

Chromodomains direct integration of retrotransposons to heterochromatin

Xiang Gao,^{1,3} Yi Hou,¹ Hirotaka Ebina,² Henry L. Levin,² and Daniel F. Voytas^{1,4}

¹Department of Genetics, Development & Cell Biology, Iowa State University, Ames, Iowa 50011, USA; ²Section on Eukaryotic Transposable Elements, Laboratory of Gene Regulation and Development, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

The enrichment of mobile genetic elements in heterochromatin may be due, in part, to targeted integration. The chromoviruses are Ty3/gypsy retrotransposons with chromodomains at their integrase C termini. Chromodomains are logical determinants for targeting to heterochromatin, because the chromodomain of heterochromatin protein 1 (HPI) typically recognizes histone H3 K9 methylation, an epigenetic mark characteristic of heterochromatin. We describe three groups of chromoviruses based on amino acid sequence relationships of their integrase C termini. Genome sequence analysis indicates that representative chromoviruses from each group are enriched in gene-poor regions of the genome relative to other retrotransposons, and when fused to fluorescent marker proteins, the chromodomains target proteins to specific subnuclear foci coincident with heterochromatin. The chromodomain of the fungal element, MAGGY, interacts with histone H3 dimethyl- and trimethyl-K9, and when the MAGGY chromodomain is fused to integrase of the *Schizosaccharomyces pombe* Tfl retrotransposon, new Tfl insertions are directed to sites of H3 K9 methylation. Repetitive sequences such as transposable elements trigger the RNAi pathway resulting in their epigenetic modification. Our results suggest a dynamic interplay between retrotransposons and heterochromatin, wherein mobile elements recognize heterochromatin at the time of integration and then perpetuate the heterochromatic mark by triggering epigenetic modification.

[Supplemental material is available online at www.genome.org.]

Mobile elements are a major component of the genetic material, and to protect genomes from the deleterious consequences of transposition, cells have evolved strategies, such as RNA interference (RNAi), to rein in mobile element amplification (Bosher and Labouesse 2000; Vastenhouw and Plasterk 2004). Such surveillance systems lead to epigenetic inactivation of mobile elements through post-transcriptional silencing (Sijen and Plasterk 2003), histone and DNA modifications (Lippman and Martienssen 2004; Martienssen et al. 2004), as well as the recently discovered piwi/Aubergine germline silencing pathway (Girard et al. 2006; Grivna et al. 2006; Vagin et al. 2006). Mobile element insertions that decorate eukaryotic genomes, therefore, frequently define unique chromatin domains. Whereas the genetic consequences of transposition in terms of mutation and genome rearrangement have long been recognized, the biological consequences of their epigenetic marks, which include effects on gene expression and the formation of heterochromatin, are only beginning to be appreciated (Slotkin and Martienssen 2007).

Underlying the genetic and epigenetic impact of mobile elements is integration site choice. Although forces such as selection or recombination contribute to the nonrandom distribution of mobile elements in eukaryotic genomes, an increasing number of studies indicate that many mobile elements target integration to specific chromosomal sites (Bushman 2003). For some elements, target sites are determined by recognizing specific DNA sequences, whereas for others, including several retrotransposons and retroviruses, chromatin impacts target site choice. The Ty1

and Ty3 retrotransposons of *Saccharomyces cerevisiae*, for example, integrate near sites of RNA polymerase III (pol III) transcription by recognizing pol III transcription complexes or chromatin states associated with pol III transcription (Yieh et al. 2000, 2002; Bachman et al. 2005; Mou et al. 2006). The Tfl retrotransposon of *Schizosaccharomyces pombe* recognizes certain RNA polymerase II (pol II) promoters (Singleton and Levin 2002; Bowen et al. 2003). Similarly, the retroviruses, which are closely related to the retrotransposons, show target site biases. Murine leukemia virus (MLV) shows a preference for promoter regions (Wu et al. 2003), and human immunodeficiency virus (HIV) integrates preferentially into actively transcribed genes at sites with transcription-associated histone modifications (Schroder et al. 2002; Wang et al. 2007).

In contrast to targeting to specific transcription complexes or chromatin states associated with transcription, the Ty5 retrotransposon of *S. cerevisiae* integrates preferentially into transcriptionally inactive regions of the genome, namely heterochromatin found at the telomeres and silent mating loci (Zou et al. 1996). Targeting to heterochromatin requires a six-amino-acid motif (the targeting domain) at the C terminus of Ty5 integrase that interacts with the heterochromatin protein Sir4 (Gai and Voytas 1998; Xie et al. 2001). This interaction tethers the Ty5 integration complex to target sites resulting in the observed target site biases. Although targeting determinants encoded by other retrotransposons remain to be identified, increasingly, it appears that integrase is the mediator of target site choice. Recent experiments in which domains of HIV and MLV integrase were swapped demonstrated that integrase is responsible for the different patterns of integration observed at genes for these two viruses (Lewinski et al. 2006), and in a separate study, sequence differences among various retroviral integrases were found to correlate with global and local integration patterns (Derse et al.

³Present address: Department of Biology, 1001 East Third Street, Indiana University, Bloomington, Indiana 47405, USA.

⁴Corresponding author.

E-mail Voytas@iastate.edu; fax (515) 294-7155.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.7146408>.

2007). Analogous to the Ty5/Sir4 relationship, an interaction between HIV integrase and the transcription factor LEDGF/p75 underlies HIV's preference for actively transcribed genes (Ciuffi et al. 2005; Llano et al. 2006; Shun et al. 2007).

The targeting determinants for Ty5 lie within the integrase C terminus, a region of the protein that is highly divergent among retrotransposons both in sequence composition and size. One exception occurs in a specific lineage of retrotransposons referred to as the chromoviruses (family *Metaviridae*) (Marin and Llorens 2000; Gorinsek et al. 2004, 2005). The integrase C termini of some chromoviruses have a chromodomain (CHD)—an ~40- to 50-amino-acid sequence motif that can interact with diverse targets, including proteins, RNA, and DNA (Brehm et al. 2004). Among the best-characterized CHD partners are methylated histone residues. The heterochromatin protein 1 (HP1) (chromobox homolog, CBX) CHD typically interacts with histone H3 methyl-K9 (Jacobs and Khorasanizadeh 2002; Nielsen et al. 2002), and the polycomb CHD interacts with H3 methyl-K27 (Fischle et al. 2003; Min et al. 2003), both of which are marks typically found in heterochromatin. In humans, the tandem CHDs of chromo-helicase/ATPase DNA-binding protein 1 (CHD1) act cooperatively to specifically bind H3 methyl-K4 (Flanagan et al. 2005), a mark characteristic of euchromatin. In each of these proteins, the CHD targets the protein to sites bearing the specific epigenetic marks. Consequently, when CHDs were identified in retrotransposon integrases, they were hypothesized to play a role in target specificity (Koonin et al. 1995; Malik and Eickbush 1999). In this study, we provide evidence that some retrotransposon CHDs do, in fact, recognize histone modifications characteristic of heterochromatin and that retrotransposon chromodomains can tether integration complexes to heterochromatin resulting in targeted integration.

Results and Discussion

Diversity of retrotransposon-encoded chromodomains

Amino acid sequences of integrase C termini were analyzed from chromoviruses present in the genomes of diverse plants, animals, and fungi. Two distinct groups of CHDs were identified (Supplemental Fig. S1). Group I CHDs display high sequence similarity to cellular (i.e., nontransposable element) CHDs and are found in diverse eukaryotes. In the HP1 CHD and its homologs, three conserved aromatic residues recognize methylated lysines on histone H3 (Jacobs and Khorasanizadeh 2002; Nielsen et al. 2002); all three residues are conserved in group I CHDs. The group II CHDs are found only in plant retrotransposons. Both sequence similarity and predicted secondary structures indicate that they are homologs of the classical, group I CHDs; however, the group II motif lacks the first conserved aromatic residue and usually the third, suggesting that they interact with different partners. A very different motif was identified at the corresponding position of the CHD among plant centromere-specific retrotransposons (CR elements) (Jiang et al. 2003). Although this group of retrotransposons has been classified as chromoviruses, the motif could not be convincingly aligned to CHDs, and results of secondary structure predictions suggest that this well-conserved motif is structurally distinct (data not shown). In the studies described below, the MAGGY retrotransposon from *Magnaportha grisea* represents elements with group I CHDs; group II elements are represented by both a Tma-like element from *Arabidopsis thaliana* (referred to hereafter as Tma) and Os_{rn} 377-208, a RIRE3-like element from

Oryza sativa (referred to hereafter as Os); the *Zea mays* CRM and *O. sativa* CRR2 are representative elements with the CR motif.

