Supporting Information

Nobuta *et al.* **10.1073/pnas.0808066105**

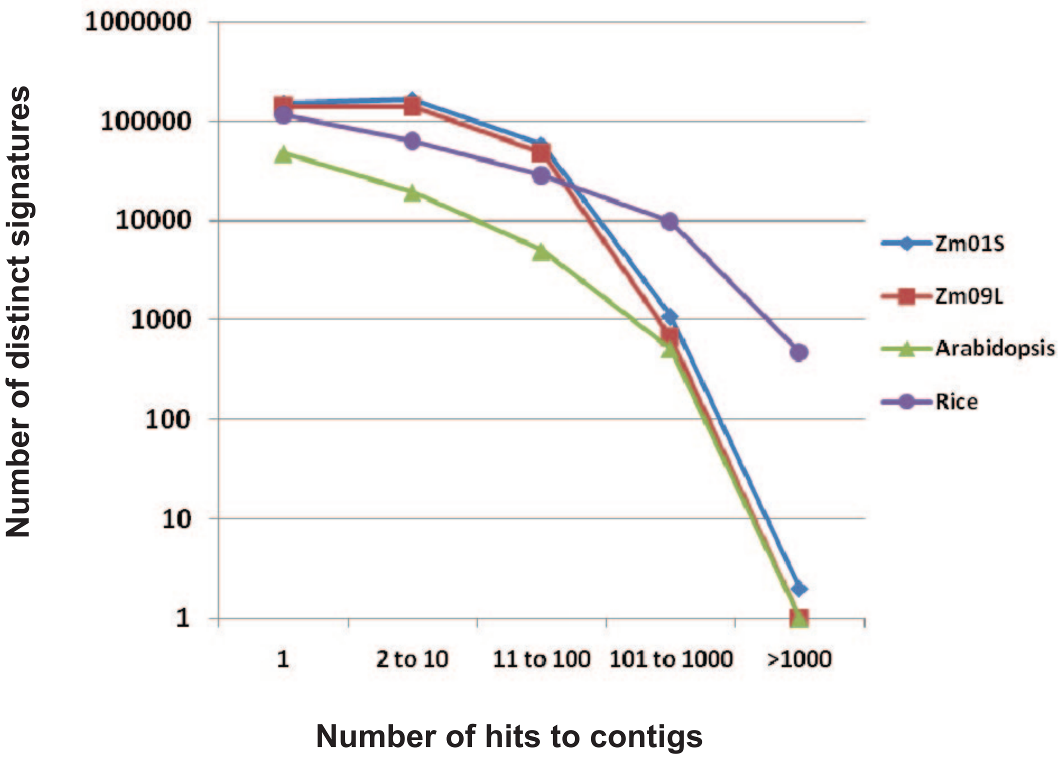
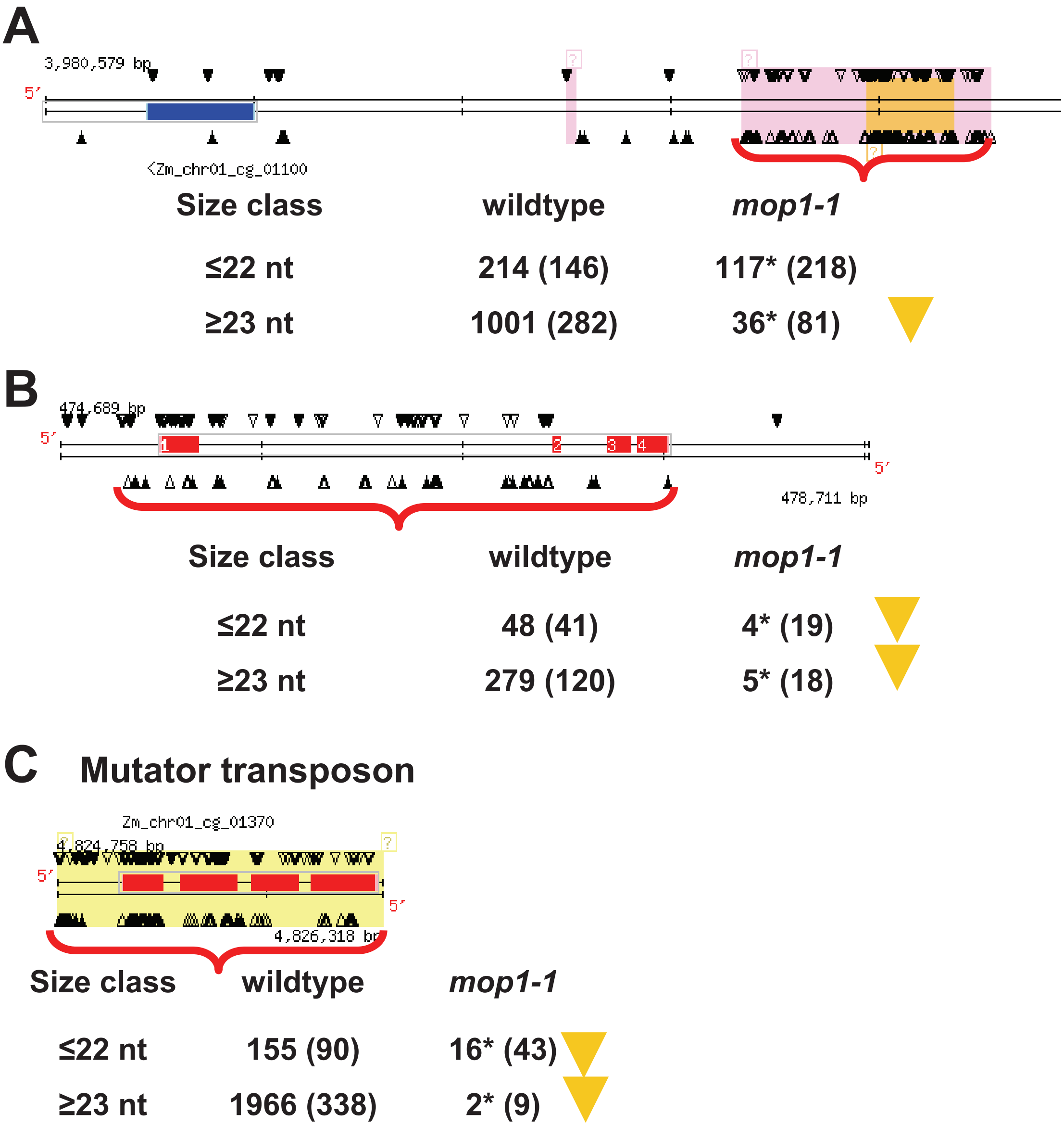


