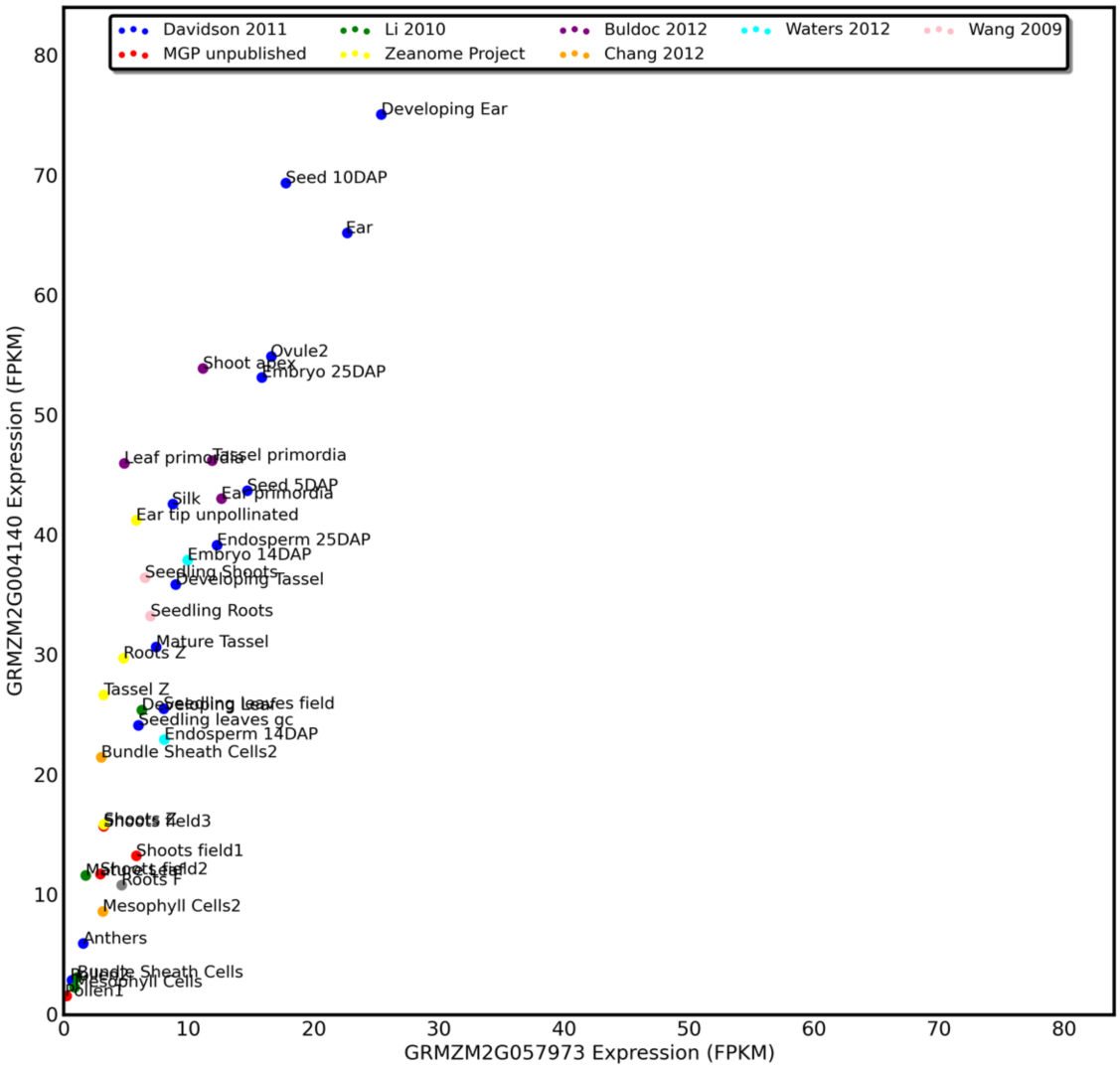
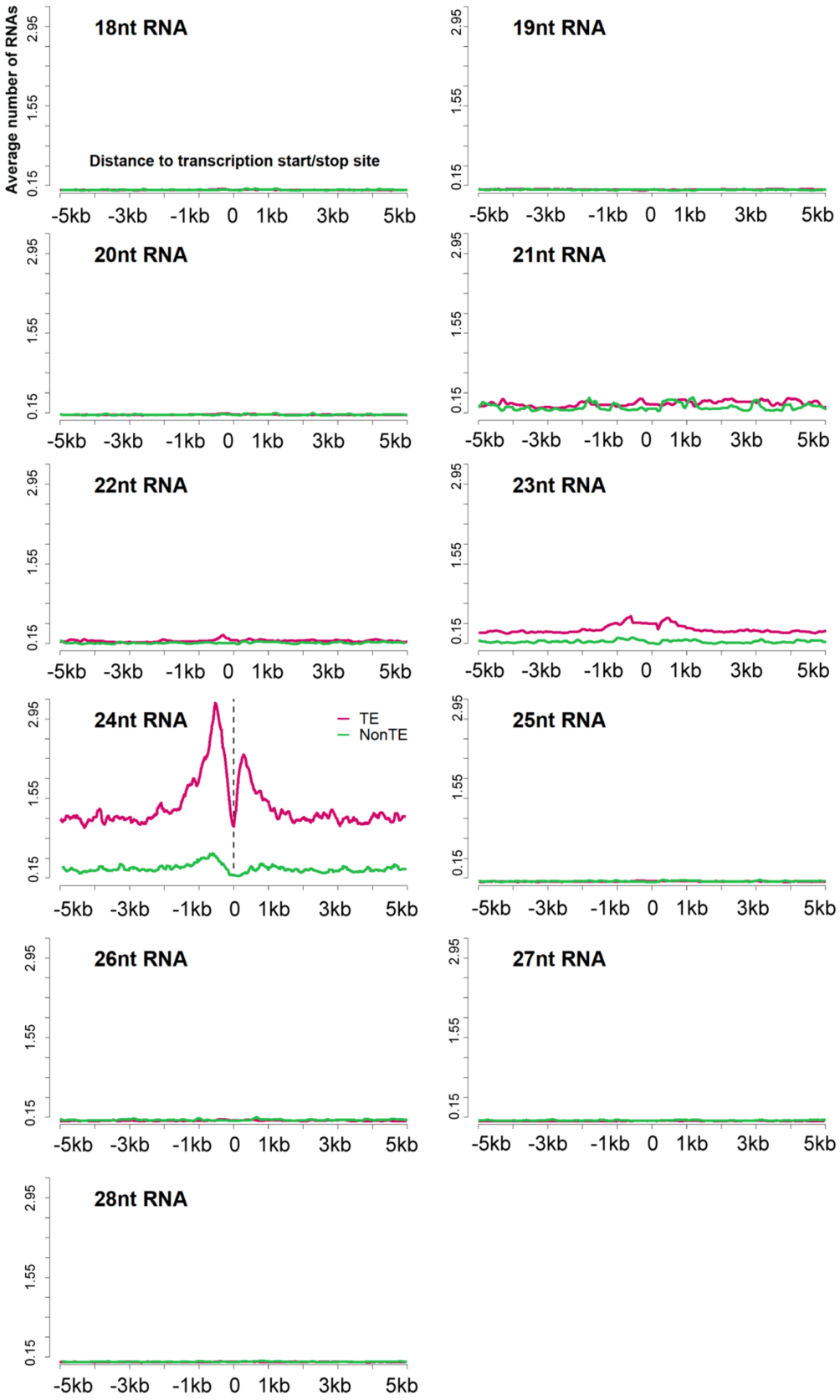
SUPPLEMENTAL FIGURES