Retrotransposons with chromodomains are associated with repetitive regions of host genomes

If CHDs are targeting determinants, chromoviruses would be predicted to have biased chromosomal distributions. In species such as *A. thaliana* where blocks of euchromatin and heterochromatin are distinct and well characterized, Ty3/*gypsy* retrotransposons (of which the chromoviruses are a member) show a significant association with heterochromatin compared with Ty1/*cop* retrotransposons (Pereira 2004; Peterson-Burch et al. 2004). This holds true for insertions of our representative group II chromovirus, Tma, which are clustered in pericentromeric regions (Supplemental Fig. S2A).

The clustering of transposable elements on all seven chromosomes of *M. grisea*—the host of the group I element, MAGGY—has previously been noted (Thon et al. 2004; Dean et al. 2005). Unfortunately, the distribution of epigenetic marks that define heterochromatin, such as histone and DNA methylation, has not been characterized in this species. In general, heterochromatin is gene poor and enriched in repeats and transposable elements (Grewal and Elgin 2007; Grewal and Jia 2007). We therefore quantitatively assessed the frequency with which MAGGY insertions are associated with genes (a proxy for euchromatin) or transposable elements (a proxy for heterochromatin). Analysis of 20-kb windows centered on MAGGY insertions revealed that, on average, MAGGY insertions are associated with 2.40 (SE 0.18) genes and 0.56 (SE 0.10) transposable elements (Fig. 1A), whereas random windows have 3.88 (SE 0.24) genes and 0.17 (SE 0.05) transposable elements ($P < 0.00001$ for both genes and transposable elements). Windows of 10 kb and 40 kb were also tested, and biases in distribution were also found to be significant (data not shown). MAGGY insertions, therefore, appear to be clustered with other transposable elements and are located in relatively gene poor regions.

Heterochromatin in rice includes the pericentromere as well as islands along the chromosomes (Houben et al. 2003). In contrast to Tma of *A. thaliana*, insertions of Os—the other representative group II element—are dispersed on the rice chromosome arms (Fig. 1B). The association of Os with genes and other transposable elements was evaluated in 40-kb windows centered on Os insertions (40-kb windows were chosen because of the larger intergenic regions in plants relative to fungi; however similar results were also obtained with 20-kb and 80-kb windows, data not shown). In regions flanking the 917 Os insertions in the rice genome, there were on average 2.96 (SE 0.07) genes and 2.93 (SE 0.06) transposable elements (Fig. 1C). In contrast, random 40-kb windows averaged 5.03 (SE 0.09) genes and 1.20 (SE 0.05) transposable elements ($P < 0.00001$ for both genes and transposable elements). As a control, we evaluated gene and transposon associations for Tos17—a Ty1/*cop* retrotransposon—which prefers to integrate near genes (Miyao et al. 2003). Consistent with this integration specificity, Tos17 insertions were associated on average with 4.90 (SE 0.25) genes and 1.10 (SE 0.12) transposable elements ($P < 0.00001$ for both transposable elements and genes), and therefore they occupy a different genome context from the Os chromovirus that is more similar to random sites ($P < 0.33$ for both transposable elements and genes). Comparisons of long terminal repeat (LTR) DNA sequence divergence indicate that insertions comprising the Tos17 and Os families are

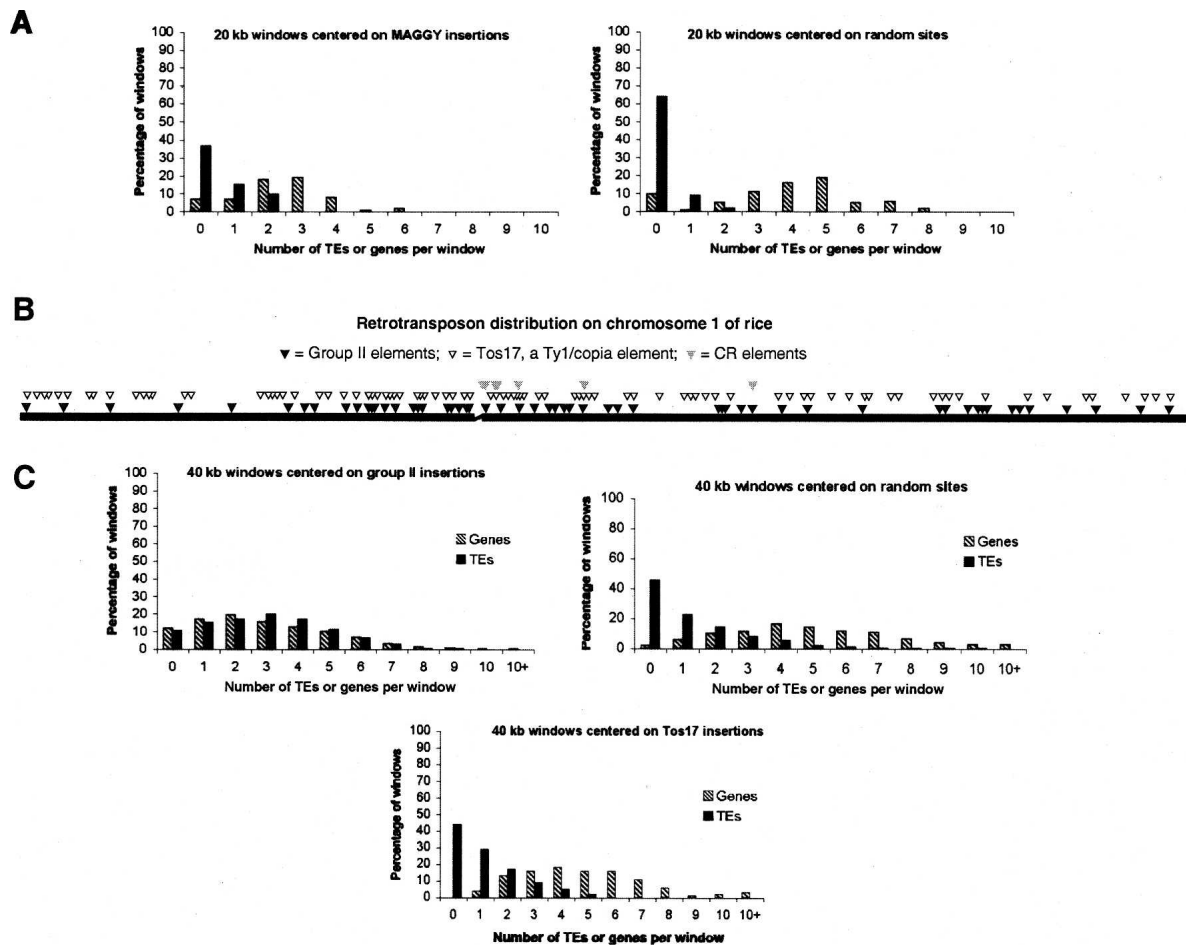


Figure 1. Retrotransposons with CHDs and CR motifs are preferentially located in gene-poor, transposon-rich regions of their host genomes. (A) The genomic distribution of *M. grisea* MAGGY (group I) insertions. For each of the 75 MAGGY insertions in the completed genome sequence (Dean et al. 2005), 20-kb windows at the site of insertion (10 kb upstream and downstream) were surveyed for genes and transposable elements. Seventy-five randomly selected sites were similarly surveyed as a control. (B) The distribution of group II, CR, and Tos17 insertions on chromosome 1 of *O. sativa*. Insertions of group II chromoviruses and Tos17 are distributed along the chromosome arms; however, as indicated in C below, these two element families occupy different genomic contexts. Only a few CR elements are found in the assembled rice genome sequence and these are near the pericentromeric regions (gap in chromosome). Many other CR elements have been revealed by FISH analysis or are present in the recent sequence of the centromere from chromosome 8 (Cheng et al. 2002; Nagaki et al. 2004). (C) Os (class II) insertions are in transposon-rich, gene-poor regions scattered across the 12 rice chromosomes. For each Os insertion site, a 40-kb window (20 kb upstream and 20 kb downstream) was surveyed for genes and transposable elements. Windows containing Os insertions have more transposable elements and fewer genes than random windows or windows centered on Tos17 insertions, a retrotransposon that integrates preferentially into gene-rich regions of the rice genome (Miyao et al. 2003).

comparable in age and very young (Supplemental Fig. 2B). These insertions, therefore, are less likely to be fixed, suggesting that target site specificity is the major underlying factor in their different genomic distributions. Although forces such as recombination and selection could contribute, they would have to act very differently on these two comparably aged families to yield such very different genomic distributions.

The CR retrotransposons got their name because of their association with centromeric repeats, and therefore they show a clear heterochromatin association (Miller et al. 1998; Cheng et al. 2002). Although there are only a few CR insertions present in the rice genome sequence, most are associated with pericentromeric regions (Fig. 1B). Many additional insertions reside in the centromeric heterochromatin, which is evidenced by the recent sequencing of a rice centromere and FISH analyses in a variety of plant species (Jiang et al. 1996; Dong et al. 1998; Cheng et al. 2002; Nagaki et al. 2004).

