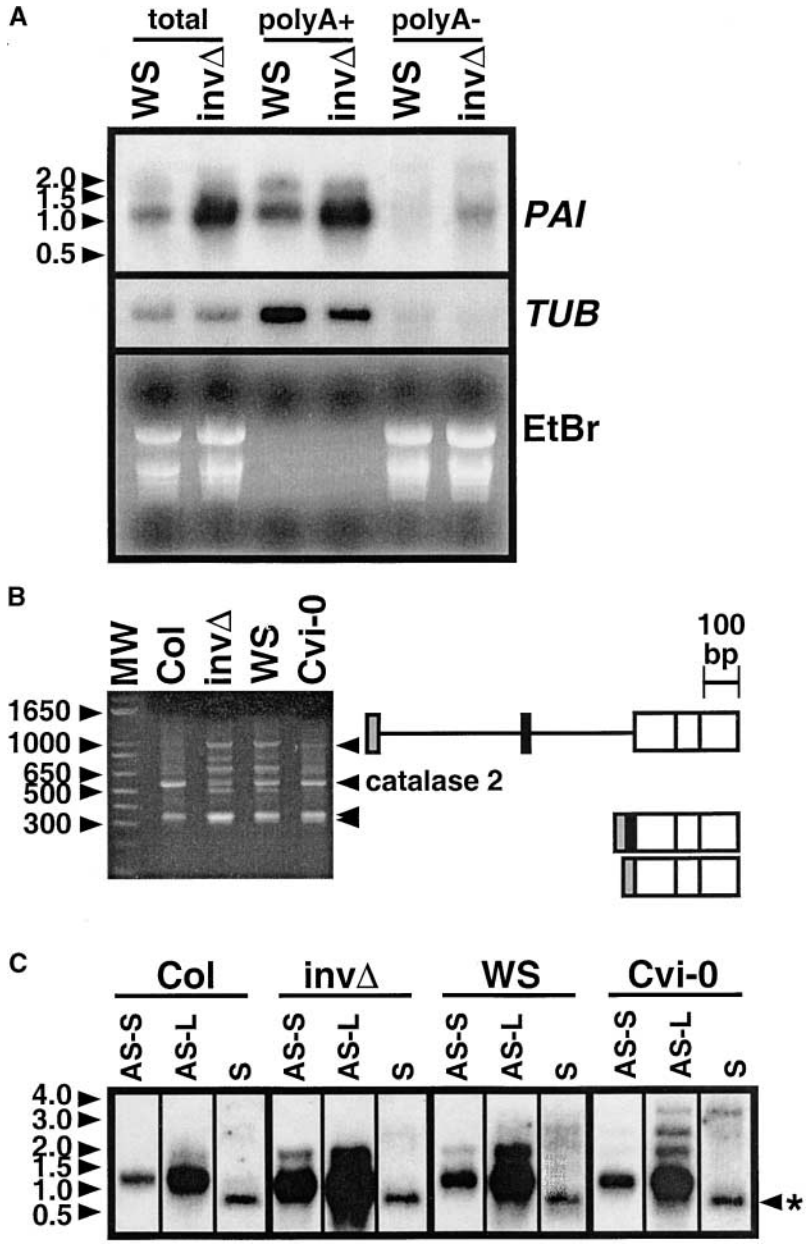

FIGURE 5.—*PAI* steady-state transcript analysis in the *invpai1-Δpai4* mutant. (A) Total RNA (total) from WS or *invpai1-Δpai4* (invΔ) was separated into polyadenylated [poly(A)+] and nonpolyadenylated [poly(A)−] fractions, gel blotted, and probed with a *PAI* probe. The blot was stripped and reprobed with β-tubulin (*TUB*) to control for loading of poly(A)+ species. The ethidium-bromide (EtBr)-stained gel is shown to control for loading of poly(A)− species. Molecular weight markers in kilonucleotides are indicated in the left margin. (B) Polyadenylated RNA prepared from the indicated strains was used as a template for RT-PCR amplification of *PAI1* transcript 5′ ends. An ethidium-bromide-stained gel of each RT-PCR reaction is shown, with molecular weight (MW) markers in leftmost lane, sizes of markers in base pairs in the left margin, and the identities of sequenced products indicated in the right margin. In the diagrams of *PAI* splice variants, the first upstream exon is represented by a shaded box, the second 26-nt upstream exon is represented by a solid box, and *PAI* exons are represented by open boxes. Lines indicate intron sequences that are retained in the longest splice variant. Note that in Cvi-0, the first upstream intron is 26 nt shorter than in WS due to two small deletions. (C) Duplicate blots of polyadenylated RNA prepared from the indicated strains electrophoresed in parallel on the same gel were hybridized with either a *PAI* antisense-strand (AS) or a *PAI* sense (S)-strand RNA probe. Both a short (AS-S) and a long (AS-L) exposure of the same antisense-strand-probed blot are shown, in the first case to resolve the predominant lower molecular weight species and in the second case to clearly show the less-abundant higher molecular weight species. A single exposure of the sense-strand-probed blot is shown. A nonspecific band detected by the sense-strand probe is marked with an asterisk in the right margin. This band serves as an internal loading control for the sense-strand blot. In addition, reprobing of both blots with a *TUB* probe indicated approximately equal loading of all samples (data not shown).