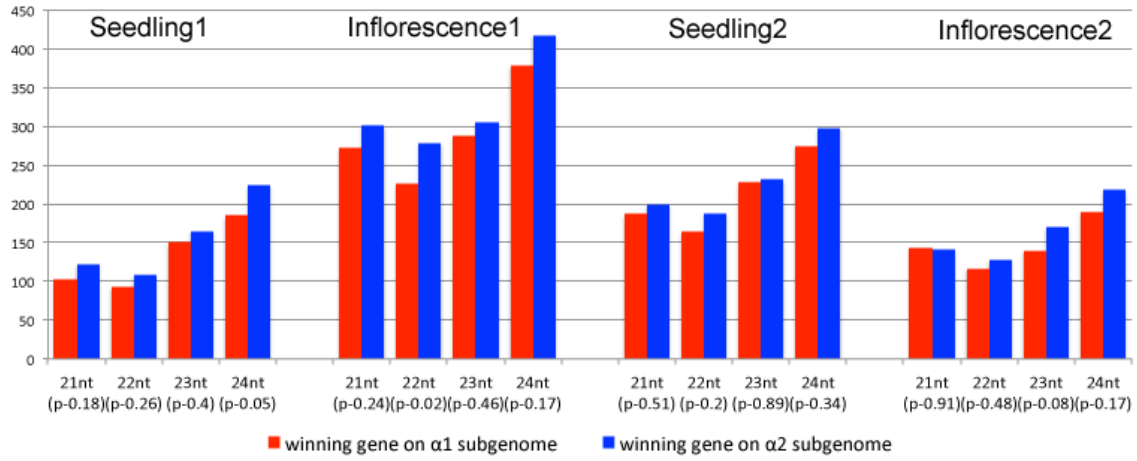


Supplemental Figure 1. Gene dominance. A 2:1 biased plot of mRNA levels, in FPKM (fragments per kilobase exon per million fragments mapped) for one pair of maize homeologous genes in 37 different developmental-endpoint experiments from 8 different laboratories using publicly available reads. Citations to these valuable maize community RNA-seq data are in qTeller-maize: http://qteller.com/qteller3/ (dot plot can be regenerated here: http://qteller.com/qteller3/scatter_plot.php?name1=GRMZM2G057973&name2=GRMZM2G004140&xmax=80&ymax=80&info=).

