Supporting Information

Yang and Bennetzen 10.1073/pnas.0905563106

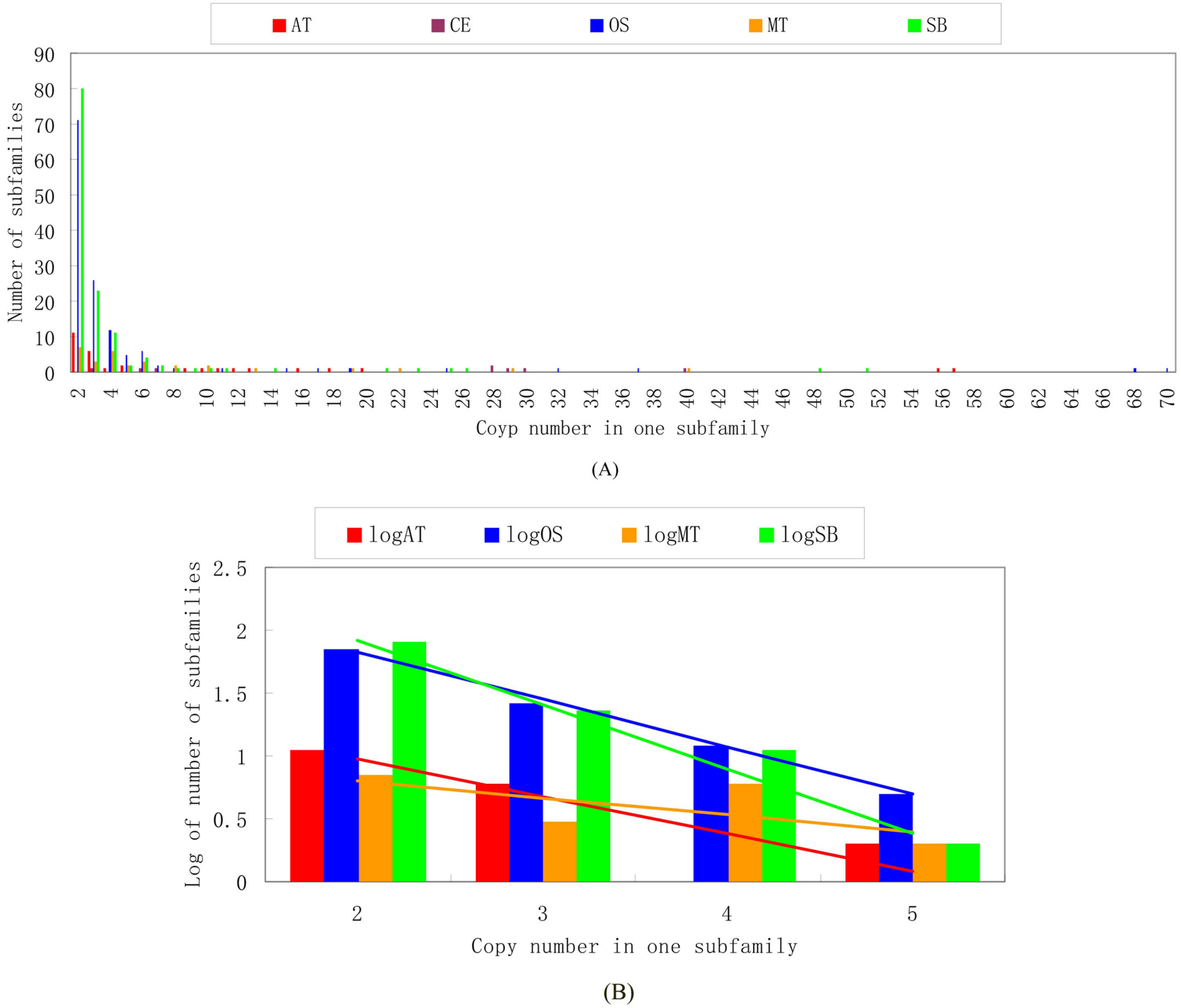

Fig. S1. Distribution of *Helitron* copy numbers in different subfamilies. (*A*) *x* axis, *Helitron* copy number in each subfamily; *y* axis, number of subfamilies with this element copy number. *Helitron* copy number results are shown for *A. thaliana*, *C. elegans*, *O. sativa*, *M. truncatula*, and *S. bicolor*. (*B*) *x* axis, *Helitron* copy numbers in each subfamily (2–5 copies shown); *y* axis, log transformed number (base 10) of the number of subfamilies with this copy number. Linear regression lines are also shown and used to estimate the number of single copy elements.