5′ RACE for parental WS (MELQUIST and BENDER 2003). As in WS, WS *invpai1-Δpai4* transcripts initiated at an upstream promoter derived from duplicated sequences of the *S15a* ribosomal protein gene that lies ~500 bp upstream of the *PAI* methylation boundary and ~850 bp upstream of the *PAI1* translational start codon (Table 1). The majority class of transcripts had 706 nt of upstream sequences removed by a single splicing event between the *S15a* first intron donor site and a cryptic acceptor site 96 nt upstream of the *PAI1* translational start codon. The other 5′ variant was spliced twice in the upstream region to retain a central 26-nt exon. In this variant, the *S15a* first intron donor site was joined to the *S15a* first intron acceptor site to remove 416 nt of intron sequences, and then a cryptic donor site was joined to the cryptic acceptor site at −96 nt to remove 264 nt of intron sequences.

Given that the processing of the *PAI1* upstream sequences involves cryptic splicing events and that the size of a transcript that failed to splice the upstream sequences would correspond to the 1900-nt species detected by RNA gel-blot analysis in WS *invpai1-Δpai4* and WS, we tested explicitly for the presence of this long 5′ splice variant using RT-PCR designed to optimize detection of the long lower-abundance species (Figure 5B). Specifically, we used a forward primer in the *S15a* first exon with a reverse primer in the *PAI* third exon, such that the putative long 5′ splice variant would yield a relatively short PCR product. These primers are predicted to yield 308- and 334-bp products from the two

TABLE 1

5′ RACE determination of *PAI* transcript ends for WS *invpai1-Δpai4*

Transcript	First upstream exon[a]	Second upstream exon[a]	First *PAI* exon[a]
PAI1	−869 to –803	−386 to –361	−96 to +10
PAI1	−869 to –803	NA	−96 to +10
PAI1	−869 to –803	NA	−96 to +10
PAI1	−868 to –803	NA	−96 to +10
PAI1	−868 to –803	NA	−96 to +10
PAI1	−868 to –803	NA	−96 to +10
PAI1	−866 to –803	NA	−96 to +10
PAI1	−850 to –803	NA	−96 to +10

NA, not applicable.

[a] Numbering is relative to the first base of the *PAI1* ATG translational start codon as +1 on genomic sequences of WS *invpai1-Δpai4*. The *PAI* sequence identity/methylation boundary ends at –355.

TABLE 2

3′ RACE determination of *PAI* transcript ends for WS *invpai1-Δpai4*

Transcript	Fourth *PAI* exon[a]	Fifth *PAI* exon[a]
PAI1	+938 to +1101	+1257 to +1506
PAI1	+938 to +1101	+1257 to +1544
PAI1	+938 to +1101	+1257 to +1544
PAI1	+938 to +1101	+1257 to +1546
PAI1	+938 to +1101	+1257 to +1546
PAI1	+938 to +1101	+1257 to +1548
PAI1	+938 to +1101	+1257 to +1548
PAI1	+938 to +1101	+1257 to +1615
PAI1	+938 to +1182	NA
PAI1	+938 to +1199	NA
PAI1	+938 to +1207	NA
PAI1	+938 to +1239	NA
PAI1	+938 to +1239	NA
PAI1	+938 to +1239	NA
PAI1	+938 to +1239	NA
PAI1	+938 to +1239	NA

NA, not applicable.

[a] Numbering is relative to the first base of the *PAI1* ATG translational start codon as +1 on genomic sequences of WS *invpai1-Δpai4*. Polyadenylation occurs after the last indicated base position. Note that the first eight transcripts have a normally spliced fourth *PAI* intron, whereas the remaining transcripts failed to remove this sequence. The *invpai1* translational stop codon is at +1471 to +1473.