Fig. S1. Match frequencies of small RNAs in maize versus those of rice and *Arabidopsis*. We determined the number of genomic matches (''hits'') to the two maize contigs and compared these values with those for 149,978 and 56,920 distinct small RNA MPSS signatures from rice and *Arabidopsis* inflorescence libraries. The small RNAs matched to the maize contigs are highly duplicated, even within 14 Mbp, compared with *Arabidopsis* and rice, as indicated by the larger number of small RNAs with 2–10 or 11–100 matches than those with just 1 match. The *y* axis is shown on a logarithmic scale. The *Arabidopsis* and rice data have been described previously (1, 2).

1. Nobuta K, *et al.* (2007) An expression atlas of rice mRNA and small RNA. *Nat Biotechnol* 25:473–477.
2. Henderson IR, *et al*. (2006) Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing, and DNA methylation patterning. *Nat Genet* 38:721–725.



A

Size class	wildtype	*mop1-1*	
≤22 nt	214 (146)	117* (218)	
≥23 nt	1001 (282)	36* (81)	▼

B

Size class	wildtype	*mop1-1*	
≤22 nt	48 (41)	4* (19)	▼
≥23 nt	279 (120)	5* (18)	▼

C **Mutator transposon**

Size class	wildtype	*mop1-1*	
≤22 nt	155 (90)	16* (43)	▼
≥23 nt	1966 (338)	2* (9)	▼

* The sum of abundance was normalized
by miRNA levels as described in the text.

Fig. S2. *mop1-1* shows reduced siRNAs of all sizes in lower copy regions. Examples of regions from the Chr. 1 maize contig show small RNAs from both the wild-type and *mop1-1* datasets mapped to specific genomic sequences indicated by the red curly braces. The images are screen shots from our small RNA viewer that indicates the location that was used in the analysis; the blue or red boxes are annotated exons, the black triangles are small RNAs (filled have only one match in the Chr. 1/9 contigs), the pink shading is a retrotransposon-like sequence, the orange shading is an inverted repeat, the yellow shading is a *Mutator* transposon. Most small RNAs in this region had one to three hits. Small RNA abundances and counts of total distinct tags for each displayed image are indicated in the table below each screen shot. The first value is the sum of abundance, and the number of corresponding distinct tags is shown in parentheses. The yellow triangle indicates a substantial reduction in *mop1-1* compared with wild type.