Retroelement chromodomains direct proteins to heterochromatin

Our bioinformatic analyses demonstrated that all five representative retrotransposons from the three groups of chromoviruses are associated with repeat-rich regions of their respective host genomes. Since some CHDs target cellular proteins to heterochromatin through interactions with histone H3 methyl-K9 (Jacobs and Khorasanizadeh 2002; Nielsen et al. 2002), we reasoned that retrotransposon CHDs may serve a similar function for integrate. To more directly test this hypothesis, we transiently expressed fusion proteins generated between yellow fluorescent protein (YFP) and retrotransposon CHDs in *A. thaliana* suspension cells and leaf protoplasts. The YFP-CHD fusions were visualized by confocal microscopy, and both types of CHDs and the CR motif showed a punctate nuclear distribution (Fig. 2A). In contrast, YFP fusions to the C terminus of integrase from Tnt1—a

Ty1/ *copia* retrotransposon associated with genes (Le et al. 2007)—were distributed throughout the nucleus. A cyan fluorescent protein (CFP) fusion to TFL2 (HP1-like homolog in *A. thaliana*) colocalized with the retrotransposon YFP-CHD fusions. When transiently expressed, TFL2 localizes to the chromocenters, the primary domain of heterochromatin in *A. thaliana* (Fransz et al. 2002; Zemach et al. 2006). We confirmed this association by showing colocalization between TFL2 and a YFP fusion to CENP-C—a kinetochore protein (Supplemental Fig. S3A).

Group I and group II CHDs have conserved aromatic amino acid residues, which in the HP1 CHD interact with methylated H3 K9 (Bannister et al. 2001; Lachner et al. 2001) (see Supplemental Fig. S1). These conserved residues were mutated in the MAGGY, Tma, and Os CHDs, and the mutations abolished the localization of all three CHDs to heterochromatin (Fig. 2B). In summary, localization experiments with representative retrotransposon CHDs and their mutants are consistent with a role for these motifs in recognizing heterochromatin and directing integration complexes to these sites.

The MAGGY chromodomain interacts with histone H3 methyl-K9

To test directly whether the MAGGY CHD interacts with histones, pull-down experiments were performed using a His₆-tagged MAGGY CHD and purified histones from calf thymus (Supplemental Fig. S4A). MAGGY specifically pulled down H3. To test the specificity of the H3 interaction, pull-downs were conducted with a calf thymus histone extract and various glutathione *S*-transferase (GST)-tagged CHDs. As indicated on stained polyacrylamide gels, the MAGGY CHD pulled down from the mixture a protein of the same size as histone H3 (Fig. 3A) that was subsequently confirmed to be H3 by mass spectrometry (data not shown). Neither GST alone nor a MAGGY CHD with mutations in conserved aromatic amino acids pulled down histone H3. Moreover, no proteins were pulled down by GST-Tma and GST-Os fusions, suggesting that the partners for these CHDs are plant specific.

To determine whether the MAGGY CHD interacts with a specific H3 modification, His-tagged wild-type and mutant MAGGY CHDs were used in pull-down assays with a series of biotinylated histone H3 modified peptides, including H3 dimethyl-R2, H3 amino acids (aa) 18–36, H3 methyl-K27, H3 dimethyl-K27, H3 trimethyl-K27, H3 aa 28–41, H3 methyl-K36, H3 dimethyl-K36, H3 trimethyl-K36, H3 aa 1–21, H3 methyl-K4, H3 methyl-R8, H3 aa 9–29, H3 trimethyl-K14, H3 methyl-R17, H3 trimethyl-K18, and H3 dimethyl-K9 (M. Bedford, pers. comm.; data not shown). The MAGGY CHD specifically pulled down the dimethyl-K9 peptide but neither the methyl-K4 peptide (Fig. 3B) nor any of the other peptides tested (data not shown). Additional experiments were performed to test for interactions with both di- and trimethyl-K9 peptides, and the MAGGY CHD interacted with both modified forms (Supplemental Fig. S4B) with some apparent preference for the trimethylated form (data not shown). In summary, the pull-down experiments indicate that like its HP1 homolog, the MAGGY CHD interacts with histone H3 and has strong specificity for H3 di- and trimethyl-K9.

H3 methyl-K9 is found in heterochromatin in most eukaryotes, including fungi, plants, and animals (Jacobs et al. 2001; Peters et al. 2001; Tamaru and Selker 2001; Gendrel et al. 2002), and the interaction between the MAGGY CHD and H3 methyl-K9 suggests that it is responsible for the observed subnuclear localization of the MAGGY CHD fusion protein. The role of H3

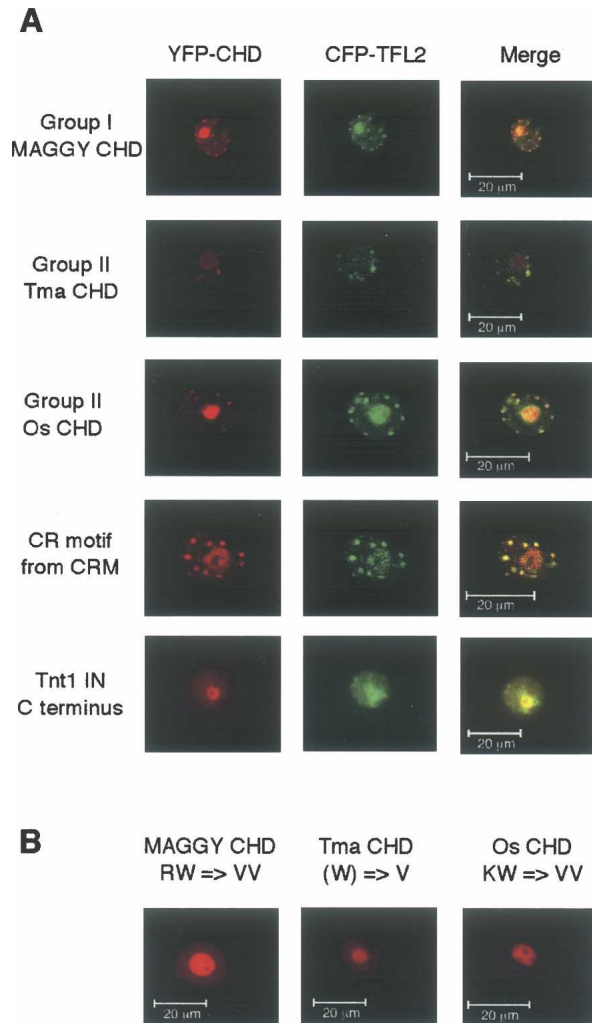


Figure 2. Subnuclear localization of retrotransposon CHDs and the CR motif. (A) YFP fusion proteins expressing retrotransposon CHDs or the CR motif localize to sites of heterochromatin in *A. thaliana* cells. Constructs expressing fusion proteins between YFP and either group I or group II CHDs or the CR motif were transformed into *A. thaliana* suspension cell protoplasts and visualized by confocal microscopy. The fusion proteins formed punctate foci in the nucleus and were enriched in the nucleolus. Localization was coincident with a fusion between CFP and TFL2—the *A. thaliana* HP1 homolog—but not the negative control, namely the Tnt1 IN C terminus. (B) Mutations in a conserved aromatic amino acid in group I and group II CHDs that is predicted to interact with the methyl group on H3 K9 abrogate subnuclear localization of YFP-CHD fusion proteins. Residues modified in the representative group I and group II CHDs are highlighted in red in Supplemental Fig. S1. In each case, residues were mutated to valine.

methyl-K9 in heterochromatin biology has been studied extensively in *S. pombe* (Grewal and Rice 2004), and in wild-type *S. pombe* strains, we observed that an enhanced green fluorescent protein (EGFP)-MAGGY CHD fusion localized to 1–3 subnuclear foci in ~70% of the nuclei examined (Fig. 3C, top; data not shown). This localization was similar to that observed for Chp2, a protein with a CHD that localizes to heterochromatin (Sadaie et al. 2004). The two proteins colocalize in both *swi6-115* and wild-type cells (Fig. 3C, bottom; data not shown). The *swi6-115* strain has a mutation in the HP1 homolog of *S. pombe* and yet normal H3 methyl-K9 in heterochromatic regions (Nakayama et al.

sequences. Integration and recombination frequencies can be calculated by comparing frequencies of G418 resistance conferred by a wild-type element and an integrase frameshift mutant for which Tf1 cDNA can only enter the genome through recombination. The integration frequency, therefore, can be calculated by subtracting the number of G418 resistant colonies observed in the frameshift mutant from the number observed in wild type. Wild-type Tf1 retroelements typically show about 20-fold more integration than recombination events (Fig. 4A).