upstream-spliced variants, a 1014-bp product from a species in which the upstream sequences are unspliced but the *PAI* first and second introns are correctly spliced, and a 1423-bp product from an unspliced template. As a control, we tested RNA prepared from the Col strain, which has a single unmethylated *PAI1* gene and lacks a detectable 1900-nt *PAI* species in RNA gel-blot analysis. We also tested RNA prepared from the Cape Verde Islands (Cvi-0) strain, which carries a methylated *PAI* inverted repeat in which the *PAI1* and *PAI4* genes lie 839 bp farther apart than in WS (Melquist *et al.* 1999). Like WS *invpai1-Δpai4* and WS, Cvi-0 produces a detectable 1900-nt transcript species, plus unique higher molecular weight species (Figure 5C). The RT-PCR analysis revealed that all four strains yielded short products corresponding to the upstream-spliced transcripts and a species at ~600 bp; however, only the three strains with methylated *PAI* inverted repeats yielded detectable amounts of an ~1000-bp product. Cloning and sequencing of RT-PCR products showed that the 600-bp product corresponded to a fortuitously amplified catalase 2 (At4g35090) transcript and that the 1000-bp product corresponded to a *PAI1* transcript that failed to splice either the *S15a* first intron or the cryptic intron in upstream sequences, but that correctly spliced the *PAI* first and second introns (Figure 5B). Other RT-PCR products were not analyzed further. The RT-PCR analysis thus shows that the 1900-nt species detected in strains with methylated *PAI* inverted repeats corresponds to a *PAI1* 5′ splice variant that fails to splice the upstream sequences between the *S15a* transcript start and more proximal *PAI* sequences. Presumably, there was a bias against recovering this long lower-abundance alternatively spliced species under the conditions we used for 5′ RACE analysis. We also tried to detect the 1900-nt species using either total or polyadenylated RNA gel-blot analysis with a probe corresponding to the upstream intron sequences, but this approach did not give a signal above background hybridization (data not shown), perhaps because the intron probe sequences contain a number of poly(T) tracts that can cross-hybridize to polyadenylated transcripts.

The WS *invpai1-Δpai4* rearrangement produces a novel prematurely polyadenylated transcript from the *invpai1* gene and suppresses readthrough transcription into palindromic Δ*pai4* sequences: 3′ RACE analysis of WS *invpai1-Δpai4* RNA revealed a novel type of *PAI1* transcript relative to parental WS (Table 2; Melquist and Bender 2003). In this species, the fourth intron failed to splice and the transcript was polyadenylated internal to the fourth intron sequences. The novel species is ~200 nt shorter than the full-length WS *PAI1* transcript and therefore likely corresponds to the lower molecular weight shift detected by gel-blot analysis of total mutant *PAI* transcripts (Figure 5A). A second type of *PAI1* 3′ end species was also detected: a full-length correctly spliced *PAI* transcript that was polyadenylated ~70 nt downstream of the translational stop codon (Table 2).

In previous analysis, we found that the longest *PAI* transcripts in WS and Cvi-0 hybridize to a *PAI* sense-strand-specific RNA probe, indicating that they are 3′ readthrough transcripts into palindromic *PAI4* sequences

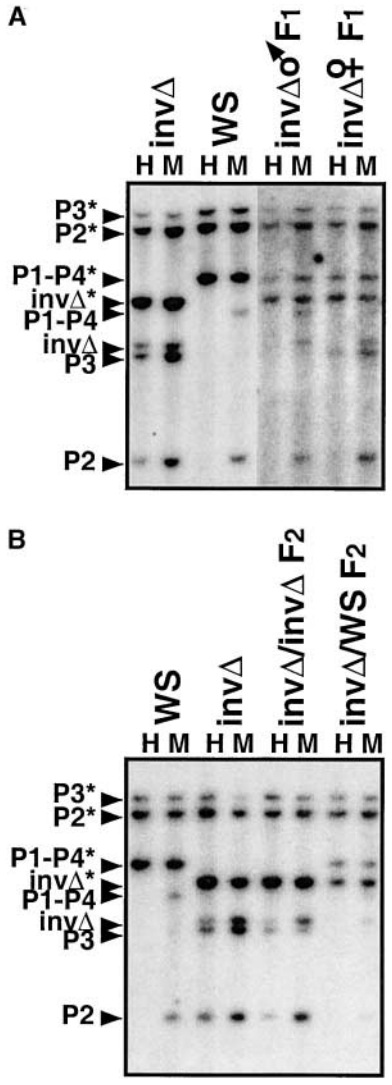

FIGURE 6.—The *invpai1-Δpai4* rearrangement can receive dense methylation conferred by the intact *PAI1–PAI4* locus in a heterozygote, but cannot maintain it once the *PAI1–PAI4* locus is removed by segregation. Genomic DNA samples prepared from the indicated strains were digested with the methylation-sensitive isoschizomers *Hpa*II (H) and *Msp*I (M) and probed with a *PAI* internal cDNA fragment. Methylated bands are denoted by asterisks. P1–P4 is *PAI1–PAI4*, invΔ is *invpai1-Δpai4*, P2 is *PAI2*, and P3 is *PAI3*. (A) Southern blot analysis of DNA from WS, WS *invpai1-Δpai4* (invΔ), or representative F_1 hybrids between these two strains made with the mutant as either the male or the female parent are shown. Similar results were obtained in three independent experiments. Note that the F_1 samples were run on the same blot as the parental samples, but a longer exposure of these samples is shown to adjust for loading differences. (B) Southern blot analysis of DNA from WS, WS *invpai1-Δpai4* (invΔ), or F_2 plants with the indicated genotypes at the *PAI1–PAI4* locus segregated from a WS × *invpai1-Δpai4* F_1 parent are shown.