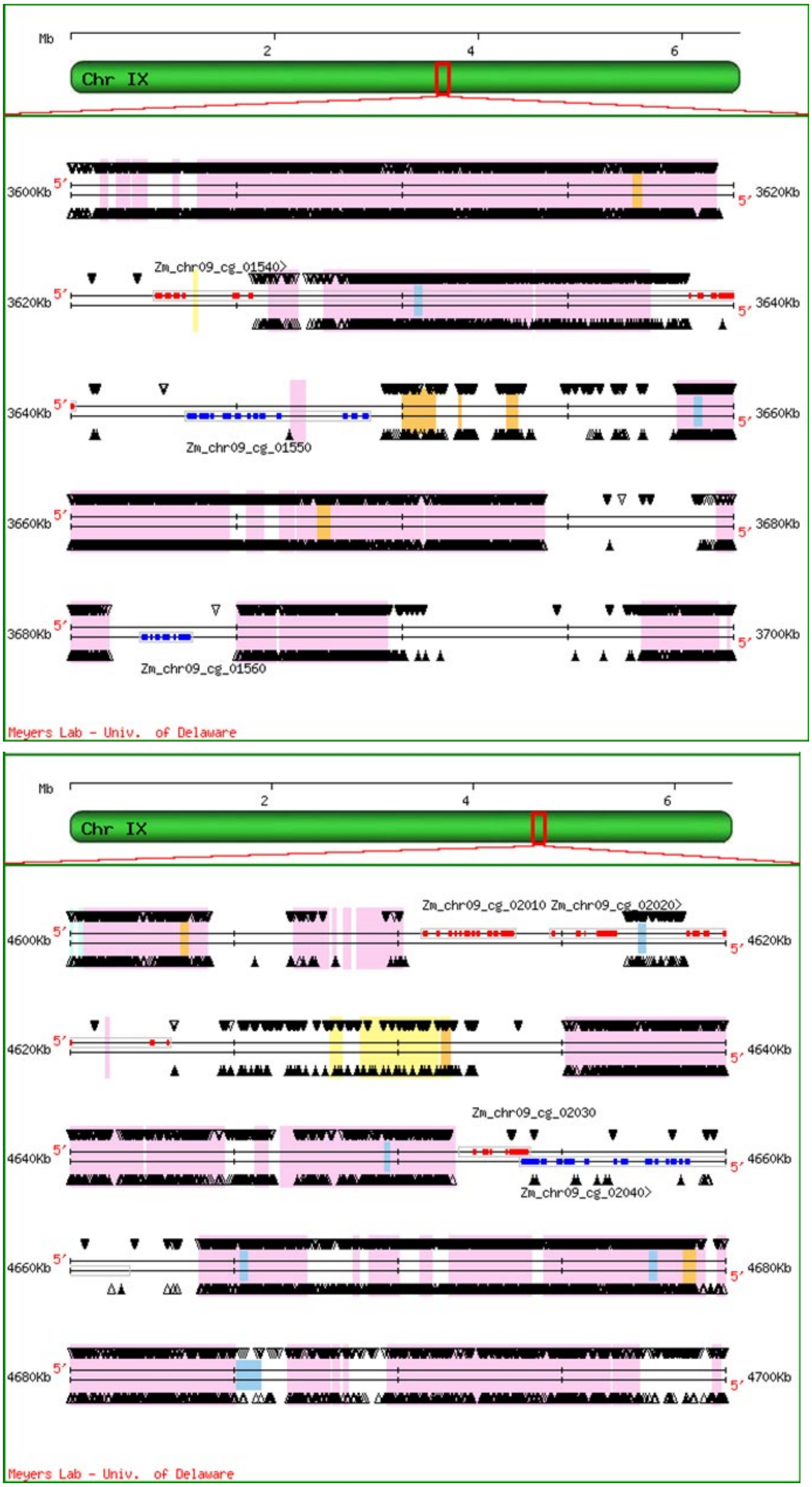


Fig. S3. Small RNA islands interspersed between annotated genes. Example regions, indicated by red squares, from the Chr. 9 maize contig, showing small RNAs from both the wild-type and *mop1-1* datasets mapped to the genomic sequence. The images are the screen shots from our maize small RNA viewer, and the descriptions of the symbols and colors are as described in Fig. S2, except for the red or blue boxes (exons, top or bottom strands, respectively), yellow shading (transposon), and blue shading (tandem repeats).

