Supplementary Figure Legends

Figure S1. Amino acid sequence alignments of retrotransposon CHDs. The alignments are similar to those described in previous studies (Gorinsek et al. 2004; Gorinsek et al. 2005; Marin and Llorens 2000), and all of the amino acid sequences were aligned together with the exception of the CR motifs, which were aligned separately. The CHDs separate into two groups based on similarity to CHDs from cellular proteins. Among conserved residues shared by group I and cellular CHDs are three aromatic amino acids that interact with methylated lysine residues on histone H3 in the HP1 CHD (arrowheads) (Jacobs and Khorasanizadeh 2002; Nielsen et al. 2002). Only one of these aromatic residues is conserved among the group II CHDs (arrowhead). Another is completely absent (X) and the third is present only in some sequences (?). Residues mutated and tested for effects on CHD subnuclear localization (Figure 2B) are shaded in dark red. Arrows above the alignment denote conserved secondary structures shared by the cellular, group I and group II CHDs. The CR retrotransposons encode a very different motif at a position in integrase that corresponds to the location of the CHDs in the group I and group II elements. The two letter abbreviations before each retrotransposon sequence designate species of origin (At, *Arabidopsis thaliana*; Cc, *Coprinopsis cinerea*; Cf , *Cladosporium fulvum*; Cr, *Chlamydomonas reihardtii*; Cn, *Cryptococcus neoformans*; Dm , *Drosophila melanogaster*; Dr, *Danio rerio*; Fo , *Fusarium oxysporum*; Ga, *Gasterosteus aculeatus*; Hv, *Hordeum vulgare*; Le, *Lycopersicon esculentum*; Lj, *Lotus japonicus*; Mg, *Magnaporthe grisea*; Mm, *Mus musculus*; Mt, *Medicago truncatula*; Os, *Oryza sativa*; Pt, *Poncirus trifoliate*; Sb, *Sorghum bicolor*; Sp, *Schizosaccharomyces pombe*; Tm,

Tricholoma matsutake; Tr, *Takifugu rubripes*; Uh, *Ustilago hordei*; Xt, *Xenopus tropicalis*; Zm, *Zea mays*). GenBank accession numbers for the retrotransposon sequences are as follows: Dm HP1a, P05205; Mm HP1a, Q61686; Sp Swi6,_P40381; Sp Chp2, CAA16917; Fo_Skippy, L34658; Cf Cft1, AF051915; Mg Pyret, Ab062507; Cr Chlamydomonas, sc_2123; Mg MGRL-3, AF314096; Uh Uhchromovir1, AC114899; Tm MarY1, AB028236; Xt Silurana, AC147356; Dr Drsushi33, AL596141; Dr Drsushi5, AL714031; Ga Gasushi, AC145765; Tr sushi, AF030881; Dr Drsushi, AL732562; Mg MAGGY, L35053; Cc Ccchromovir, AACS01000152; Lj Ljchro, AP004915; Os Osr35, AC068924; Os rn_377-208, AK068625; Zm 1, AF466646; Zm Reina, U69258; At 1, AP002071a; Os Osr34, AP004365a; Oz RIRE3, AC119148; Zm Tekay, AF050455; Os dagul, AC087542; Sb RetroSor2, AF061282; At Tma, AF147263; At Tma3-1, AC005965; Pt 1, AF506028; Pt 2, AF506028; Le galariel, AF119040; Oz CRR2, AK068116; Zm CRM, AC152494; Hv Cereba, AY040832; Mt 1, AC131249; Lj 1, AP004525; At CRA2, At2g06890; At CRA3, AC007887). Asterisks indicated those elements studied in detail.