(MELQUIST and BENDER 2003; Figure 5C). The shift upwards in the readthrough transcripts in Cvi-0 *vs.* WS reflects the different spacing between *PAI1* and *PAI4* in the two strains. To similarly determine whether the *invpai1-Δpai4* mutant transcript population included readthrough transcripts with palindromic *PAI4* sequences, which might be selected against during 3′ RACE analysis, we performed RNA gel-blot analysis of polyadenylated RNA with a sense-strand-specific RNA probe made on a *PAI* cDNA template (Figure 5C). The probe (MATERIALS AND METHODS) has 347 bp of homology with the *PAI4* sequences remaining in the *Δpai4* rearrangement. As a control for nonspecific hybridization, we also analyzed RNA prepared from the Col strain, which completely lacks *PAI4* inverted repeat sequences. In WS *invpai1-Δpai4* RNA, the sense-strand probe hybridized to a non*PAI* species only at ~800 nt, as also observed in other samples tested, including Col. Thus, the WS *invpai1-Δpai4* mutant does not produce detectable transcripts that contain palindromic *PAI4* material.

The WS *invpai1-Δpai4* locus can receive dense methylation triggered by the parental WS *PAI1–PAI4* locus, but cannot maintain dense methylation in its absence: In previous work we showed that the WS *PAI1–PAI4* locus is a potent trigger of dense CG and non-CG *PAI* methylation (LUFF *et al.* 1999). To determine whether the rearranged *invpai1-Δpai4* locus can be a target for this dense methylation, we generated *invpai1-Δpai4/PAI1–PAI4* F_1 heterozygotes by crossing WS *invpai1-Δpai4* to parental WS, either with the mutant as the male parent and WS as the female parent or vice versa. DNA prepared from cauline leaves of individual F_1 plants was analyzed by *Hpa*II/*Msp*I Southern blot relative to DNA prepared from the parental strains to assess changes in *PAI* methylation patterns. Regardless of whether the mutant was the male or the female parent in the cross, F_1 plants yielded DNA that showed increased resistance of the *invpai1-Δpai4* locus to cleavage by *Hpa*II relative to the *invpai1-Δpai4* homozygous parent DNA (Figure 6A). Similarly, the *PAI2* locus showed increased resistance to *Hpa*II cleavage. These patterns indicate that the partially demethylated loci inherited from the mutant parent rapidly acquire increased methylation during the F_1 generation.

Interestingly, the *PAI3* locus reproducibly showed a different pattern of remethylation. For F_1 plants generated with the mutant as the male parent, this locus displayed increased resistance to *Hpa*II cleavage, similarly to the other *PAI* loci. However, for F_1 plants generated with the mutant as the female parent, *PAI3* retained susceptibility to *Hpa*II cleavage similar to that observed in the *invpai1-Δpai4* mutant parent. This pattern suggests that the wild-type *PAI1–PAI4* locus is not an efficient trigger of *PAI3* methylation when inherited from the male parent in a cross and ties in with two previous observations. First, in experiments in which demethylated *PAI* genes are remethylated, *PAI3* is typically slower to remethylate than the other loci (LUFF *et al.* 1999; MALAGNAC *et al.* 2002). Second, in experiments where the WS *PAI1–PAI4* locus is used to trigger *de novo* methylation of the allelic unmethylated Col *PAI1* gene

in F_1 heterozygous plants, Col *PAI1* is not efficiently methylated when the WS *PAI1–PAI4* locus is inherited from the male parent in the cross (Luff *et al.* 1999). A possible explanation for the different potency of the *PAI1–PAI4* methylation trigger when inherited from the male parent *vs.* the female parent is suppressed transcription of the paternally inherited genome during the early cell divisions in the F_1 embryo (Vielle-Calzada *et al.* 2000); in this view, there would be a lag in the production of methylation-triggering RNA from the *PAI1–PAI4* locus when inherited from the male.

To further determine whether dense methylation can be maintained on the mutant *invpai1-Δpai4* locus after the *PAI1–PAI4* locus is segregated away, DNA prepared from pooled homozygous *invpai1-Δpai4/invpai1-Δpai4* F_2 progeny segregated from F_1 plants made with the mutant as the male parent was tested by *Hpa*II/*Msp*I Southern blot assay. This analysis showed that the segregated F_2 homozygotes reverted to the partially methylated pattern seen in the *invpai1-Δpai4* parent (Figure 6B). In contrast, DNA prepared from pooled *invpai1-Δpai4/PAI1–PAI4* heterozygous F_2 sibling plants maintained dense methylation on all three *PAI* loci at a slightly higher level than that observed in the F_1 parent, as indicated by increased resistance to *Msp*I cleavage. This pattern is consistent with a progressive increase in methylation upon inbreeding in the presence of the intact *PAI1–PAI4* locus. Similar results were obtained for methylation analysis of individual F_2 plants of each genotype (data not shown).