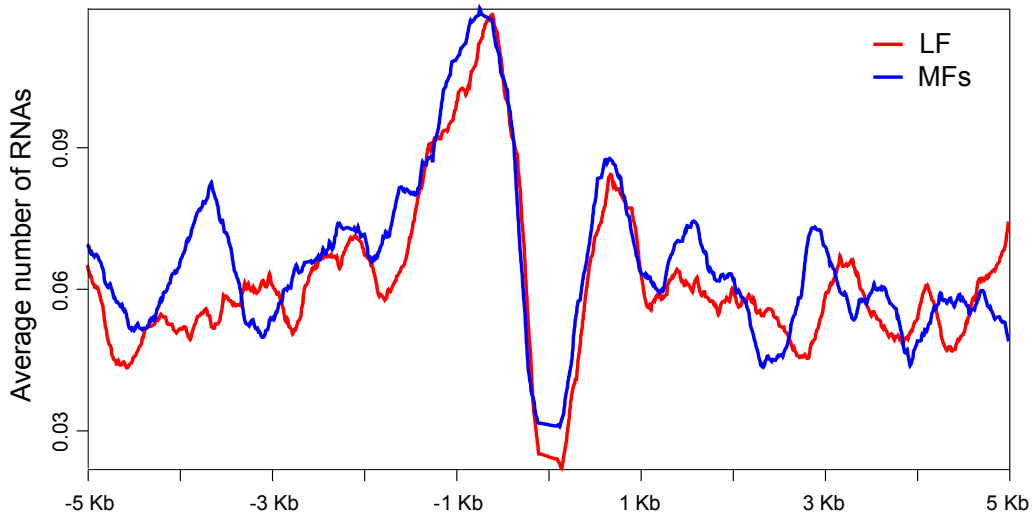


Supplemental Figure 2. The average number of small RNAs with different lengths (18-28nt) mapped to TEs and non TE genomic sequences. The overall distribution of

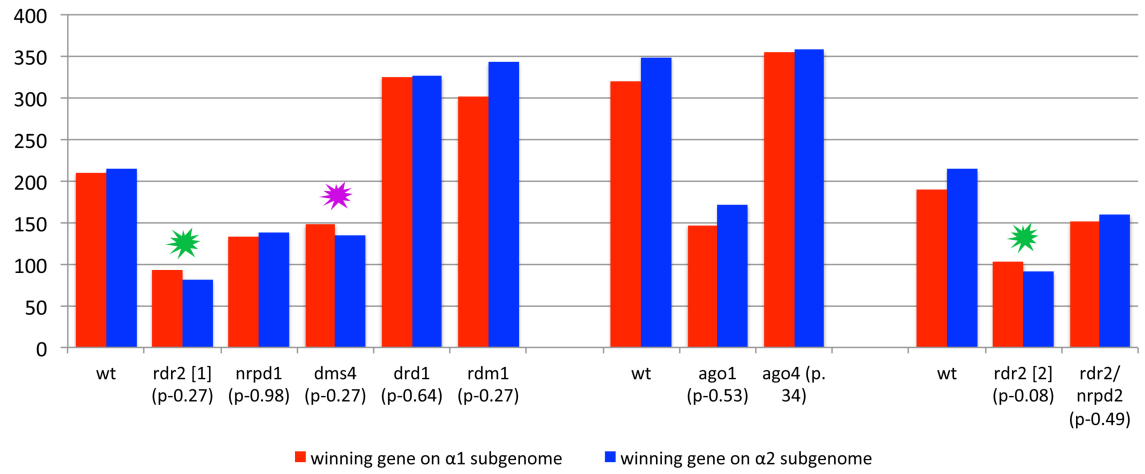
level of targeting of 24nt RNA molecules to flanking sequences of the average *B. rapa* gene, with (green) and without (red) known transposons being hardmasked. Y axis: number of unique-sequence RNAs that mapped perfectly were averaged in a 100bp sliding window moving in 10bp increments through each flanking region of a of *B.rapa* genes, where only unique-sequence molecules were used. All genes have an arabidopsis ortholog. X-axis: kilobase pairs (kb) from the start of transcription, indicated by the dashed line.



Supplemental Figure 3. The 500 bp upstream of the UTR of genes on subgenome 2 tend to have more total (TE-targeted and non-TE targeted) small RNAs bound than do their homeologs in four independent experiments: the homeolog on α2 wins the horserace more often than the gene on α1 repeatedly. The Y axis is the number of horserace wins in units "nucleotides that are hit by smRNAs at least once". The X axis denotes the size class of siRNAs. "p" = p-value by X^2 assuming 1:1 ratio.

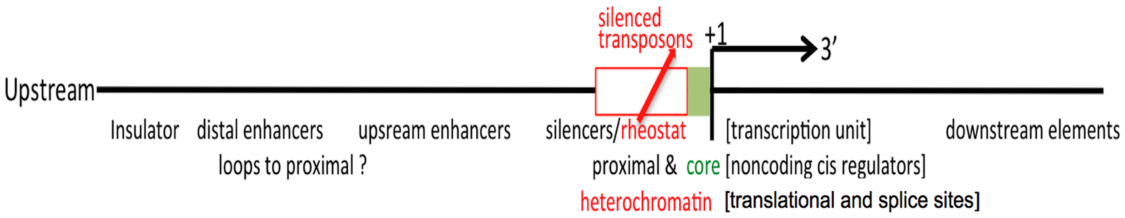


Supplemental Figure 4. Distribution of 24nt small RNAs corresponding to simple repeats around *B. rapa* doublet genes. No bias is seen between LF or MF subgenomes. Y axis: number of unique-sequence 24nt RNAs that mapped perfectly were averaged in a 100bp sliding window moving in10bp increments through each flanking regions of *B.rapa* genes, where only unique-sequence 24 bp molecules were used. See Textbox 1A. Genes from subgenome LF and MFs were calculated separately. All genes have an arabidopsis ortholog. X-axis: kilobasepairs (kb) from the start of transcription, indicated by the dashed line.