We anticipated that expression of the *neo* gene within Tf1 might be compromised if Tf1 variants with a MAGGY CHD integrated into heterochromatin. We therefore used for our experiments a *swi6-115* host strain, which has a mutation in HP1 (W269R). Swi6 protein levels are dramatically reduced in this strain, and marker genes inserted into centromeric repeats are expressed (Nakagawa et al. 2002). However, and importantly for our experiments, histone modifications that characterize heterochromatin are not altered in *swi6-115* strains (Nakayama et al. 2001), although they tend to be restricted to sites where hetero-

chromatin nucleates, such as CenH at the mating loci and the dg-dh repeats in the outer centromere (Noma et al. 2004). As described above, the subnuclear localization of a MAGGY CHD-EGFP fusion is indistinguishable between wild type and a *swi6-115* background, and the MAGGY CHD colocalizes with the heterochromatin marker Chp2 in *swi6-115* cells (Fig. 3C).

In initial experiments, we replaced the Tf1 CHD with the CHD from MAGGY; however, the modified element was incapable of transposition (data not shown). The Tf1 integrase CHD is located at the very C terminus of the Tf1 polypeptide, and as an alternative strategy, we simply fused the MAGGY CHD to the integrase C terminus, thereby creating a Tf1 element with tandem CHDs (Tf1-Mac). A variant of Tf1-Mac was also created that carries the integrase frameshift mutation (Tf1-Mac-fs). Although the addition of the MAGGY CHD reduced transposition frequencies sevenfold compared with wild type, Tf1-Mac still produced fourfold more integration than recombination events in a *swi6-115* strain (Fig. 4A).

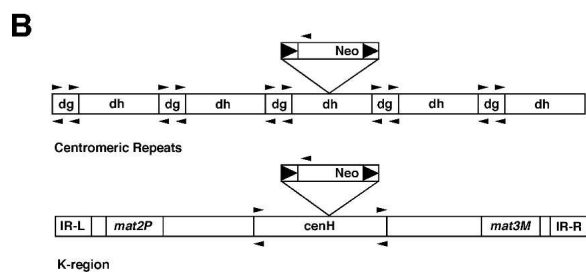
To determine whether Tf1-Mac integrates into heterochromatin, Tf1-Mac transposition was induced, and DNA was prepared from the resulting G418 resistant colonies. PCR experiments were then carried out using a Tf1 primer and either (1) one of four primers at the two ends of the centromeric dg repeats or (2) one of four primers at the ends of K-region at the mating locus (Fig. 4B). For wild-type Tf1, 2400 G418 colonies were screened. On the basis of the frequency of integration relative to recombination (Fig. 4A), this approximates 2267 integration events. No insertions of wild-type Tf1 elements were detected into either the outer centromere or the mating locus. For Tf1-Mac, 2200 G418 resistant colonies were screened representing ~1650 integration events. In this case, eight insertions were identified in heterochromatin (Fig. 4C; data not shown). These eight insertions were cloned and all eight arose by transposition, as they were flanked by target site duplications (data not shown). These results suggest that the addition of the MAGGY CHD to Tf1 integrase alters the Tf1 integration pattern.

Since the majority of the Tf1-Mac insertions did not inte-

A

Strain	Construct	Frequency <i>neo</i> resistance	Transposition: recombination ¹
<i>swi6-115</i>	Tf1	9.8E-3 ± 1.2E-3	18:1
	Tf1-fs	5.4E-4 ± 1.20E-4	
	Tf1-Mac	1.5E-3 ± 2.57E-4	4:1
	Tf1-Mac-fs	3.5E-4 ± 4.98E-5	
<i>clr4Δ</i>	Tf1-Mac	2.2E-3 ± 4.7E-3	4:1
	Tf1-Mac-fs	4.7E-4 ± 1.6E-4	

¹Transposition:recombination is calculated as the ratio of the frequency *neo* resistance for the wild type element (Tf1 or Tf1-Mac) to the frameshift mutant (Tf1-fs or Tf1-Mac-fs).



C

Strain	Construct	Transposition: recombination	Heterochromatin insertions/total	Normalized for recombination ¹
<i>swi6-115</i>	Tf1	18:1	0/2400	0/2267
	Tf1-Mac	4:1	8/2200	8/1650
<i>clr4Δ</i>	Tf1-Mac	4:1	0/2000	0/1566

¹The targeting pattern of Tf1-Mac is significantly different from Tf1 in the *swi6-115* strain ($p < 0.001$); the Tf1-Mac targeting pattern in *clr4Δ* is significantly different from the Tf1-Mac targeting pattern in the *swi6-115* strain ($p < 0.008$).

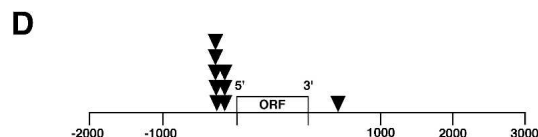


Figure 4. Target site choice of the Tf1 retrotransposon in *S. pombe* is altered by adding the MAGGY CHD to integrase. (A) Transposition and cDNA recombination frequencies of Tf1 and Tf1-Mac. The MAGGY CHD was fused to the C terminus of Tf1 integrase, creating Tf1-Mac. Tf1-Mac transposes, but at a lower frequency than wild-type Tf1. Integrase frameshift mutants (Tf1-fs and Tf1-Mac-fs) abolish Tf1 or Tf1-Mac integration, and therefore serve to infer frequencies of cDNA recombination. (B) A screen for insertions of Tf1 and Tf1-Mac in heterochromatin. A diagram of *S. pombe* centromeric repeats and mating type region depicting the location of the PCR primers used to identify heterochromatic Tf1 insertions. (C) Targeting of Tf1-Mac to heterochromatin. Centromeric heterochromatin in *S. pombe*, which is enriched in histone H3 methyl-K9, becomes a target site for Tf1-Mac in the *swi6-115* strain. In the *swi6-115* strain, there are eight integration events into heterochromatin out of 2200 Tf1-Mac transposition events. This number of transposition events includes elements incorporated into the genome by either integration or recombination. The number that arose specifically by integration was inferred by normalizing the data using the value for the ratio of transposition to recombination (4:1). The normalized heterochromatin targeting frequency is eight out of 1650 integration events. Tf1-Mac does not target into heterochromatin in the *clr4Δ* strain, which lacks H3 methyl-K9. (D) Native Tf1 target specificity is retained by the Tf1-Mac construct. Fusion of the MAGGY CHD to Tf1 does not abolish Tf1's native targeting specificity in the *swi6-115* strain. Sites of insertion for nine randomly selected Tf1 integration events were determined by inverse PCR. Eight were found within promoters and one near a transcription terminator, consistent with previously documented Tf1 targeting patterns (Behrens et al. 2000; Singleton and Levin 2002).

grate into the dg repeats or K-region, we used inverse PCR to determine the genomic integration sites of randomly selected Tf1-Mac transposition events (Fig. 4D). Of the nine integration events recovered, all were near genes and most occurred within the preferred window for Tf1 integration, namely upstream of genes transcribed by RNA polymerase II. The observed pattern was indistinguishable from wild-type Tf1 (Behrens et al. 2000; Singleton and Levin 2002). We concluded that native Tf1 targeting mechanisms determine the integration site for the majority of Tf1-Mac insertions and that the addition of the MAGGY CHD directs ~1/200 integration events to the regions of heterochromatin surveyed in our screen. In terms of possible target sites, the intergenic regions in euchromatin approximate 7 Mb, whereas our PCR screen surveyed less than 0.1 Mb of heterochromatin. Therefore, in terms of availability alone, there are many more euchromatic integration sites for Tf1 relative to sites in heterochromatin. The efficiency differences of the two targeting mechanisms could also be due to the difference in affinity between targeting determinants and their cellular partners, which cannot be evaluated because the interacting partner of Tf1's chromodomain remains unknown.

It is possible that addition of the MAGGY CHD impaired Tf1's native targeting mechanism resulting in the eight heterochromatic Tf1-Mac insertions, rather than directing targeting to heterochromatin through interaction between MAGGY CHD and histone H3 methyl-K9. To control for this possibility, we first tested transposition of a Tf1-Mac variant with mutations in conserved residues that recognize H3 methyl-K9 (the RW-to-VV mutant). This element, however, was compromised for transposition (data not shown), and so as an alternative control, we tested transposition of Tf1-Mac in a *clr4Δ* strain. This strain lacks histone H3 methyl-K9 (Nakayama et al. 2001) and abrogates the subnuclear localization of the MAGGY EGFP-CHD (Supplemental Fig. S3B). Transposition and recombination frequencies were comparable in the *swi6-115* and *clr4Δ* strains (Fig. 4A). A screen of 2000 G418 resistant colonies approximating 1566 integration events failed to recover any Tf1-Mac insertions in heterochromatin in the *clr4Δ* strain (Fig. 4C). The Tf1-Mac targeting pattern in the *swi6-115* strain is significantly different than in the *clr4Δ* strain ($P < 0.008$). This control suggests that Tf1-Mac integrates into heterochromatin because of the interaction between the MAGGY CHD and histone H3 methyl-K9.