The WS *invpai1-Δpai4* locus cannot trigger *de novo* methylation of an unmethylated *PAI2* target gene: As another test for whether the rearranged *invpai1-Δpai4* locus could produce a *PAI* methylation signal, we asked whether the locus could trigger *de novo* methylation of an unmethylated *PAI2* gene. When the parental WS *PAI1–PAI4* locus is combined via genetic crosses with an unmethylated *PAI2* gene from the Col strain, the Col *PAI2* gene becomes methylated *de novo*, achieving a methylation density similar to that seen for WS *PAI2* by the F_3 or F_4 generation of inbreeding by self-pollination (Luff *et al.* 1999). Similarly, when a WS *pai1* missense mutant *pai1-PAI4* locus is combined with an unmethylated *PAI2* gene from the Landsberg *erecta* (Ler) strain, the Ler *PAI2* gene becomes methylated *de novo*, with a progressive increase in methylation upon inbreeding (Malagnac *et al.* 2002). In this second case, because of the defect in PAI1 enzyme, the progressive accumulation of *PAI2* methylation and transcriptional silencing can be monitored by visual inspection for a blue fluorescent PAI-deficient phenotype.

To test the *invpai1-Δpai4* locus for *de novo* methylation of *PAI2*, we crossed the mutant with wild-type Col and identified seven F_2 individuals that were homozygous for the *invpai1-Δpai4* locus and homozygous for the *PAI2* gene inherited from Col. None of these plants was fluorescent when newly segregated, and representative lines did not acquire a fluorescent phenotype diagnostic of *PAI2* silencing even after three generations of inbreeding by self-pollination to the F_5 generation. *Hpa*II/*Msp*I Southern blot analysis did not reveal any evidence of *de novo PAI2* methylation (Figure 7). A control cross between the WS *pai1* missense mutant and wild-type Col generated eight F_2 individuals that were homozygous for the *pai1-PAI4* mutant locus and homozygous for the *PAI2* gene inherited from Col. All eight plants were fluorescent when newly segregated. By the F_5 generation, representative inbred lines were phenotypcially similar to the WS *pai1* missense mutant (Figure 4) and displayed acquisition of *de novo PAI2* methylation as monitored by *Hpa*II/*Msp*I Southern blot analysis (Figure 7). These results indicate that the *invpai1-Δpai4* locus cannot trigger new methylation on a *PAI* target sequence, in contrast to the unrearranged *PAI1–PAI4* or *pai1-PAI4* loci.

It should be noted that in the F_5 inbred plants from the control cross between the WS *pai1* missense mutant and wild-type Col, the Col *PAI2* locus did not achieve fully dense methylation relative to the parental WS *PAI2* locus (Figure 7). A similar pattern was previously observed for the analogous experiment done with a cross between the WS *pai1* missense mutant and wild-type Ler (Malagnac *et al.* 2002). This resistance of *PAI2* to full methylation could reflect a selection bias against the

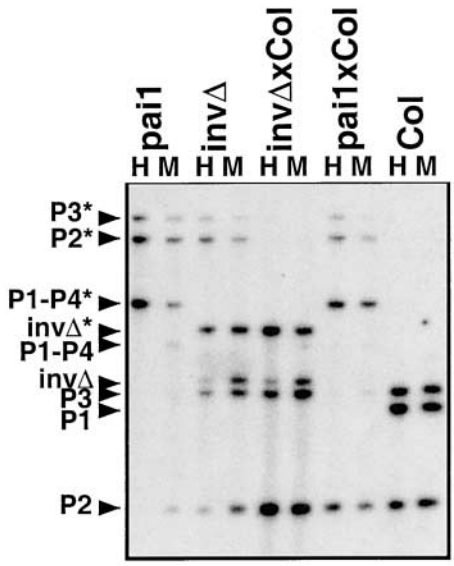

Figure 7.—The *invpai1-Δpai4* rearrangement is defective for *de novo* methylation of a *PAI2* methylation target inherited from Col. Genomic DNA samples prepared from WS *pai1* (pai1), WS *invpai1-Δpai4* (invΔ), F_5 generation tissue from a representative *invpai1-Δpai4/invpai1-Δpai4* Col *PAI2*/Col *PAI2* hybrid (invΔxCol), F_5 generation tissue from a representative WS *pai1-PAI4*/WS *pai1-PAI4* Col *PAI2*/Col *PAI2* hybrid (pai1xCol), and wild-type Col were digested with the methylation-sensitive isoschizomers *Hpa*II (H) and *Msp*I (M) and probed with a *PAI* internal cDNA fragment. Methylated bands are denoted by asterisks. P1–P4 is *PAI1–PAI4*, invΔ is *invpai1-Δpai4*, P2 is *PAI2*, and P3 is *PAI3*.

most strongly silenced PAI-deficient cells in the *pai1* mutant background.