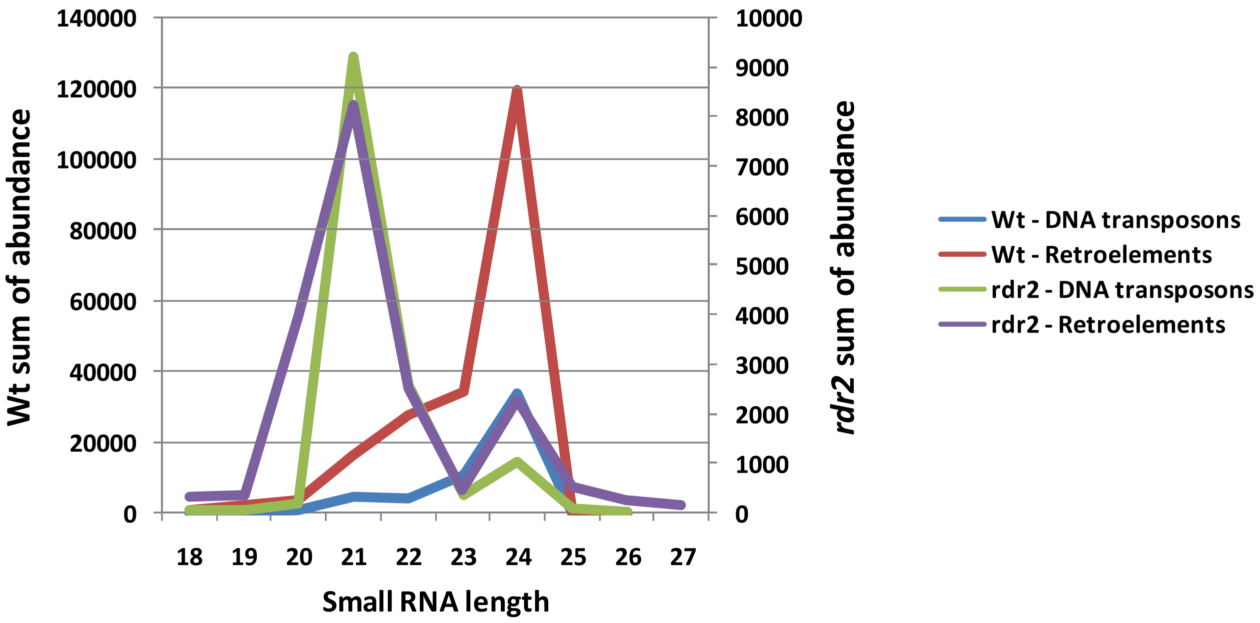


Fig. S4. Small RNA sizes from wild-type and *rdr2 Arabidopsis* matching retrotransposons and DNA transposons. The size distribution of *Arabidopsis* small RNAs mapping to LTR retrotransposons (predominantly high-copy elements) or DNA transposons (predominantly low-copy elements) in the *Arabidopsis* genome is shown. The *y* axis indicates the sum of the abundance for the set of distinct small RNAs matching these regions. Note that wild-type *Arabidopsis* is plotted on the left axis, and *rdr2* on the right axis using different scales because of the dramatic difference in their abundances.

Table S1. Summary statistics of SBS libraries of maize immature ear

Group	Library	SBS reads[*]	Distinct signatures[†]	Proportion distinct[‡]
1	Wildtype	5,640,318	3,640,512	0.65
2	*mop1-1*	7,249,635	2,405,543	0.34
	Totals[§]	12,889,953	5,135,191[¶]	
	Intersection of groups 1 and 2		42,944	

[*] Reads $\geq$15 bp in length (after trimming of the 3′ adapter) from the sum of two (wild-type) or three (*mop1-1*) SBS flow cells or sequencing reactions.

[†] Number of different sequences found within the set.

[‡] Number of distinct signatures divided by the total sequenced.

[§] Sums of the different libraries.

[¶] Union of groups 1 and 2.

Table S2. Summary statistics of SBS libraries of immature *Arabidopsis* flowers

Group	Library	SBS reads[*]	Distinct signatures[†]	Proportion distinct[‡]	SBS reads, genome matched[§]	Distinct signatures, genome matched[§]
3	Wildtype	1,674,296	974,050	0.58	1,041,431	493,633
4	*rdr2*	4,720,974	369,816	0.08	2,232,154	115,419
	Totals[¶]	6,395,270	1,315,980[‖]			
Intersection of groups 3 and 4			27,886			

[*] Reads ≥15 bp in length (after trimming of the 3′ adapter) from the sum of two (wild-type) or three (*mop1-1*) SBS flow cells or sequencing reactions.

[†] Number of different sequences found within the set.

[‡] Number of distinct signatures divided by the total sequenced.

[§] These values were calculated after matching to the *Arabidopsis* genomic sequence (TAIR v7). These values exclude sequences matching t/r/sn/snoRNAs which numbered 1,651,983 total reads (35,025 distinct sequences) in the *rdr2* library, and 82,830 total reads (18,493 distinct sequences) in the wild-type library.

[¶] Sums of the different libraries.

[‖] Union of groups 3 and 4.

Table S3. Small RNAs from wild-type and *mop1-1* maize matched to two large maize contigs

		Wild type		*mop1-1*	
Maize genome contig*	Total contig size, bp	No. distinct small RNAs	Total hits†	No. distinct small RNAs	Total hits†
Zm01S	7,822,695	186,074	1,233,770	239,567	1,901,466
Zm09L	6,560,930	171,432	994,690	205,990	1,502,480
Total	14,383,625				

*Maize contigs from Bruggman *et al*. [Bruggman R, *et al*. (2006) Uneven chromosome contraction and expansion in the maize genome. *Genome Res* 16:1241–1251].

†Number of locations at which small RNAs were matched within each contig.