We are aware that binding specificities differ for some CHDs depending on whether they are assessed in vitro or in vivo. For example, the *Arabidopsis* TFL2 CHD binds H3 di- and trimethyl-K9 as well as trimethyl-K27 in vitro but primarily recognizes H3 trimethyl-K27 in vivo (Turck et al. 2007; Zhang et al. 2007). We recognize that our heterologous experimental system for measuring MAGGY CHD function (i.e., *S. pombe* vs. *M. grisea*) has limitations; however, both our in vitro biochemical and in vivo genetic experiments suggest that the MAGGY CHD interacts exclusively with H3 di- and trimethyl K9. We note that in the filamentous fungus, *Neurospora crassa*, which shares a common ancestor with *M. grisea* (200 million years ago), the HP1 homolog interacts with H3 trimethyl-K9 both in vivo and in vitro (Freitag et al. 2004).

Some HP1 isoforms in mammals and *Drosophila melanogaster* target proteins to euchromatin where they repress gene expression (Hiragami and Festenstein 2005). Such differences in targeting may be due to sequences outside of the chromodomain, and we cannot rule out the possibility that other sequences in integrase play a role in targeting. Moreover, different targeting

patterns among chromoviruses could result from different cellular partners of their CHDs. Despite these possibilities, our experiments demonstrate that the chromovirus CHD itself can target integration, and this observation—coupled with our data indicating that chromoviruses are enriched in gene poor regions, GFP-CHD fusion proteins localize to heterochromatin, and CHDs recognize histone modifications—support the conclusion that in their native context, chromovirus CHDs perform a targeting function.

Heterochromatin and transposable elements

Targeted integration into domains of heterochromatin likely has significant consequences for both the retrotransposon and their hosts. In plants, for example, retroelement activity is one of the most important factors contributing to genome expansion (San-Miguel et al. 1996), and rampant transposition would be predicted to expand domains of heterochromatin. Support for this comes from analysis of pericentromeric heterochromatin in *A. thaliana*, which has undergone about a 6- to 10-fold expansion within the last 5 million years relative to related species (Hall et al. 2006). In contrast, the size of euchromatin in these species is largely unchanged.

Selective targeting to heterochromatin might benefit the mobile element by allowing it to avoid negative selection arising from insertion into genes (Boeke and Devine 1998). Although the mobile element may be silenced, this would not necessarily compromise its ability to transpose, especially for those that insert in facultative heterochromatin. There are also many examples of genes that reside in constitutive heterochromatin (Grewal and Elgin 2007). Among these are the *S. cerevisiae* Ty5 retrotransposons, which are transcriptionally activated 20-fold by the pheromone response pathway despite their location in heterochromatin (Ke et al. 1997).

Whereas we provide evidence that retrotransposon CHDs target integration to heterochromatin by recognizing epigenetic marks, abundant evidence has accumulated over the past several years indicating that transposable element insertions themselves perpetuate the epigenetic marks that define heterochromatin (Slotkin and Martienssen 2007). The RNAi pathway keeps transposable element insertions quiescent and is triggered by double-stranded RNA that arises from read-through transcription of nested transposable elements (Fig. 5). siRNAs are then generated by Dicer (DICER1) through cleavage of dsRNA, which ultimately leads to the silencing of the mobile element by DNA and histone methylation. As suggested by our data, the epigenetically marked transposable elements now become targets of additional integration events, and this self-perpetuating mechanism creates a genomic safe haven for the mobile elements and reinforces the expansion and persistence of domains of heterochromatin. Although it is clear that forces such as recombination and selection impact localization of mobile elements in heterochromatin, we believe we are only beginning to appreciate the extent to which integration site choice contributes to the landscape of the eukaryotic genome.

Methods

Strains and plasmids

YFP and CFP coding sequences from plasmids pSKY36 and pSKC36 (gift of S. Howell, Iowa State University) (Banno et al. 2001) were amplified by primers DVO3302 and DVO3303 and

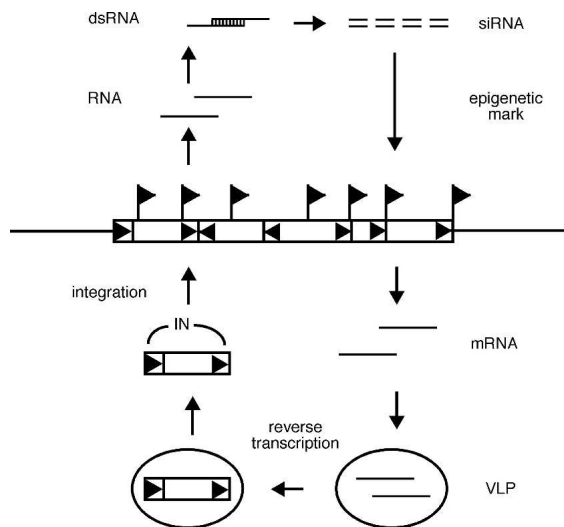


Figure 5. Self-perpetuating model for heterochromatin expansion. Repetitive elements such as retrotransposons produce dsRNAs that trigger the RNAi pathway. This results in the targeting of DNA and histone methyltransferases to retrotransposons resident on host chromosomes. Histone modification (e.g., H3 methyl-K9) establishes heterochromatin and, in turn, creates recognition sites for retrotransposon integrases. The targeting of retrotransposons to the domain of heterochromatin serves to reinforce the epigenetic mark.

cloned into the XbaI and SacII sites of pYH37 (Havecker et al. 2005) to create pYH48 and pYH49, respectively. (The DNA sequences of all primers used in this study are available upon request.) The SV40 large T antigen nuclear localization signal (NLS) (MAPKKKRKV) (Kalderon et al. 1984). TFL2 was PCR amplified using pAVA121NF as a template (gift of V. Gaudin; Gaudin et al. 2001) and primers DVO3259 and DVO3260. The PCR fragment was cloned into the XmaI and SacI sites of pYH49 to create pYH94. The MAGGY, Tma, and Os CHDs were amplified by PCR from genomic DNA extracted from *M. grisea* (gift of R. Dean, North Carolina State University), *A. thaliana*, *O. sativa*, and *Z. mays*, respectively (primers DVO3439 and DVO3440, MAGGY; DVO3484 and DVO3486, Tma; DVO3489 and DVO3490, Os; DVO3327 and DVO3328, CRM). All of the above PCR products were cloned into the XmaI and SacI sites of pYH48 to create pXG147 (MAGGY CHD), pXG156 (Tma3 CHD), pXG158 (Os CHD), and pYH67 (CRM CR motif).

To create His₆-tagged CHDs for the pull-down experiments, the coding sequences of the CHDs were PCR amplified (using primers DVO3567 and DVO3440, MAGGY; DVO3520 and DVO3486, Tma; DVO3522 and DVO3488, Os) and cloned into the NdeI and SacI sites of pET28b+ (Novagen). GST fusion constructs were generated by cloning PCR products (generated with primers DVO3622 and DVO3440, MAGGY; DVO3623 and DVO3486, Tma; DVO3624 and DVO3490, Os) into the NcoI and SacI sites of pET42 (Novagen). Mutant chromodomains were generated by site-specific PCR mutagenesis (Ausubel et al. 1987) (primers DVO3557 and DVO3558 for the MAGGY R28W29 to VV mutation [amino acid positions are according to sequence in Supplemental Fig. S1; spaces are not counted]; primers DVO3559 and DVO3560 for the Tma W31 to V mutation; primers DVO3561 and DVO3562 for the K28W29 to VV mutation). The BL21-codon plus *Escherichia coli* strain (Invitrogen) was used for protein expression.

The MAGGY CHD was fused to the C terminus of Tf1 integrase by two-step PCR-based gene construction (Dillon and

Rosen 1990) using primers DVO3378–DVO3582. The PCR product encoding the gene fusion was used to replace the NarI and BsrGI restriction fragment of the Tf1 element in pHL414-2 (Levin et al. 1993), thereby creating Tf1-Mac (pXG175). The gene fusion was also cloned into the Tf1 frameshift mutant in pHL431 (Hoff et al. 1998) to create Tf1-Mac-fs (pXG194). The *swi6-115* mutant (YB429: *h90 ade6-M216 leu1-32 urs4-DS/E swi6-115*) was used to test Tf1-Mac target specificity (gift of J. Nakayama; Nakayama et al. 2001). yXG150 was derived from SPM1049 (*h90 ade6-M210 leu1-32 uea4-D18 clr4::KanMX6*) (gift of J. Nakayama; Nakayama et al. 2001) to detect Tf1-Mac transposition in the *clr4Δ* strain. Because Tf1 carries a marker gene that confers kanamycin resistance after transposition, the KanMX6 construct was deleted in yXG150. This was accomplished by first carrying out a two-step PCR amplification to create a *LEU2* gene with 200 bp of flanking KanMX6 sequence. Primers used for the amplification were DVO4040–DVO4044. The PCR product was then used for targeted gene replacement.