DISCUSSION

Cytosine methylation is an important regulator of gene expression and genome stability in mammals and plants. A key question is how methylation is targeted to specific regions of the genome. In plants, sequences that produce dsRNA, including RNA viruses and transcribed inverted repeats, can be potent triggers for methylation of homologous genomic DNA sequences. For example, the transcribed endogenous *PAI1–PAI4* inverted repeat in the WS strain of Arabidopsis triggers dense methylation on *PAI* target sequences at unlinked positions in the genome (Luff *et al.* 1999). The ability of this inverted repeat to signal methylation is suppressed by silencing the upstream promoter that drives transcription through the locus, indicating an RNA-based DNA methylation mechanism (Melquist and Bender 2003). In this work, we characterized *invpai1-*Δ*pai4*, a mutant variant of WS with an intact promoter region but an internal deletion and rearrangement of the central sequences in the *PAI1–PAI4* locus (Figure 1). The mutant locus confers reduced density of methylation on WS *PAI* sequences without affecting methylation patterning elsewhere in the genome (Figure 2; Figure 3). Moreover, the mutant locus is defective in triggering *de novo* methylation on a previously unmethylated *PAI* target sequence (Figure 7). The rearrangement activates a novel premature polyadenylation site and suppresses readthrough into palindromic *PAI4* sequences (Table 2; Figure 5). These findings support the view that the *PAI* methylation defect in the rearrangement is conferred by reduced production of a dsRNA trigger for methylation due to an altered *PAI* polyadenylation profile.

In previous work, we found considerable natural variation in *PAI1–PAI4* inverted repeat structures among WS and other wild isolates of Arabidopsis (Melquist *et al.* 1999). Variations included different amounts of central sequences separating *PAI1* and *PAI4* and flanking duplications that could be explained as gene conversion events. The *invpai1-*Δ*pai4* rearrangement characterized here likely reflects the genetic instability of the locus. This rearrangement was isolated from an EMS-mutagenized seed population, but presumably occurred either as an indirect consequence of the mutagenesis or as an unrelated event. A possible explanation for the mutant structure is that an initial inversion event occurred by pairing and homologous recombination between *PAI1* and *PAI4*, followed by a deletion event extending from the recombination breakpoint in *PAI4* into the 3′ end of the gene (Figure 1). Interestingly, the deletion breakpoint at the 3′ end of the gene is 3 bp from the point where the palindromic and potentially aligned sequences between *PAI1* and *PAI4* end. It should be noted that an inversion between *PAI1* and *PAI4* without the accompanying central deletion and consequent reduced *PAI2* methylation and transcriptional silencing would likely yield a severe PAI-deficient mutant by introducing the inactivating fifth exon mutation into the expressed *PAI1* gene. In fact, such a mutant could potentially be an embryo-lethal tryptophan auxotroph. Thus, inversions at this locus might occur at a relatively high frequency, but not be recovered as viable plants.

The *PAI1* promoter region presents a complex arrangement of a proximal *PAI* promoter fused to an upstream *S15a*-derived promoter. In WS and WS *invpai1-*Δ*pai4*, in which the proximal *PAI* promoter sequences are methylated, only transcripts that initiate at the upstream promoter are detected by 5′ RACE (Melquist and Bender 2003; Table 1). In contrast, in Col, in which the proximal *PAI* promoter sequences are unmethylated, both upstream and proximal transcript starts are detected by 5′ RACE (Melquist and Bender 2003). Although we have not quantified the levels of transcripts initiating at each start site in Col, two lines of evidence suggest that the more proximal site is used preferentially when the promoter region is unmethylated. First, RNA gel-blot analysis shows that Col has an overall lower accumulation of *PAI* transcripts than does WS (Bender and Fink 1995; Figure 5C). Furthermore, on the basis of cDNA abundance and RACE analysis, approximately half of the accumulated Col *PAI* transcripts correspond to *PAI2* species initiating at the proximal *PAI* promoter (Melquist *et al.* 1999; Melquist and Bender 2003). These data can be accounted for by proposing that most Col *PAI1* transcripts initiate at the proximal *PAI* promoter even though this promoter is weaker than the distal *S15a* promoter. Second, the low-abundance 1900-nt splice variant that initiates at the upstream start site is not detected in Col either by RNA gel blot or by RT-PCR (Figure 5), suggesting lower expression from this upstream start site when the proximal start site is unmethylated and available than in strains like WS and Cvi-0 in which the proximal start site is methylated and silenced.

The *invpai1-*Δ*pai4* mutant illustrates that downstream sequences can influence the choice of splicing events in a plant transcript. In this mutant, all the sequences upstream of the *invpai1* fifth exon 9-bp deletion introduced by the inversion event are identical to those in the wild-type *PAI1* transcript, and yet uniquely in the mutant the fourth exon fails to splice at a high frequency (Table 2). Therefore, the fifth exon polymorphism and/or the novel rearranged sequences downstream of the *invpai1* stop codon contribute to the efficiency of fourth intron splicing. The retention of fourth intron sequences has the consequence of revealing a new polyadenylation region, resulting in the novel prematurely terminated transcripts. In those mutant transcripts that do splice out the fourth intron, polyadenylation occurs ~70 bp downstream of the translational stop codon

(Table 2) *vs.* ~130 bp downstream of the translational stop codon in wild-type WS (Melquist and Bender 2003).