Supplemental Figure 5. *rdr2* and *dms4* mutants remove the 24nt small RNA targeting bias from arabidopsis subgenomes. In order to discover what proteins and pathways are involved in the maintenance of 24nt small RNA recessive subgenome bias, we used a series of mutant datasets obtained from Gene Expression Omnibus (GEO accessions in Methods) that relate to the RdDM (RNA-directed DNA methylation) small RNA processing pathway in arabidopsis (26). Our question was simple: do any of these mutants undo the 24nt small RNA bias toward Subgenome 2? We found that only two mutant homozygotes removed 24nt small RNA subgenome bias: mutants of *RNA-DEPENDENT RNA POLYMERASE2* (*RDR2*), and of *DEFECTIVE IN MERISTEM SILENCING 4* (*DMS4*). The mutant *rdr2-1* is a T-DNA insert into the arabidopsis gene *At4g11130*, and specifies a transcript level below the detection limit (27); for this reason, we call *rdr2-1* a "knockout" of function. Total 24nt small RNAs (both TE-and non-TE-targeted combined) are lost in the *rdr2* mutant, such that there are fewer α2 small RNAs in the mutant than in α1. While the X^2 values were not significant, this phenomenon was observed in two independent *rdr2* mutant studies (green stars); its repeatability suggests that the effects of the *rdr2* mutant on 24nt small RNA bias are authentic. DMS4 is a potential regulator of polymerase abundance or activity, that interacts with both Pol II and Pol V in vivo (28). The *dms4-1* homozygous mutant is a loss-of-function mutant due to an G-to-A conversion at the splice-site acceptor of the sixth intron in *At2g30280*. Its phenotype is late flowering and pleiotropic (28, 29). In our study, the *dms4-1* mutant (purple star), showed a similar result to the *rdr2* mutant. Taking these

mutant data together, we suggest that smRNA processing may be involved in maintaining genome dominance, and this certainly fits with our result that either smRNA degree of targeting or smRNA coverage marks genes on the non-dominant, post-polyploid subgenome. Total (TE- and non-TE targeted) 24nt small RNAs for each mutant dataset, 500bp upstream of the 5'UTR. Each wild-type (wt) is specific for the study with which it is aggregated. Y-axis: number of horse race "winners" after the pair-wise comparison of 24bp RNA targeting to the -1 to -500 bp region (Supplemental Information 1). Chi-square probabilities (p) of the observed winning gene numbers being different from the expected WT ratio, our null hypothesis, is shown below each pair of bars.



Supplemental Figure 6. On the basis of data presented, we hypothesize that a regulatory element (in red) must exist along with the canonical elements of a typical differentially expressed eukaryotic gene (2), or at least a higher plant gene. The core promoter (green) is defined as a TATA and an upstream activator sequence (UAS) and that is often close to a "silencer" element, all 100-200 bp upstream of +1, the transcription start site. Our added feature (red), the rheostat, or cluster of transposons, is denoted by the electrical symbol for a variable resistor.

SUPPLEMENTAL TABLES

Table 1A: arabidopsis α pairs TE-targeted 24nt smRNA reads

TE-targeted 24nt smRNA reads	# pairs	p value X2 of win	a1-located gene wins	a2-located gene wins	p by X2 of a 1:1	margin of win (%)
all α-pairs	303	no limit	140	163	0.2	no winner
α1 expressed>α2	169	no limit	71	98	0.04	α2 wins by 28%
α1 expressed>α2	92	<0.05	34	58	0.01	α2 wins by 41%
α1 expressed>α2	55	<0.001	15	40	0	α2 wins by 63%
α2 expressed>α1	133	no limit	68	65	0.8	α1 wins by 5%
α2 expressed>α1	67	<0.05	32	35	0.7	no winner
α2 expressed>α1	40	<0.001	18	22	0.5	no winner

Table 1B: arabidopsis α pairs TE-targeted 24nt smRNA coverage (bp)

TE-targeted 24nt smRNA coverage (bp)	# pairs	p value X2 of win	a1-located gene wins	a2-located gene wins	p by X2 of a 1:1	margin of win (%)
all α-pairs	303	no limit	140	163	0.2	no winner
α1 expressed>α2	169	no limit	71	98	0.04	α2 wins by 28%
α1 expressed>α2	131	<0.05	53	78	0.03	α2 wins by 32%
α1 expressed>α2	104	<0.001	39	65	0.01	α2 wins by 40%
α2 expressed>α1	133	no limit	68	65	0.8	no winner
α2 expressed>α1	101	<0.05	50	51	0.9	no winner
α2 expressed>α1	72	<0.001	37	35	0.8	no winner

Table 1C: arabidopsis α pairs TE coverage (bp)