Figure S2. Retrotransposons with CHDs and CR motifs are preferentially located in gene-poor, transposon-rich regions of their host genomes. **A)** The genomic distribution of Ty1/*copia* (yellow), group II (red) and CR (blue) retrotransposon insertions in the *A. thaliana* genome. Sequences representing the C-termini of Tos17 (Ty1/*copia*), Tma (group II), and CR element integrases were used to query the annotated *A. thaliana* genome sequence by tblastn. Sequence hits with E-values lower then 1E-4 and over 70% identity and 90% coverage were considered. The positions of insertions are

plotted on each of the five *A. thaliana* chromosomes. Note that both types of elements are enriched in pericentromeric regions (depicted in black), which are the major domains of heterochromatin in *A. thaliana* (Fransz et al. 2002). In contrast, Ty1/*copia* retrotransposon insertions extend further out along the chromosome arms (Peterson-Burch et al. 2004). The breaks in the chromosome denote the centromeres, which were not sequenced. **B)** DNA sequence divergence between LTRs of full-length Tos17 and Os elements. The 5' and 3' LTR sequences from each element were parsed using a Perl script, and divergence was estimated using the Jukes-Cantor model (Jukes and Cantor 1969).

Figure S3. Further characterization of the subnuclear localization of chromovirus CHDs. **A)** The LHP1 protein of *A. thaliana* co-localizes with CENP-C in protoplasts. CENP-C is a marker for chromocenters in *A. thaliana* (Shibata and Murata 2004). **B)** Subnuclear localization of the MAGGY CHD in *S. pombe* requires conserved residues in the CHD (first panel). Neither the MAGGY CHD nor Chp2 localize to heterochromatin in a *clr4Δ* strain, which lacks H3 methyl-K9 (Nakayama et al. 2001) (second and third panels). **C)** Unlike the MAGGY CHD, the Tma and Os CHDs do not localize to heterochromatin in *S. pombe*.

Figure S4. The MAGGY CHD recognizes histone H3 methyl-K9. **A)** MAGGY CHD interacts with purified histone H3. Individually purified histones from calf thymus were used in pull-down experiments with a His_6-tagged MAGGY CHD. The pull-down products were separated by SDS-PAGE and visualized by Coomassie blue staining

(upper panel). A specific interaction was only observed between the MAGGY CHD and histone H3. The weak interaction with H4 is non-specific, since it is observed with control proteins that lack the MAGGY CHD (lower panel). M, a histone mixture. **B)** The MAGGY CHD interacts with both H3 dimethyl-K9 and H3 dimethyl-K9 peptides. The experiment was performed as described in the Figure 3B legend. Specifically, biotin-labeled histone peptides were incubated with His_6-tagged MAGGY CHD, and then pulled-down with streptavidin agarose beads. Pull-down reactions were separated by SDS PAGE and transferred to nylon membranes. The presence of the CHD was detected using an anti-His_6 antibody.

Supplementary References

Fransz, P., J.H. De Jong, M. Lysak, M.R. Castiglione, and I. Schubert. 2002. Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. *Proc Natl Acad Sci U S A* **99:** 14584-14589.

Gorinsek, B., F. Gubensek, and D. Kordis. 2004. Evolutionary genomics of chromoviruses in eukaryotes. *Mol Biol Evol* **21:** 781-798.

Gorinsek, B., F. Gubensek, and D. Kordis. 2005. Phylogenomic analysis of chromoviruses. *Cytogenet Genome Res* **110:** 543-552.

Jacobs, S.A. and S. Khorasanizadeh. 2002. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295:** 2080-2083.

Jukes, T.H. and C. Cantor. 1969. Evoltion of protein molecules. In *Mammalian protein metabolism*, pp. 21-132. Academic Press, New York.

Marin, I. and C. Llorens. 2000. Ty3/Gypsy retrotransposons: description of new Arabidopsis thaliana elements and evolutionary perspectives derived from comparative genomic data. *Mol Biol Evol* **17:** 1040-1049.

Nakayama, J., J.C. Rice, B.D. Strahl, C.D. Allis, and S.I. Grewal. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292:** 110-113.

Nielsen, P.R., D. Nietlispach, H.R. Mott, J. Callaghan, A. Bannister, T. Kouzarides, A.G. Murzin, N.V. Murzina, and E.D. Laue. 2002. Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416:** 103-107.