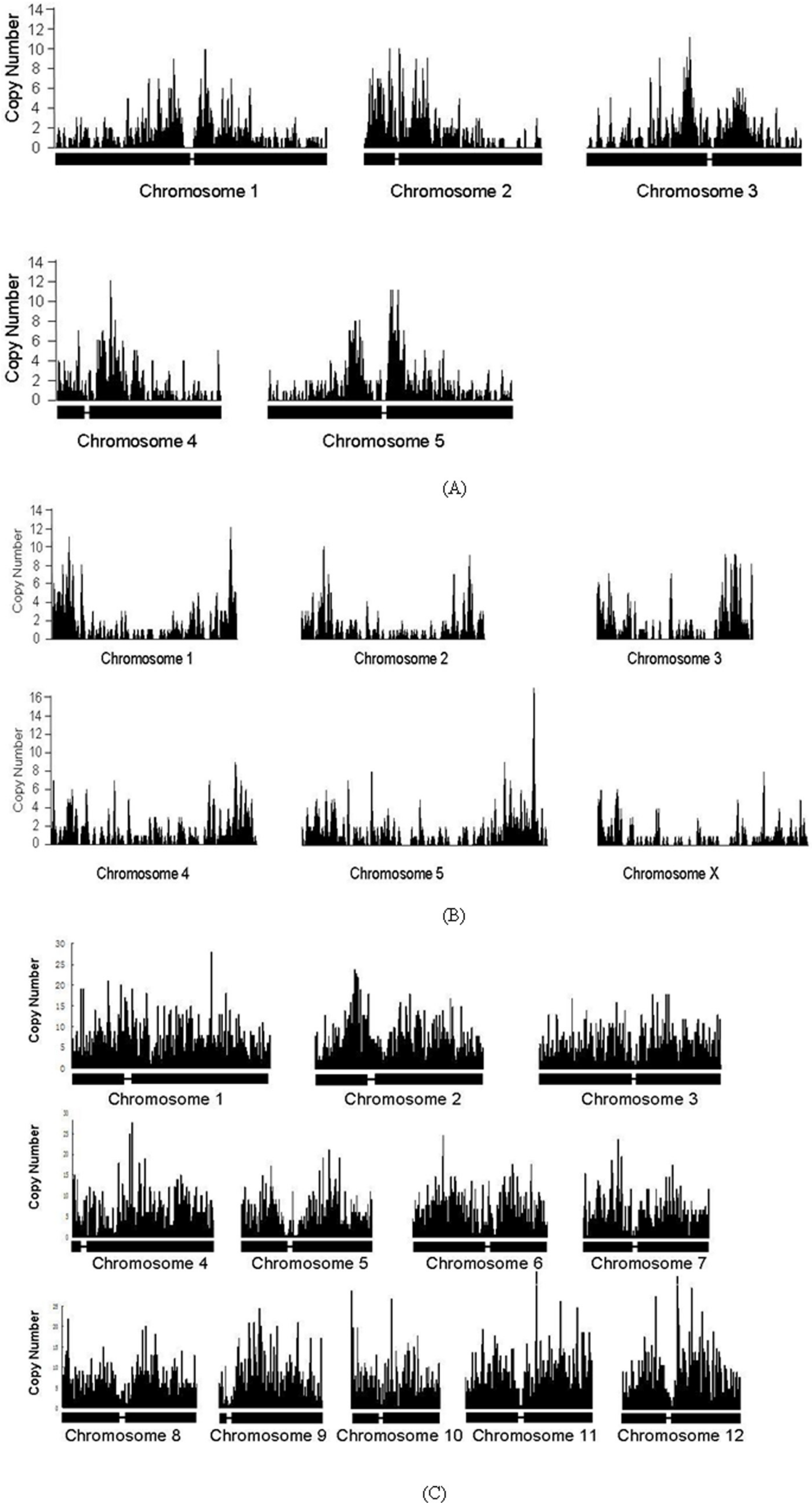

Fig. S2. Distribution of *Helitrons* along the chromosomes of *A. thaliana* (*A*), *C. elegans* (*B*), and *O. sativa* (*C*). The *x* axis shows the chromosome locations and the *y* axis indicates the copy numbers of *Helitrons* (both intact and truncated elements) per 100 kb. Narrow lines in *A* and *C* on the *x* axis indicate the positions of centromeres.