TE coverage (bp)	# pairs	p value X2 of win	a1-located gene wins	a2-located gene wins	p by X2 of a 1:1	margin of win (%)
all α-pairs	634	no limit	286	348	0.01	α2 wins by 18%
α1 expressed>α2	337	no limit	151	186	0.06	no winner
α1 expressed>α2	250	<0.05	108	142	0.03	α2 wins by 24%
α1 expressed>α2	201	<0.001	85	116	0.0	α2 wins by 27%
α2 expressed>α1	297	no limit	135	162	0.1	no winner
α2 expressed>α1	214	<0.05	93	121	0.06	no winner
α2 expressed>α1	161	<0.001	70	91	0.1	no winner

Supplemental Table 1: Horserace experiments comparing genes per subgenomes as targets for 24nt RNA reads (1A), 24nt RNA coverage in base-pairs (bp) (2B), and TE coverage (bp) only (1C) in *A. thaliana* α pairs. p-value X2 of win means the significance of winners of both FPKM and small RNA horseraces. See supplemental Information 1 and/or Methods for an exact description of the units measured to determine the win.

***B. rapa* TEs 1kb upstream of TSS (units: bp coverage)**	**LF>MF**	**MF>LF**	**total**	**% LF>MF**	**% MF>LF**	**pval**	**Margin %**
all TEs	1963	2254	4217	0.47	0.53	7.42358E-06	13
MITES (Harbinger/TcMariner/Pogo)	240	261	501	0.48	0.52	0.348136085	8
LINES (L1/Penelope)	489	611	1100	0.44	0.56	0.000234666	27
LTR (Copia/Gypsy/Pao)	1596	1946	3542	0.45	0.55	4.08043E-09	22
DNA/hAT	371	423	794	0.47	0.53	0.064977782	13
DNA/En-Spm	168	169	337	0.5	0.5	0.956557944	0
DNA/MuDR	117	157	274	0.43	0.57	0.015671027	33
nonMITE DNA (hAT, En-Spm, MuDR)	583	684	1267	0.46	0.54	0.004547096	17
Unknown	1045	1133	2178	0.48	0.52	0.059346439	8
SINES	224	272	496	0.45	0.55	0.031141211	24
RC/Helitron	55	82	137	0.4	0.6	0.021067887	50
simple repeat	48	46	94	0.51	0.49	0.836568881	4

Supplemental Table 2. The relative genome preference for different *B. rapa* transposon families. Units are in TE bp coverage, 1kb upstream of the TSS. Preference is variable among different TE families.

Supplemental Information 1

The units of 24 bp small RNA targeting

These units are used on the Y-axes of our figures 3 and 4 and for tables 3 and 4, and have specific meanings.

Moving average experiments: For the distribution of small RNAs targeting unique sequence around the start of transcription (TSS) of the composite gene.
Average number of 24 bp smRNA molecules in a 100 bp window that match 100% to the gene—but hit the genome once only--and anywhere in the 100 bp window, advancing over the gene in 10 bp increments. This average-of-all-gene value varies from 0-3 molecules. Larger numbers of particular 24bp molecules are excluded as outliers beyond our standard deviations (see Methods).

Horserace experiments: For homeologous gene FPKM horseraces.
Number of fragments per Kb of smRNA per million mapped reads that locate to the 500 bp stretch from -1 to -500, inclusive, where TSS =0. The homeologous gene with the most smRNA reads wins, even if by 1. Either a "win" or "loss" is recorded. Exclusion of bonanza amplifications is assured because all perfectly stacked reads are reduced to one only. The winning gene is called dominant, and this gene usually resides on the dominant subgenome, but not always.

Supplemental Information 2

One scenario for the origin of genome dominance.

Step 1. **Parents diverge, making their "tradeoffs" independently**. Two species exist with genomes of "identical" gene content, but differ significantly in transposon load. A species may have fewer transposons partly because it more efficiently silences transposons (3). When those silenced transposons are upstream of genes (i.e. in the rheostat position), these genes express less message. Since transposons are distributed throughout the entire genome, the net effect of this process is an average, global repression of mRNA levels. Thus, each species makes a tradeoff between any negative consequences of gene expression silencing and any benefits of transposon silencing. In contrast, the species in which selection against transposon activity is less severe will generally accumulate mRNAs to a higher level, because silencing of its transposons, and the genes just downstream from them, is less efficient. Thus, the genome of this species would be expected to have less transposon silencing, with less of an effect on nearby genes, and more TEs near genes (in the rheostat position).

Step 2. **The wide cross.** Parents cross and a hybrid emerges as a fertile plant. The hybrid is a novel organism. Because of mismatches between the small RNA populations and their targets, regulation of both transposons and nearby genes becomes chaotic, with many transposons and genes escaping RdDM. As a consequence, transposons are awakened in somatic tissues of F1 plant. If the cross is not too wide, the chaotic loss of silencing can lead to luxuriance (the growth manifestation of *heterosis)* due to increased flux in gene dosage limited pathways caused by the release of epigenetic regulation of genes near transposon patches (4).