Several interrelated mechanisms could contribute to the increased *PAI* transcript accumulation observed in the *invpai1-Δpai4* mutant *vs.* wild-type WS (Figure 5). First, the novel polyadenylation sites used in the *invpai1* transcripts (Table 2) could create messages that are more resistant than the wild-type *PAI1* species to RNA degradation. Second, the population of long readthrough transcripts detected in WS (Figure 5C) is presumably converted to shorter species in the mutant by activation of new polyadenylation signals, such that the heterogeneous long population is collapsed down into a more homogenous short population. This "collapsed" population is more likely than the original dsRNA population to accumulate to higher steady-state levels because it is no longer a substrate for dicing into siRNAs. Third, low levels of siRNAs produced by dicing readthrough transcripts could contribute to a low level of *PAI* transcript degradation via RNAi, and this effect would be blocked by readthrough suppression. Once dicer mutants that affect siRNA production are identified in Arabidopsis (Finnegan *et al.* 2003), we can use such mutants to determine the contribution of RNAi to *PAI* steady-state message levels in WS.

When transcription of WS *PAI1–PAI4* is suppressed by a transgene that triggers methylation and silencing of the *S15a*-derived promoter sequences upstream of *PAI1*, methylation on the singlet genes *PAI2* and *PAI3* is reduced, but methylation on the *PAI–PAI4* inverted repeat itself is not significantly affected (Melquist and Bender 2003; Figure 4). In contrast, the *invpai1-Δpai4* mutant displays reduced methylation on all *PAI* loci, including the inverted repeat (Figures 2–4). We previously proposed two models to account for the findings in the WS(S15aIR) promoter silenced strain: either the *PAI1–PAI4* inverted repeat maintains its own methylation via an RNA-independent mechanism such as a DNA structure signal or the *PAI1–PAI4* inverted repeat is more sensitive than the singlet *PAI* genes to RNA-based methylation signals (Melquist and Bender 2003). In the second model, the transcriptional activity of *PAI1–PAI4* and/or the unusual structure of the locus could contribute to its stronger susceptibility. The findings for the *invpai1-Δpai4* mutant can be accommodated by either model. In the first case, the alteration in the overall palindromic structure of the mutant locus could affect a methylation mechanism that responds to intrinsic structural cues. In the second case, stronger suppression of the RNA signal in the *invpai1-Δpai4* mutant *vs.* the promoter-silenced strain WS(S15aIR) could drop the RNA signal below the threshold needed to maintain methylation at the inverted repeat; furthermore, the altered structure of the locus could reduce its interactions with RNA.

The *PAI1–PAI4* locus provides general insights into how methylation patterns are established and maintained on endogenous methylation targets such as transposable elements. Our previous work showed that this single locus can generate an aberrant RNA signal sufficient for methylation of all related sequences in the genome, even though the aberrant RNA is not sufficient for effective RNAi (Luff *et al.* 1999; Melquist and Bender 2003). By analogy, for dispersed transposable elements, only one or a few transcribed elements could generate an RNA signal sufficient for methylating and silencing the entire group of related sequences. Here, our studies of the *invpai1-Δpai4* mutant show that methylation imprints established by an RNA signal are difficult to completely remove, even when the methylation trigger locus is rearranged to severely suppress the production of the signal. In particular, in contrast to a mutant with complete deletion of the *PAI1–PAI4* locus (Bender and Fink 1995), the *invpai1-Δpai4* mutant does not yield *PAI*-demethylated progeny at a detectable frequency. The stability of the residual methylation in *invpai1-Δpai4* therefore implies that low levels of an RNA trigger for DNA methylation persist in this mutant. In fact, the levels of the RNA trigger are likely to be extremely low for the following reasons: *PAI* dsRNA cannot be detected in the mutant by RNA gel blot (Figure 5C); the increase in steady-state levels of *PAI* transcripts suggests a loss of *PAI*-directed RNAi (see above); the loss of non-CG methylation patterning (Figures 2 and 3) diagnostic of RNA-directed DNA methylation (Pélissier *et al.* 1999) suggests an RNA signal deficiency; and the mutant locus is not able to produce a strong enough signal for *de novo* methylation of an unmethylated *PAI2* target (Figure 7). The *invpai1-Δpai4* mutant thus illustrates that when superimposed on systems that maintain preexisting methylation imprints, even a severely impaired RNA signal for DNA methylation can promote stable methylation patterning.

This work was supported by National Institutes of Health grant GM-61148 and March of Dimes grant FY00-655 to J.B. and by National Institutes of Environmental Health Sciences Training Grant ES 07141 to S.M.

LITERATURE CITED

Bartee, L., and J. Bender, 2001 Two *Arabidopsis* methylation-deficiency mutations confer only partial effects on a methylated endogenous gene family. Nucleic Acids Res. **29:** 2127–2134.