Constructs to monitor subnuclear localization in *S. pombe* were generated by PCR-amplifying chromodomains of MAGGY, Tma, and Os (DVO3718 and DVO3721, MAGGY; DVO3719 and DVO3521, Tma; DVO3720 and DVO3523, Os). The 5' primer for each amplification was designed to encode the SV40 large T antigen NLS (MAPKKKRKV) (Kalderon et al. 1984). PCR products were cloned into the NdeI and BamHI of pYB232 to generate fusions with EGFP under the control of the *nmt* promoter (gift of J. Nakayama; Sadaie et al. 2004). As a control for the subnuclear localization experiments, monomeric red fluorescent protein (mRFP) was PCR-amplified (DVO4076 and DVO4077) using pFA6a-mRFP-hph as a template (gift of T. Toda; Asakawa et al. 2005). The PCR product was cloned into the NdeI site of pREP42 (Maundrell 1993) to generate pXG204. The Chp2 coding sequence was PCR-amplified (DVO4156 and DVO4157) using pYB232 as a template. The product was digested with XhoI and BamHI and cloned into the Sall and BamHI sites of pXG204. To monitor the subnuclear localization of chromodomains in *S. pombe*, the strain yXG126 was derived from spYB399 (pYB232/*h90 ade6-M216 leu1-32 ura4-DS/E*) (Sadaie et al. 2004) by losing plasmid pYB232.

DNA sequence analysis

DNA sequences of chromoviruses and chromodomains from cellular genes were retrieved from the National Center for Biotechnology Information (NCBI). The chromovirus integrase C termini were used to query rice (Sasaki et al. 2002) and *A. thaliana* (The Arabidopsis Genome Initiative 2000) genome sequences by TBLASTN to identify sites of chromovirus insertions. Retrotransposon distributions were mapped on chromosomes using a BioPerl script. The sequence alignment of chromodomains was generated by ClustalX (Thompson et al. 1997) and with manual adjustments. Secondary structure predictions of retrotransposon chromodomains was carried out using Jpred software (Cuff et al. 1998).

To assess the genomic distribution of MAGGY insertions, the amino acid sequence of MAGGY integrase was used to query the *M. grisea* genome (NCBI, 2003 version) (Dean et al. 2005) by TBLASTN (Altschul et al. 1997). Sequence hits with *E*-values $< 1 \times 10^{-4}$ and $> 70\%$ identity and 90% coverage were collected (75 elements). Using the coding sequence (CDS) annotations from GenBank, the numbers of genes or transposable elements were then counted in 10-kb windows both upstream and downstream of the insertion site. A similar analysis was conducted with 75 random sites selected by a random number generator. For the *O. sativa* genome, the locations of chromodomain-encoding retrotransposons and Tos17 (AAP53905) were mapped

against the protein database annotated by The Institute for Genomic Research (TIGR) (Yuan et al. 2005). For these analyses, the amino acid sequence of the *rn_377-208* (Os) and *Tos17* integrase C termini were used to query the rice protein database by BLASTP (Altschul et al. 1997). Insertions were identified by the following cutoff criteria: 1×10^{-4} , 30% identify, and 90% alignment coverage for *rn_377-208* (917 elements); 1×10^{-4} , 40% identity, and 90% alignment coverage for *Tos17* (106 elements). Annotated coding regions were evaluated to determine the number of genes adjacent to retrotransposon insertions. Annotations were manually checked to ensure that each annotation corresponded to a unique insertion. It is likely that many degenerate insertions and solo LTRs are not included in the available annotations, and so the actual number of transposable elements adjacent to a given insertion is likely an underestimate; however, this underestimate only strengthens the conclusion that chromodomain-encoding retrotransposons are associated with other mobile elements. Analyses were also performed in which the chromovirus group under question was excluded from the data set to control for self-association. For rice, when other group II elements were eliminated from the 40-kb intervals, the number of transposable elements per interval dropped to 2.6 from 2.9, which is still significantly different ($P < 0.00001$) from the number of transposable elements associated with random sites or *Tos17* elements. For *MAGGY* and *Tos17*, no other *MAGGY*- or *Tos17*-like elements were present in the intervals analyzed. The significance of the association of retrotransposons with genes or transposable elements was evaluated by the t-test.

The LTR sequences of the *Tos17* and rice group II chromoviruses were collected as described (Peterson-Burch et al. 2004). Comparisons of the 5' and 3' LTR sequences were performed to determine the number of substitutions per neutral site (Jukes and Cantor 1969). This provided an estimate of the age of the elements, since the two LTRs are typically identical at the time of integration.

In vivo protein localization

Established protocols were used for transformation and heterologous protein expression in *S. pombe* and *A. thaliana* suspension cells (Moreno et al. 1991; Havecker et al. 2005). Images were collected using a Leica upright confocal microscope with a 60 \times water objective lens and 4 \times zoom lens at the Image Analysis Facility, College of Veterinary Medicine, Iowa State University or with Nikon Eclipse E800 microscope at the CBS Imaging Center, University of Minnesota.

Pull-down assays

His₆- and GST-CHD fusion proteins were purified from *E. coli* extracts with Ni-agarose beads and glutathione-sepharose as described by the manufacturer (Sigma). For pull-down assays with the calf thymus histones (Worthington), GST-tagged chromodomains were incubated with the histone extract (5 μ g) in BTP buffer (25 mM BTP at pH 6.8, 1M KCl, 0.5% Triton X-100) for 2 h and washed six times with binding buffer (Huyen et al. 2004). Samples were separated by SDS-PAGE, and gels were stained with SYPRO Ruby Protein Gel Stain (Invitrogen). To identify interactions with specific H3 modifications, His₆-tagged fusion proteins were incubated with biotinylated H3 peptides and streptavidin agarose in binding buffer (150 mM NaCl, 50 mM HEPES at pH 7.5, 0.1% Tween, and 10% glycerol) for 2 h and washed with binding buffer four times. Bound proteins were eluted with SDS loading buffer at 37°C for 30 min, followed by immunoblotting using an anti-His₆ antibody (Qiagen). Similarly, GST-tagged fusion chromodomain proteins were incubated with biotinylated

H3 peptides with di- and tri-methyl H3 K9, pulled down by streptavidin agarose beads, and detected by immunoblotting using anti-GST antibodies.

Transposition assays and target site analyses

Quantitative Tf1 transposition assays were performed as previously described (Levin 1995) except that cells were scraped from EMM 5-FOA plates, diluted, and plated to YES medium with G418 and 5-FOA to select transposition events. To determine the total number of cells, dilutions were plated on YES medium without G418. The ratio of the number of G418, 5-FOA resistant cells to the total cell number was taken as the transposition frequency.

To identify insertions in *S. pombe* heterochromatin, DNA was purified (Ausubel et al. 1987) from pools of 20 colonies growing on YES + G418 + 5-FOA plates. The DNA was used as a template in PCR reactions with a Tf1 primer that was not complementary to *S. pombe* Tf2 elements (DVO3639). Eight primers were used to survey Tf1 insertions in the outer pericentromeric region and mating loci (DVO3790–DVO3793 and DVO3797–DVO3800), each of which was paired with DVO3936 in separate PCR reactions. DVO3791 and DVO3792 were used to amplify *dh* repeats as a control for DNA quality and PCR conditions. The PCR products were examined on 0.8% agarose gels and sequenced with the Tf1 primer DVO3731, which is contained in DVO3639. The resulting 5' flanking sequences were searched against the NCBI nonredundant database using BLASTN (Altschul et al. 1997) to identify Tf1 insertion sites. DNA sequences flanking the 3' end of Tf1 insertions were determined by PCR using a primer that recognizes a downstream genomic sequence and the Tf1 primer DVO3743. The PCR products were then sequenced, and the 5' and 3' flanking sequences made it possible to detect target site duplications. The Fisher exact test was performed to assess the significance of the frequency of insertions in heterochromatin.

To detect the genome-wide insertion sites of Tf1 and Tf1-Mac, inverse PCR reactions were carried out with DNA prepared from randomly selected colonies growing on YES plus G418 plates (Ochman et al. 1988). Primers used for the inverse PCR were DVO3639 and DVO3743. PCR products were sequenced and insertion sites were determined as described above. Recombination events with endogenous Tf2 sequences or solo LTR were not mapped. For each integration event, the distance to the nearest neighboring genes was determined using the *S. pombe* genome annotation (Wood et al. 2002).