Table S4. Small RNAs from SBS libraries matching different types of genomic sequences

	No. of distinct signatures			Sum of abundance*		
Type	Wild type	*mop1-1*	*mop1-1*/WT ratio	Wild type	*mop1-1*	*mop1-1*/WT ratio
≤20 nt						
Gene	1,101	1,575	1.4	421	465	1.1
Intergenic regions	15,136	25,793	1.7	17,164	29,401	1.7
Tandem repeats	1,915	4,071	2.1	478	906	1.9
Inverted repeats	4,735	8,373	1.8	1,835	2,057	1.1
Retrotransposons†	11,486	20,470	1.8	11,195	20,385	1.8
Transposon†	626	1,213	1.9	187	168	0.9
		Average‡=	1.8			1.4
21 nt						
Gene	2,225	4,562	2.1	1,554	2,923	1.9
Intergenic regions	33,914	82,441	2.4	61,731	202,467	3.3
Tandem repeats	3,974	10,303	2.6	1,278	5,086	4.0
Inverted repeats	9,864	24,218	2.5	4,052	12,849	3.2
Retrotransposons†	26,598	69,286	2.6	34,603	128,399	3.7
Transposon†	1,284	3,064	2.4	404	1,026	2.5
		Average‡=	2.4			3.1
22 nt						
Gene	546	1,606	2.9	263	645	2.5
Intergenic regions	6,902	28,870	4.2	8,312	43,718	5.3
Tandem repeats	839	4,149	4.9	210	1,360	6.5
Inverted repeats	2,020	8,947	4.4	625	2,837	4.5
Retrotransposons†	5,424	24,128	4.4	5,787	30,319	5.2
Transposon†	266	1,116	4.2	58	194	3.3
		Average‡=	4.2			4.6
23 nt						
Gene	238	157	0.7	126	42	0.3
Intergenic regions	3,579	2,452	0.7	4,212	2,617	0.6
Tandem repeats	394	377	1.0	111	86	0.8
Inverted repeats	1,109	801	0.7	481	171	0.4
Retrotransposons†	2,666	1,970	0.7	2,591	1,901	0.7
Transposon†	138	84	0.6	49	10	0.2
		Average‡=	0.7			0.5
24 nt						
Gene	5,181	2,812	0.5	6,128	1,405	0.2
Intergenic regions	113,262	56,253	0.5	201,524	76,956	0.4
Tandem repeats	9,829	6,857	0.7	5,895	2,181	0.4
Inverted repeats	28,420	17,265	0.6	25,291	6,413	0.3
Retrotransposons†	87,612	47,476	0.5	132,467	59,235	0.4
Transposon†	3,337	2,041	0.6	2,570	605	0.2
		Average‡=	0.6			0.3
≥25 nt						
Gene	257	129	0.5	205	49	0.2
Intergenic regions	5,297	3,187	0.6	6,174	3,662	0.6
Tandem repeats	533	431	0.8	248	177	0.7
Inverted repeats	1,512	1,036	0.7	688	332	0.5
Retrotransposons†	4,082	2,597	0.6	4,184	2,693	0.6
Transposon†	187	124	0.7	80	26	0.3
		Average‡=	0.7			0.5

*Sum of hits-normalized abundances for all locations of all matching signatures. These values were not normalized based on the miRNA levels.

†Numbers of retrotransposons and transposons can include sequences annotated as genes as well as those intergenic regions identified as retrotransposons and transposons by low stringency analysis with RepeatMasker.

‡Note that the average is not weighted for the relative proportion of sequences in each category.

Table S5. Ratios of genomic matches (hits) for small RNAs matched to maize contigs from wild type and *mop1-1*

Ratio of hits*	WT ≤22 nt	*mop1-1* ≤22 nt	WT ≥23 nt	*mop1-1* ≥23 nt
1 vs. (11–100)	2.201	1.702	3.556	1.774
(1–5) vs. (11–100)	4.109	3.133	6.168	3.305
(1–10) vs. (11–100)	4.591	3.575	6.766	3.779

*The ratio of small RNAs was determined for small RNAs of each size class, comparing the number of those with low hits (<10 genomic matches in the contigs) with those with high hits (>10 genomic matches in the contigs). The data for these calculations are plotted in Fig. 3C.