Step 3. **Allotetraploidy and the origin of genome dominance.** Whole genome doubling sometimes occurs during this temporary heterotic stage, creating a new heterotic allotetraploid. Germline-specific mechanisms operating in the soma and the gametophyte ensure rapid reestablishment of transposon and gene silencing (5, 6) and selection results in a new tradeoff between transposon silencing and gene silencing. In this new nuclear environment, **we predict that the subgenome with the lower average number of TEs in the rheostat position upstream of the gene, as inherited from its parent in Step 2, will have a higher average level of expression vs. it's homeolog tens of millions of years later**. Subgenome dominance is initially then a consequence of epigenetic modifications acting on genetic differences between parental genomes. However, over time, members of the submissive genome accumulate new TE insertions as a consequence of relaxed selection on these genes, resulting in an ongoing over-representation of TEs near these genes, even as the original population of TEs near these genes is lost. As discussed in the text, *inbreeding depression=genome dominance*; both are hypothetical outcomes of RdDM of transposons in the rheostat position that exert negative position effects on adjacent genes. Previous phenomenological work in the 70's by the Drew Schwartz laboratory (7, 8) predicted that a programmable, quantitative, competitive level of gene regulation must exist.

Supplemental Information 3

Sometimes results are such that they may well explain—without experimental confirming data—more than just the questions posed originally. Such is the case with our exploration of the cause of genome dominance. Our results, conclusions and working hypotheses (Textbox 2 and Figure 8 with the rheostat) provide a gratuitous explanation of heterosis—a grail like problem in genetics and agriculture--and also of the C-value paradox, a longstanding problem in genome size evolution. There's so much information necessary to define historically these problems, we now choose to largely ignore the mountains of previous work and reviews. Our work explains these difficult problems, in whole or part, but we do not have data in direct support of our ideas. Thus, the best our work can achieve is beauty, plus a heuristic value perhaps, with regard to our extrapolations beyond our experimental systems. Nevertheless, the following paragraphs put forth solutions to these time-honored questions.

The heterosis problem. Heterosis has one particularly defining characteristic: wide hybrids often express a burst of vigor that cannot be captured during inbreeding, as if the vigor were a property of the hybrid *per se.*

Both the heterosis phenomenon and, as we have explained, genome dominance are consequences of a wide cross. In a previous publication (4) we argued that inbreeding depression, the flip side of heterosis, is the same mechanism as genome dominance. This mechanism is smRNA directed silencing of genes with transposons upstream by position effect. Our argument includes several references to results from those studying heterosis, and finding epigenetic involvements--specifically 24nt RNA involvement. Our conclusions involving genome dominance extrapolate to the heterosis problem. Genome dominance is explained by RdDM of adjacent transposons and downstream position effects. *Heterosis can be seen as the out-of-control period immediately following the wide cross* when genome dominance is not yet established by epigenetic control mechanisms. The result is luxurious flux through normally dosage-limited metabolic pathways. It makes sense the wide hybrid would be able to escape the Gaut lab tradeoffs because the smRNAs elicited to control the transposons in one parent are not expected to work properly on the transposons of both parents, and temporary escape from position-effect down-regulation seems likely. Eventually, as the wide hybrid goes tetraploid and as that tetraploid is inbred, the Gaut-lab tradeoffs eventually fine-tune the genome, and inbreeding depression counters heterosis. Sometimes the sheer beauty of a hypothesis demands that it be put forth, even though it is undoubtedly over-reaching our data. Our solution, where we credit the Gaut lab for giving us the idea of transposon-gene expression tradeoffs, has been rendered as a cartoon with citations (4) following the reasoning that something really important is best approached with humor.

The C-value paradox. The term "C-value paradox" was coined by C.A. Thomas (9). The paradox, exhibited especially in plants, is that haploid genome contents for species in the same families—like different sorts of beans or grasses – differ by several fold and this bulk DNA difference is not accounted for by polyploidization and is not related by any sort of obvious differences in morphological or behavioral complexity. Transposons are known to be a primary sort of DNA that makes up these differences in C-value. Transposons, because they have a selfish life cycle of their own, have been called "junk" in relation to the host organism. This report not only illuminates a bulk regulatory role for transposons, but also specifically implicated upstream transposons in a position effect mechanism to down regulate genes in relation to their homeologous gene or to other genes obviously in the same functional network. Said in another way, transposon bulk balances the expression of homeologous genes—or perhaps alleles, homeologous genes, other paralogous genes and perhaps other genes in the same functional network. This balance of gene expression has been shown to be particularly important to phenotype; the gene balance hypothesis derives from the central premise that haploinsufficiency and triploinsufficiency negative phenotypes are common, and maintenance of gene product balance among genes that act together in multimers, or function together in networks, is accurately preserved by purifying selection. Hollister and Gaut (3) report that tradeoffs between transposon

silencing and average gene expression are made differently in differ C-value species (like *A. thaliana* and *A. lyrata*) and they do this by modifying the strength of silencing. The balance of expression among genes is, according to our scenario (Textbox 2 and Discussion), regulated by the coverage of transposons upstream of genes. This bulk function might be achieved by using a few transposons or by using many; it's *the balance not the total amount of transposons that is under selection*. Total amounts of gene expression are adjusted using different mechanisms, like adjustments to the smRNA silencing network, but no matter whether silencing of transposons is heavy or light, the balances of gene expression remain the same. Clearly, one can imagine a lineage getting stuck in a cycle of transposon blooms where total transposon load increases but purifying selection maintains transposon balances among dose-sensitive genes; the result of this is increase of C-value with the excess DNA bulk functioning. The transposons used to regulate gene balance are "junk" if seen in one way only. If they could be removed simultaneously from two or more loci at once, then balances remain unchanged and the removed DNA might be seen as junk. However, the transposon DNA did function, and simultaneous removals are difficult to accomplish. So, when balance *per se* is the point of purifying selection, and transposon bulk upstream of genes is the mechanism, the essential paradox of the C-value differences is solved: transposon DNA functions as the target of RdDM-mediated silencing, and down regulates genes close by, but the total amount of transposon needed to accomplish this function is of little importance. It's only the *balance* of transposon bulk among networked genes that is under strong purifying selection.