Acknowledgments

We thank J. Nakayama for providing several *S. pombe* strains, M. Bedford for help in testing interactions between CHDs and modified histones, and M. Sanders for assistance with the microscopy. This work was supported by National Institutes of Health Grant GM061657 (D.V.) and the Intramural Research Program of the NIH from the National Institute of Child Health and Human Development (H.L.).

References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- Asakawa, K., Toya, M., Sato, M., Kanai, M., Kume, K., Goshima, T., Garcia, M.A., Hirata, D., and Toda, T. 2005. Mal3, the fission yeast EB1 homologue, cooperates with Bub1 spindle checkpoint to

- prevent monopolar attachment. *EMBO Rep.* **6**: 1194–1200.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. 1987. *Current protocols in molecular biology*. Greene/Wiley Interscience, New York.
- Bachman, N., Gelbart, M.E., Tsukiyama, T., and Boeke, J.D. 2005. TFIIB subunit Bdp1p is required for periodic integration of the Ty1 retrotransposon and targeting of *lsw2p* to *S. cerevisiae* tDNAs. *Genes & Dev.* **19**: 955–964.
- Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**: 120–124.
- Banno, H., Ikeda, Y., Niu, Q.W., and Chua, N.H. 2001. Overexpression of *Arabidopsis* ESR1 induces initiation of shoot regeneration. *Plant Cell* **13**: 2609–2618.
- Behrens, R., Hayles, J., and Nurse, P. 2000. Fission yeast retrotransposon Tf1 integration is targeted to 5' ends of open reading frames. *Nucleic Acids Res.* **28**: 4709–4716.
- Boeke, J.D. and Devine, S.E. 1998. Yeast retrotransposons: Finding a nice quiet neighborhood. *Cell* **93**: 1087–1089.
- Bosher, J.M. and Labouesse, M. 2000. RNA interference: Genetic wand and genetic watchdog. *Nat. Cell Biol.* **2**: E31–E36. doi: 10.1038/35000102.
- Bowen, N.J., Jordan, I.K., Epstein, J.A., Wood, V., and Levin, H.L. 2003. Retrotransposons and their recognition of pol II promoters: A comprehensive survey of the transposable elements from the complete genome sequence of *Schizosaccharomyces pombe*. *Genome Res.* **13**: 1984–1997.
- Brehm, A., Tuffeland, K.R., Aasland, R., and Becker, P.B. 2004. The many colours of chromodomains. *BioEssays* **26**: 133–140.
- Bushman, F.D. 2003. Targeting survival: Integration site selection by retroviruses and LTR-retrotransposons. *Cell* **115**: 135–138.
- Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C.R., Gu, M., Blattner, F.R., and Jiang, J. 2002. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* **14**: 1691–1704.
- Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F. 2005. A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* **11**: 1287–1289.
- Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M., and Barton, G.J. 1998. JPred: A consensus secondary structure prediction server. *Bioinformatics* **14**: 892–893.
- Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.R., Pan, H., et al. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**: 980–986.
- Derse, D., Crise, B., Li, Y., Princler, G., Lum, N., Stewart, C., McGrath, C.F., Hughes, S.H., Munroe, D.J., and Wu, X. 2007. Human T-cell leukemia virus type 1 integration target sites in the human genome: Comparison with those of other retroviruses. *J. Virol.* **81**: 6731–6741.
- Dillon, P.J. and Rosen, C.A. 1990. A rapid method for the construction of synthetic genes using the polymerase chain reaction. *Biotechniques* **9**: 298–300.
- Dong, F., Miller, J.T., Jackson, S.A., Wang, G.L., Ronald, P.C., and Jiang, J. 1998. Rice (*Oryza sativa*) centromeric regions consist of complex DNA. *Proc. Natl. Acad. Sci.* **95**: 8135–8140.
- Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & Dev.* **17**: 1870–1881.
- Flanagan, J.F., Mi, L.Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. 2005. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**: 1181–1185.
- Fransz, P., De Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I. 2002. Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci.* **99**: 14584–14589.
- Freitag, M., Hickey, P.C., Khalfallah, T.K., Read, N.D., and Selker, E.U. 2004. HP1 is essential for DNA methylation in neurospora. *Mol. Cell* **13**: 427–434.
- Gai, X. and Voytas, D.F. 1998. A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. *Mol. Cell* **1**: 1051–1055.
- Gaudin, V., Libault, M., Pouteau, S., Juul, T., Zhao, G., Lefebvre, D., and Grandjean, O. 2001. Mutations in LIKE HETEROCHROMATIN PROTEIN 1 affect flowering time and plant architecture in *Arabidopsis*. *Development* **128**: 4847–4858.
- Gendrel, A.V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R.A. 2002. Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene DDM1. *Science* **297**: 1871–1873.
- Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**: 199–202.
- Gorinsek, B., Gubensek, F., and Kordis, D. 2004. Evolutionary genomics of chromoviruses in eukaryotes. *Mol. Biol. Evol.* **21**: 781–798.
- Gorinsek, B., Gubensek, F., and Kordis, D. 2005. Phylogenomic analysis of chromoviruses. *Cytogenet. Genome Res.* **110**: 543–552.
- Grewal, S.I. and Elgin, S.C. 2007. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**: 399–406.
- Grewal, S.I. and Jia, S. 2007. Heterochromatin revisited. *Nat. Rev. Genet.* **8**: 35–46.
- Grewal, S.I. and Rice, J.C. 2004. Regulation of heterochromatin by histone methylation and small RNAs. *Curr. Opin. Cell Biol.* **16**: 230–238.
- Grivna, S.T., Beyret, E., Wang, Z., and Lin, H. 2006. A novel class of small RNAs in mouse spermatogenic cells. *Genes & Dev.* **20**: 1709–1714.
- Hall, A.E., Kettler, G.C., and Preuss, D. 2006. Dynamic evolution at pericentromeres. *Genome Res.* **16**: 355–364.
- Havecker, E.R., Gao, X., and Voytas, D.F. 2005. The Sireviruses, a plant-specific lineage of the Ty1/copia retrotransposons, interact with a family of proteins related to dynein light chain 8. *Plant Physiol.* **139**: 857–868.
- Hiragami, K. and Festenstein, R. 2005. Heterochromatin protein 1: A pervasive controlling influence. *Cell. Mol. Life Sci.* **62**: 2711–2726.
- Hizi, A. and Levin, H.L. 2005. The integrase of the long terminal repeat-retrotransposon tf1 has a chromodomain that modulates integrase activities. *J. Biol. Chem.* **280**: 39086–39094.
- Hoff, E.F., Levin, H.L., and Boeke, J.D. 1998. *Schizosaccharomyces pombe* retrotransposon Tf2 mobilizes primarily through homologous cDNA recombination. *Mol. Cell Biol.* **18**: 6839–6852.
- Houben, A., Demidov, D., Gernand, D., Meister, A., Leach, C.R., and Schubert, I. 2003. Methylation of histone H3 in euchromatin of plant chromosomes depends on basic nuclear DNA content. *Plant J.* **33**: 967–973.
- Huyen, Y., Zgheib, O., Ditullio Jr., R.A., Gorgoulis, V.G., Zacharatos, P., Petty, T.J., Sheston, E.A., Mellert, H.S., Stavridi, E.S., and Halazonetis, T.D. 2004. Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**: 406–411.
- Jacobs, S.A. and Khorasanizadeh, S. 2002. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**: 2080–2083.
- Jacobs, S.A., Taverna, S.D., Zhang, Y., Briggs, S.D., Li, J., Eissenberg, J.C., Allis, C.D., and Khorasanizadeh, S. 2001. Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. *EMBO J.* **20**: 5232–5241.
- Jiang, J., Nasuda, S., Dong, F., Scherrer, C.W., Woo, S.S., Wing, R.A., Gill, B.S., and Ward, D.C. 1996. A conserved repetitive DNA element located in the centromeres of cereal chromosomes. *Proc. Natl. Acad. Sci.* **93**: 14210–14213.
- Jiang, J., Birchler, J.A., Parrott, W.A., and Dawe, R.K. 2003. A molecular view of plant centromeres. *Trends Plant Sci.* **8**: 570–575.
- Jukes, T.H. and Cantor, C. 1969. Evolution of protein molecules. In *Mammalian protein metabolism* (ed. H.N. Munro), pp. 21–132. Academic Press, New York.
- Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. 1984. A short amino acid sequence able to specify nuclear location. *Cell* **39**: 499–509.
- Ke, N., Irwin, P.A., and Voytas, D.F. 1997. The pheromone response pathway activates transcription of Ty5 retrotransposons located within silent chromatin of *Saccharomyces cerevisiae*. *EMBO J.* **16**: 6272–6280.
- Koonin, E.V., Zhou, S., and Lucchesi, J.C. 1995. The chromo superfamily: New members, duplication of the chromo domain and possible role in delivering transcription regulators to chromatin. *Nucleic Acids Res.* **23**: 4229–4233.
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**: 116–120.
- Le, Q.H., Melayah, D., Bonnivard, E., Petit, M., and Grandbastien, M.A. 2007. Distribution dynamics of the Tnt1 retrotransposon in tobacco. *Mol. Genet. Genomics* **278**: 639–651.
- Levin, H.L. 1995. A novel mechanism of self-primed reverse transcription defines a new family of retroelements. *Mol. Cell Biol.* **15**: 3310–3317.
- Levin, H.L., Weaver, D.C., and Boeke, J.D. 1993. Novel gene expression mechanism in a fission yeast retroelement: Tf1 proteins are derived from a single primary translation product. *EMBO J.* **12**: 4885–4895.
- Lewinski, M.K., Yamashita, M., Eberman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannehalli, S., et al. 2006. Retroviral DNA integration: viral and cellular determinants of