Bartee, L., F. Malagnac and J. Bender, 2001 *Arabidopsis cmt3* chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes Dev. **15:** 1753–1758.

Béclin, C., S. Boutet, P. Waterhouse and H. Vaucheret, 2002 A branched pathway for transgene-induced RNA silencing in plants. Curr. Biol. **12:** 684–688.

Bender, J., and G. R. Fink, 1995 Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. Cell **83:** 725–734.

Bird, A., 2002 DNA methylation patterns and epigenetic memory. Genes Dev. **16:** 6–21.

Dalmay, T., A. Hamilton, S. Rudd, S. Angell and D. C. Baulcombe, 2000 An RNA-dependent RNA polymerase gene in *Arabidopsis*

is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell **101:** 543–553.

Dalmay, T., R. Horsefield, T. H. Braunstein and D. C. Baulcombe, 2001 *SDE3* encodes an RNA helicase required for post-transcriptional gene silencing in *Arabidopsis.* EMBO J. **20:** 2069–2077.

Fagard, M., S. Boutet, J.-B. Morel, C. Bellini and H. Vaucheret, 2000 AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptonal gene silencing in plants, quelling in fungi, and RNA interference in animals. Proc. Natl. Acad. Sci. USA **97:** 11650–11654.

Finnegan, E. J., R. Margis and P. M. Waterhouse, 2003 Posttranscriptional gene silencing is not compromised in the *Arabidopsis CARPEL FACTORY* (*DICER-LIKE1*) mutant, a homolog of Dicer-1 from *Drosophila.* Curr. Biol. **13:** 236–240.

Haughn, G. W., and C. Somerville, 1986 Sulfonylurea-resistant mutants in *Arabidopsis thaliana.* Mol. Gen. Genet. **204:** 430–434.

Jeddeloh, J. A., J. Bender and E. J. Richards, 1998 The DNA methylation locus *DDM1* is required for maintenance of gene silencing in *Arabidopsis.* Genes Dev. **12:** 1714–1725.

Kato, M., A. Miura, J. Bender, S. E. Jacobsen and T. Kakutani, 2003 Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis.* Curr. Biol. **13:** 421–426.

Luff, B., L. Pawlowski and J. Bender, 1999 An inverted repeat triggers cytosine methylation of identical sequences in *Arabidopsis.* Mol. Cell **3:** 505–511.

Malagnac, F., L. Bartee and J. Bender, 2002 An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J. **21:** 6842–6852.

Matzke, M., A. J. M. Matzke and J. M. Kooter, 2001 RNA: guiding gene silencing. Science **293:** 1080–1083.

Melquist, S., and J. Bender, 2003 Transcription from an upstream promoter controls methylation signaling from an inverted repeat of endogenous genes in *Arabidopsis.* Genes Dev. **17:** 2036–2047.

Melquist, S., B. Luff and J. Bender, 1999 Arabidopsis *PAI* gene arrangements, cytosine methylation and expression. Genetics **153:** 401–413.

Mette, M. F., W. Aufsatz, J. van der Winden, M. A. Matzke and A. J. M. Matzke, 2000 Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. **19:** 5194–5201.

Miura, A., S. Yonebayashi, K. Watanabe, T. Toyama, H. Shimada *et al.*, 2001 Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis.* Nature **411:** 212–214.

Mourrain, P., C. Béclin, T. Elmayan, F. Feuerbach, C. Godon *et al.*, 2000 *Arabidopsis SGS2* and *SGS3* genes are required for posttranscriptional gene silencing and natural virus resistance. Cell **101:** 533–542.

Pélissier, T., S. Thalmeir, D. Kempe, H.-L. Sänger and M. Wassenegger, 1999 Heavy *de novo* methylation at symmetrical and nonsymmetrical sites is a hallmark of RNA-directed DNA methylation. Nucleic Acids Res. **27:** 1625–1634.

Schramke, V., and R. Allshire, 2003 Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. Science **301:** 1069–1074.

Sijen, T., I. Vijn, A. Rebocho, R. van Blokland, D. Roelofs *et al.*, 2001 Transcriptional and posttranscriptional gene silencing are mechanistically related. Curr. Biol. **11:** 436–440.

Singer, T., C. Yordan and R. A. Martienssen, 2001 Robertson's *Mutator* transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA Methylation* (*DDM1*). Genes Dev. **15:** 591–602.

Vaistij, F. E., L. Jones and D. C. Baulcombe, 2002 Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. Plant Cell **14:** 857–867.

Vielle-Calzada, J.-P., R. Baskar and U. Grossniklaus, 2000 Delayed activation of the paternal genome during seed development. Nature **404:** 91–94.

Volpe, T. A., C. Kidner, I. M. Hall, G. Teng, S. I. S. Grewal *et al.*, 2002 Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science **297:** 1833–1837.

Communicating editor: B. Bartel