METHODS

Brassica rapa sample preparation and small RNA sequencing

Small RNA was extracted from three organs—root, stem and leaf—from *B. rapa* accession Chiifu greenhouse-grown 5-leaf plants. The three samples were also used for the mRNA-Seq (below). Small RNA libraries were constructed following the manufacturer's instructions (Illumina GAII). Total RNAs were purified and precipitated using the RNA fragmentation kit (Ambion). The precipitated small RNAs were linked by 3' adaptors. After gel extraction, 5' adapters were added to the RNA segments. Then products were amplified by PCR and gel extracted for sequencing. 36bp single-end reads were generated using the Illumina HiSeq™ 2000 platform. We obtained 14.8 to17.7M reads for each of the three RNA samples. These reads have been deposited in http://brassicadb.org/brad/datasets/sRNA-seqLSR/. Raw reads of small RNA sequence with at least 6bp of 3' adaptor sequence were subjected to adaptor trimming. We used the tool "cutadapt" to trim both 5' and 3' adaptors (10). Clean reads were then aligned to the *B. rapa* genome by Bowtie; we kept only the uniquely and perfect mapped reads (11). We often masked the *B. rapa* v. Chiifu reference genome before use. Mask 1: transposon sequences were masked using RepeatMasker (http://www.repeatmasker.org/) Mask 2. Everything except those TEs identified by Mask 1 were masked. Mask 3: sequences appeared more than 50 times in the genome were hardmasked (12). Mask4: specific transposon families, by hard masking all TEs present two or more times in the genome, if they show sequence identity more than 80% with coverage more than 80% Mask5: simple sequences were hardmasked with NCBI WindowMasker (13).

Identification of UTR regions for *B. rapa* gene models

UTR regions were not predicted with the released gene models of *B.rapa* (14). Here we used all of the mRNA-Seq data available to identify the UTR sequences. Seven previously reported mRNA-seq datasets were downloaded from sources described previously (14, 15); there are 228.87M pair-end reads and 21.83Gb of sequences. We first aligned these reads to the genome of *B. rapa* using SOAP (16), allowing at most two mismatches, and only reported the best hit. After the alignment, we extracted the coordinates of the genomic regions mapped by the mRNA reads. We then compared the coordinates of *B. rapa* gene models with that of the mapped genomic regions, extending the 5' and 3' exons of each gene up- and downstream until they no longer corresponded to known translated sequence. These extended regions were considered as the transcribed but untranslated sequences of genes (5' and 3' UTR). For each *B. rapa* transcriptional unit, we extracted 5Kb of sequence upstream of 5'UTR and downstream of 3'UTR. Thus, we extracted 5KB flanking noncoding sequences. For two adjacent genes with an intergenic space less than 10Kb, we then first located the middle position between the two neighbor genes (calculated from the last untranscribed nucleotide of the first gene to the first untranscribed nucleotide of the second gene), the first half part is the one flanking region of first gene, the latter half part is one flanking region of the second gene, both half parts are equal in size and less than 5Kb. These flanking regions were the subject of our investigations on TE distribution and small RNA targeting around *B. rapa* genes.

The average number of small RNAs mapped to the flanking sequences of *B. rapa* genes

We calculated the number of RNAs mapped in the 100bp sliding window moving in 10bp steps into both 5' and 3' flanking regions of each *B. rapa* gene. A "mapped sequence" needed to be a perfect match. RNA reads were sometimes mapped to the *B. rapa* genome after one of the 5 different sorts of hard masking. A masked nucleotide was turn into "n". When data is hardmasked, the positions are left the same, so no data are lost; each point in each line is how many bp from the start of the TSS. The mapping data for each individual gene was then averaged across all *B. rapa* genes (Supplemental Figure 1). In this way, the average gene in any particular genome, masked or unmasked, could be compared. In this way, we found out if transposons or simple sequences comprise small RNA targets. Outliers, removed from the data of Figures 3 and 4 (but not horseraces) were either >1.5x above the 3rd quartile or >1.5x below the 1st quartile (Box Plot parameters, where x is the distance between 1st and 3rd quartiles). This would be approximately 2.7 standard deviations from the mean were the distribution normal.

Subgenome comparison of small RNA targeting to TEs in *B. rapa*

To estimate the mapping level of RNAs to TEs around genes from different subgenomes, we studied the RNA reads uniquely and perfect mapped a *B. rapa* genome where only TEs were left to align (Mask 2). A 100bp sliding window moving in 10bp steps was used to screen across the 5' and 3' flanking regions of *B. rapa* genes. Within each 100bp window, we counted the number of RNA reads mapped to the TE sequences, with "mapped" defined exactly as was done previously. The number of mapped RNAs was then averaged across the subset of *B. rapa* genes from subgenome LF and MF respectively and considered as the RNA mapping levels to TEs in each 100bp sliding window in the flanking regions of these genes (Supplemental Figure 2).