Peterson-Burch, B.D., D. Nettleton, and D.F. Voytas. 2004. Genomic neighborhoods for Arabidopsis retrotransposons: a role for targeted integration in the distribution of the Metaviridae. *Genome Biol* **5:** R78.

Shibata, F. and M. Murata. 2004. Differential localization of the centromere-specific proteins in the major centromeric satellite of Arabidopsis thaliana. *J Cell Sci* **117:** 2963-2970.

Figure S1 Gao *et al.*

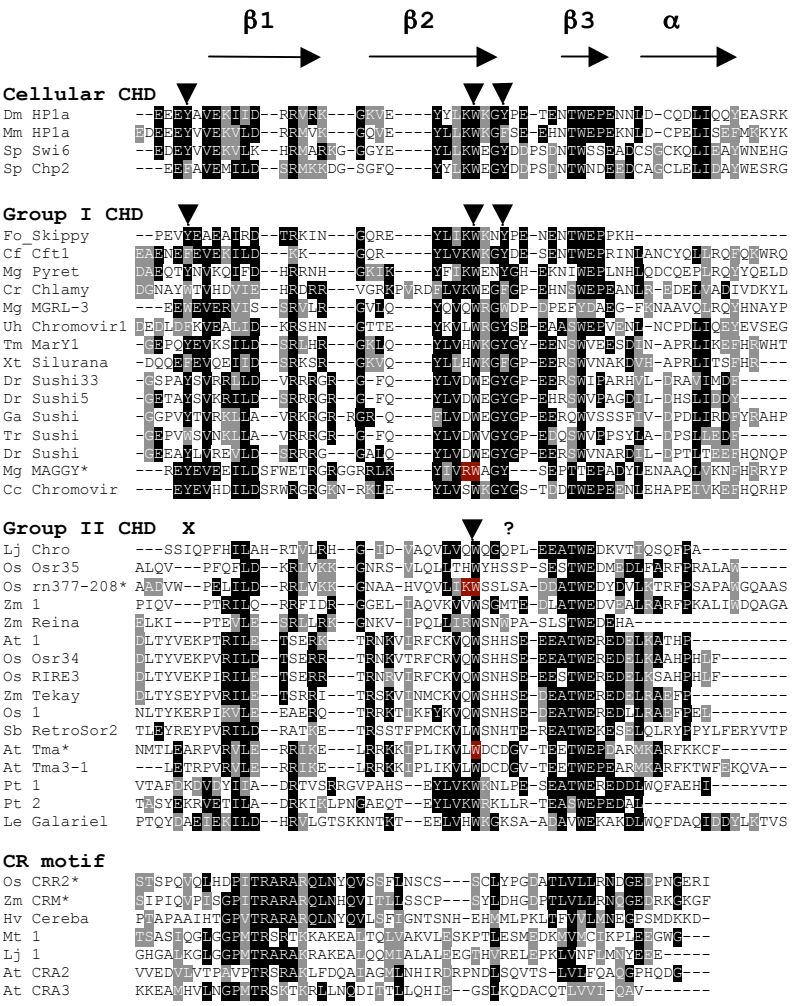
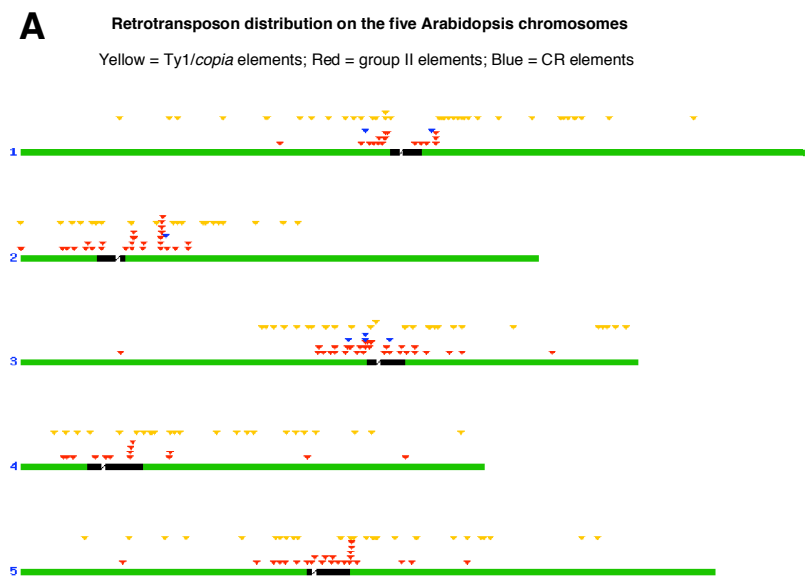