- target-site selection. *PLoS Pathog* **2**: e60. doi: 10.1371/journal.ppat.0020060.
- Lippman, Z. and Martienssen, R. 2004. The role of RNA interference in heterochromatic silencing. *Nature* **431**: 364–370.
- Llano, M., Saenz, D.T., Meehan, A., Wongthida, P., Peretz, M., Walker, W.H., Teo, W., and Poeschla, E.M. 2006. An essential role for LEDGF/p75 in HIV integration. *Science* **314**: 461–464.
- Malik, H.S. and Eickbush, T.H. 1999. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. *J. Virol.* **73**: 5186–5190.
- Marin, I. and Llorens, C. 2000. Ty3/Gypsy retrotransposons: Description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. *Mol. Biol. Evol.* **17**: 1040–1049.
- Martienssen, R., Lippman, Z., May, B., Ronemus, M., and Vaughn, M. 2004. Transposons, tandem repeats, and the silencing of imprinted genes. *Cold Spring Harb. Symp. Quant. Biol.* **69**: 371–379.
- Maudrell, K. 1993. Thiamine-repressible expression vectors pREP and pRIP for fission yeast. *Gene* **123**: 127–130.
- Miller, J.T., Dong, F., Jackson, S.A., Song, J., and Jiang, J. 1998. Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* **150**: 1615–1623.
- Min, J., Zhang, Y., and Xu, R.M. 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes & Dev.* **17**: 1823–1828.
- Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K., and Hirochika, H. 2003. Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* **15**: 1771–1780.
- Moreno, S., Klar, A., and Nurse, P. 1991. Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.* **194**: 795–823.
- Mou, Z., Kenny, A.E., and Curcio, M.J. 2006. Hos2 and Set3 promote integration of Ty1 retrotransposons at tRNA genes in *Saccharomyces cerevisiae*. *Genetics* **172**: 2157–2167.
- Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., Henikoff, S., Buell, C.R., and Jiang, J. 2004. Sequencing of a rice centromere uncovers active genes. *Nat. Genet.* **36**: 138–145.
- Nakagawa, H., Lee, J.K., Hurwitz, J., Allshire, R.C., Nakayama, J., Grewal, S.I., Tanaka, K., and Murakami, Y. 2002. Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. *Genes & Dev.* **16**: 1766–1778.
- Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**: 110–113.
- Nakayashiki, H., Kiyotomi, K., Tosa, Y., and Mayama, S. 1999. Transposition of the retrotransposon MAGGY in heterologous species of filamentous fungi. *Genetics* **153**: 693–703.
- Nielsen, P.R., Nietispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V., and Laue, E.D. 2002. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**: 103–107.
- Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S.I. 2004. RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat. Genet.* **36**: 1174–1180.
- Ochman, H., Gerber, A.S., and Hartl, D.L. 1988. Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**: 621–623.
- Pereira, V. 2004. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biol.* **5**: R79. doi: 10.1186/gb-2004-5-10-r79.
- Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., et al. 2001. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**: 323–337.
- Peterson-Burch, B.D., Nettleton, D., and Voytas, D.F. 2004. Genomic neighborhoods for *Arabidopsis* retrotransposons: A role for targeted integration in the distribution of the Metaviridae. *Genome Biol.* **5**: R78. doi: 10.1186/gb-2004-5-10-r78.
- Sadaie, M., Iida, T., Urano, T., and Nakayama, J. 2004. A chromodomain protein, Chp1, is required for the establishment of heterochromatin in fission yeast. *EMBO J.* **23**: 3825–3835.
- SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., et al. 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**: 765–768.
- Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., et al. 2002. The genome sequence and structure of rice chromosome 1. *Nature* **420**: 312–316.
- Schroder, A., Shinn, P., Chen, H., Berry, C., Ecker, J., and Bushman, F. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**: 521–529.
- Shun, M.C., Raghavendra, N.K., Vandegraaff, N., Daigle, J.E., Hughes, S., Kellam, P., Cherepanov, P., and Engelman, A. 2007. LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes & Dev.* **21**: 1767–1778.
- Sijen, T. and Plasterk, R.H. 2003. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**: 310–314.
- Singleton, T.L. and Levin, H.L. 2002. A long terminal repeat retrotransposon of fission yeast has strong preferences for specific sites of insertion. *Eukaryot. Cell* **1**: 44–55.
- Slotkin, R.K. and Martienssen, R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8**: 272–285.
- Tamaru, H. and Selker, E.U. 2001. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* **414**: 277–283.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.
- Thon, M.R., Martin, S.L., Goff, S., Wing, R.A., and Dean, R.A. 2004. BAC end sequences and a physical map reveal transposable element content and clustering patterns in the genome of *Magnaporthe grisea*. *Fungal Genet. Biol.* **41**: 657–666.
- Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. 2007. *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet.* **3**: e86. doi: 10.1371/journal.pgen.0030086.
- Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**: 320–324.
- Vastenhouw, N.L. and Plasterk, R.H. 2004. RNAi protects the *Caenorhabditis elegans* germline against transposition. *Trends Genet.* **20**: 314–319.
- Wang, G.P., Ciuffi, A., Leipzig, J., Berry, C.C., and Bushman, F.D. 2007. HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **17**: 1186–1194.
- Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., et al. 2002. The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**: 871–880.
- Wu, X., Li, Y., Crise, B., and Burgess, S.M. 2003. Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**: 1749–1751.
- Xie, W., Gai, X., Zhu, Y., Zappulla, D.C., Sternglanz, R., and Voytas, D.F. 2001. Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. *Mol. Cell. Biol.* **21**: 6606–6614.
- Yieh, L., Kassavetis, G., Geiduschek, E.P., and Sandmeyer, S.B. 2000. The Brf and TATA-binding protein subunits of the RNA polymerase III transcription factor IIIB mediate position-specific integration of the gypsy-like element, Ty3. *J. Biol. Chem.* **275**: 29800–29807.
- Yieh, L., Hatzis, H., Kassavetis, G., and Sandmeyer, S.B. 2002. Mutational analysis of the transcription factor IIIB-DNA target of Ty3 retroelement integration. *J. Biol. Chem.* **277**: 25920–25928.
- Yuan, Q., Ouyang, S., Wang, A., Zhu, W., Maiti, R., Lin, H., Hamilton, J., Haas, B., Sultana, R., Cheung, F., et al. 2005. The institute for genomic research Osa1 rice genome annotation database. *Plant Physiol.* **138**: 18–26.
- Zemach, A., Li, Y., Ben-Meir, H., Oliva, M., Mosquna, A., Kiss, V., Avivi, Y., Ohad, N., and Grafi, G. 2006. Different domains control the localization and mobility of LIKE HETEROCHROMATIN PROTEIN1 in *Arabidopsis* nuclei. *Plant Cell* **18**: 133–145.
- Zhang, X., Germann, S., Blus, B.J., Khorasanizadeh, S., Gaudin, V., and Jacobsen, S.E. 2007. The *Arabidopsis* LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat. Struct. Mol. Biol.* **14**: 869–871.
- Zou, S., Ke, N., Kim, J.M., and Voytas, D.F. 1996. The *Saccharomyces* retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. *Genes & Dev.* **10**: 634–645.

Received September 16, 2007; accepted in revised form December 20, 2007.



Chromodomains direct integration of retrotransposons to heterochromatin

Xiang Gao, Yi Hou, Hiroataka Ebina, et al.

Genome Res. 2008 18: 359-369

Access the most recent version at doi:[10.1101/gr.7146408](https://doi.org/10.1101/gr.7146408)

Supplemental Material <http://genome.cshlp.org/content/suppl/2008/04/21/gr.7146408.DC2>
<http://genome.cshlp.org/content/suppl/2008/02/26/gr.7146408.DC1>

References This article cites 97 articles, 43 of which can be accessed free at:
<http://genome.cshlp.org/content/18/3/359.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Affordable, Accurate
Sequencing.



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>