TE distribution around *B. rapa* genes

We also used the 100bp sliding window—with 10bp step moving into the 5' and 3' flanking regions—to estimate the TE density around each *B. rapa* gene. In each 100bp window, we calculated the ratio of TE nucleotides for each gene and then averaged the ratio across subsets of *B. rapa* genes. One subset was genes from subgenome LF, for example, or those genes from either of the subgenomes MF that were being dominated in a doublet, for another example. The averaged ratio was determined as the TE density in the flanking region of these subsets of *B. rapa* genes. TE annotation was that used in the *B. rapa* var. Chiifu release paper (14).

***B. rapa* mRNA levels**

The three mRNA-seq seedling datasets—root, stem and leaf (the same samples from which the small RNA-seq datasets were generated)—were generated as part of a previous study (15). Reads were mapped onto *B. rapa* v. Chiifu using the qTeller method described in http://qteller.com/arabidopsis/faq.php using GSNAP (17), and expression levels of individual genes were quantified using Cufflinks (18).

Arabidopsis Small RNA-seq analysis
For all small RNA-seq studies, reads were mapped using the Bowtie program (19). Only unique and perfectly mapped reads were included; no mismatches were allowed). Arabidopsis small RNA reads were mapped to unmasked TAIR10 using Bowtie, as described above. Arabidopsis raw reads were downloaded from NCBI's Sequence Read Archive. For TE-derived small RNA analysis, the TAIR10 genome was masked by RepeatMasker for all except transposon sequence, then the small RNA reads were mapped using Bowtie as described. The following datasets were used: GEO accessions GSE32284 (Ma Lab, National Institute of Biological Sciences, Beijing), GSE36424 (Meyers Lab, University of Delaware), GSE34207 (Meyers Lab, University of Delaware), GSE28591 (Wang Lab, Rockefeller University).

Arabidopsis TE analysis
TAIR10 transposable element fasta sequence was obtained from arabidopsis.org (ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release/) and these fasta sequences were used as the library against which RepeatMasker identified TE coordinates in the Arabidopsis TAIR10 genome. Identifying TE-specific small RNAs, we used the masking scheme described above for *B. rapa*. We used the coverageBed script from BEDTools (20) to find TE coverage 500bp upstream of the 5'UTR for the genes in this study.

Arabidopsis mRNA expression
Arabidopsis mRNA-seq expression reads were mapped onto TAIR10 using qTeller methods http://qteller.com/arabidopsis/faq.php. Arabidopsis raw reads were downloaded from NCBI's Sequence Read Archive. The following datasets were used: GEO accessions GSE37644, GSE38286, GSE38286 (Jacobsen Lab, UCLA), GSM799868 (Ma Lab, National Institute of Biological Sciences, Beijing).

Supplemental Data References

1. Pikaard JRHCS (2011) Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. *Nature Reviews Molecular Cell Biology* (12):483-492.
2. Levine M, and Tijan, R. (2003) Transcription regulation and animal diversity. *Nature* 424:147-151.
3. Hollister JD & Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* 19(8):1419-1428.
4. Freeling M*, et al.* (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Current opinion in plant biology* 15:131-139.
5. Slotkin RK*, et al.* (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136(3):461-472.
6. Li H, Freeling M, & Lisch D (2010) Epigenetic reprogramming during vegetative phase change in maize. *Proceedings of the National Academy of Sciences of the United States of America* 107(51):22184-22189.

7. Schwartz D (1971) Genetic control of alcohol dehydrogenase--a competition model for regulation of gene action. *Genetics* 67:411-425.
8. Freeling M (1975) Further studies on the balance between Adh1 and Adh2 in maize: gene competitive programs. *Genetics* 81:641-654.
9. Thomas CA, Jr. (1971) The genetic organization of chromosomes. . *Annu Rev Genet* (5):237-256.
10. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:10-12.
11. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
12. Lyons E*, et al.* (2008) Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol* 148(4):1772-1781.
13. Morgulis A, Gertz EM, Schaffer AA, & Agarwala R (2006) WindowMasker: window-based masker for sequenced genomes. *Bioinformatics* 22(2):134-141.
14. Wang X*, et al.* (2011) The genome of the mesopolyploid crop species Brassica rapa. *Nat Genet* 43(10):1035-1039.
15. Cheng F*, et al.* (2012) Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One* 10.1371/journal.pone.0036442.
16. Li R, Li Y, Kristiansen K, & Wang J (2008) SOAP: short oligonucleotide alignment program. *Bioinformatics* 24(5):713-714.
17. Wu TDaW, C.K. (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. . *Bioinformatics* 21:1859-1875.
18. Roberts A, Pimentel, H., Trapnell, C., Pachter, L. (2011) Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* (10.1093/bioinformatics/btr355).
19. Langmead B, Trapnell, C., Pop, M., Salzberg, S.L.. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. . *Genome Biol* 10(25).
20. Quinlan ARH, I.R. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841-842.