Figure S2 Gao *et al.*



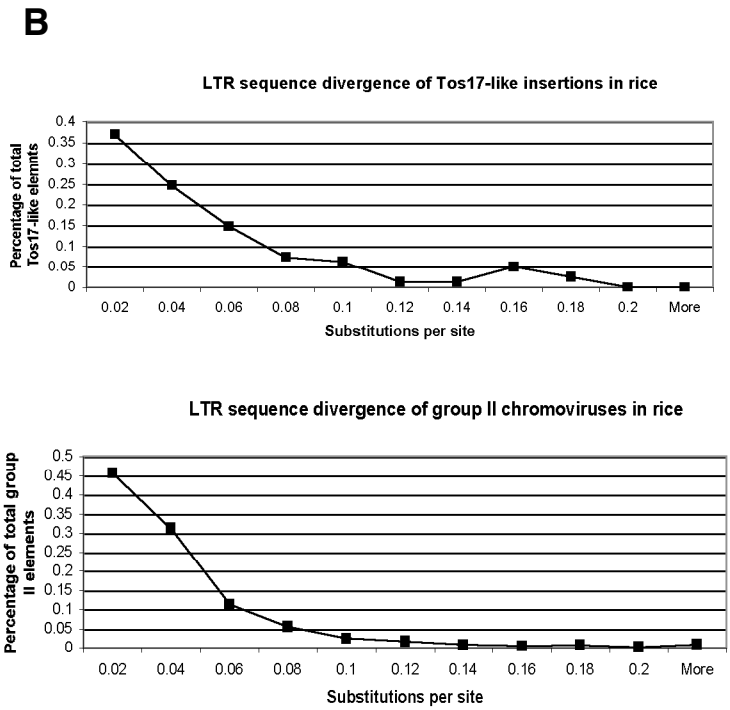


Figure S3 Gao *et al*.

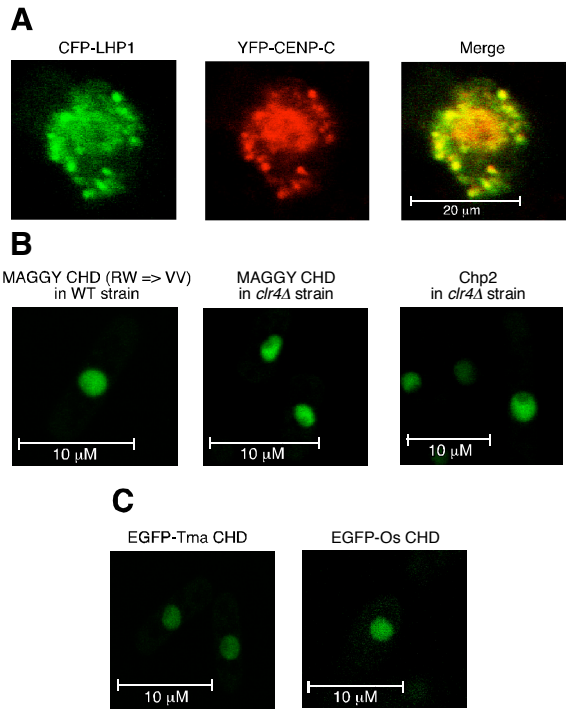


Figure S4 Gao *et al*.