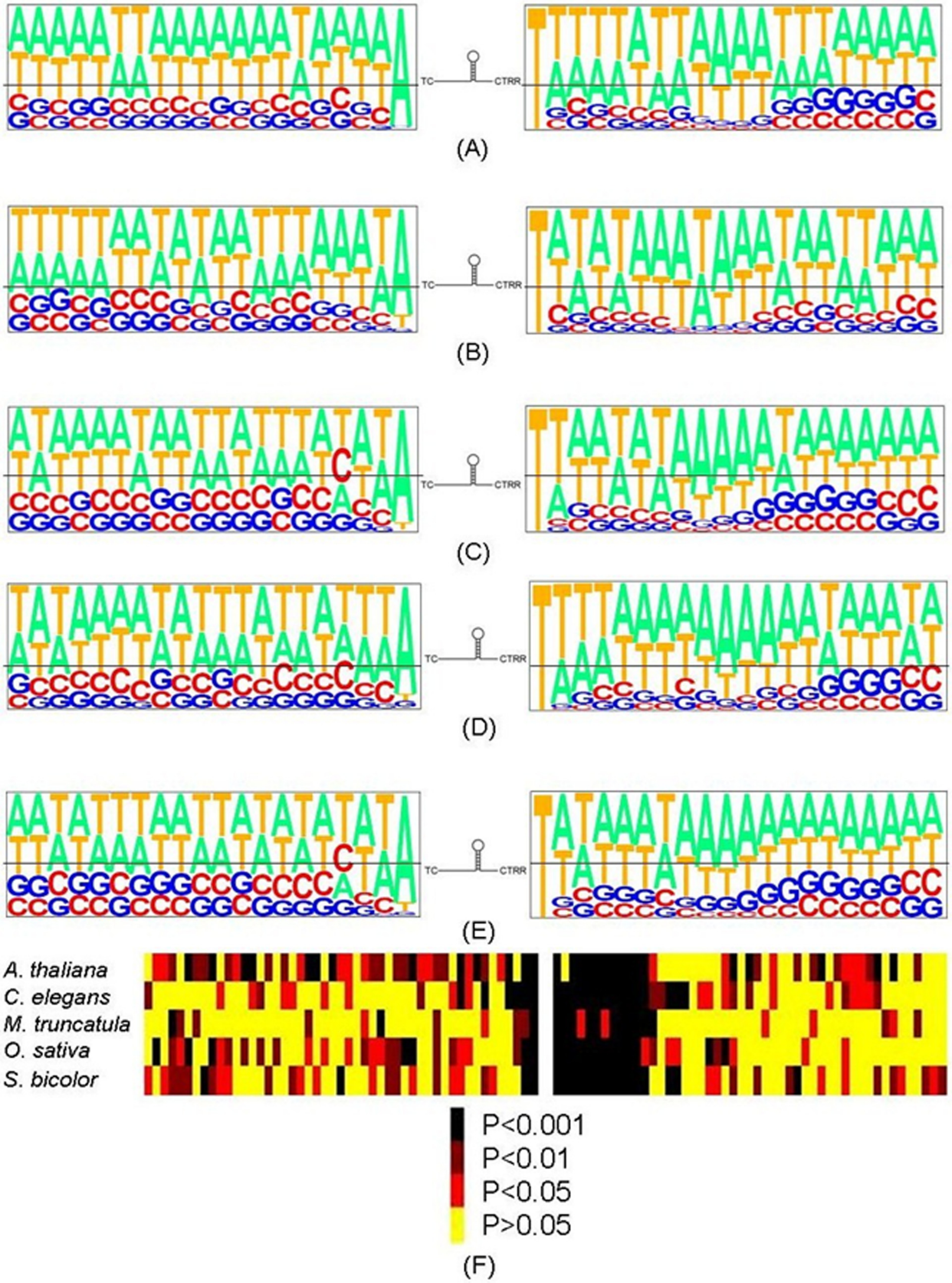

Fig. S3. Insertion site base composition. *Left* side of *Helitrons*, 20 bp upstream of *Helitron* insertion sites. *Right* side of *Helitrons*, 20 bp downstream of *Helitron* insertion sites. Black lines in *A*–*E* indicate average GC content of the genome. (*A*) *Helitrons* of *A. thaliana*. (*B*) *Helitrons* of *C. elegans*. (*C*) *Helitrons* of *O. sativa*. (*D*) *Helitrons* of *M. truncatula*. (*E*) *Helitrons* of *S. bicolor*. (*F*) χ^2 tests of GC content for 50 bp upstream and downstream of *Helitron* insertion sites on each position compared to a random AT flanking site from the same genome. The *P* value is indicated by color, with black representing the most significant and yellow the least significant associations.

Table S1. *Helitron* families in *A. thaliana*

Family name	No. of intact element	Size range, bp	Original name(s) of known *Helitron(s)*	Putative autonomous elements	Gene fragment(s)
Hug	195	360–2,734	ATREP1, ATREP2, ATREP3, ATREP5, ATREP11, AtREP21, HELITRONY1A		+
Hali	27	775–1,709	ATREP6, ATREP9		
Hari	22	642–1,500	ATREP10		
Hane	15	1,346–2,111	AtREP20	+	
Hani	11	606–655	ATREP13		
Hup	3	2,106–2,544	ATREP4		
Hame	3	1,340–1,628	ATREP12		
Hip	2	1,189–1,219			
Hake	2	3,389–3,395			+
Haki	1	817–903	ATREP7		
Total	281	360–3,395			

Table S2. ***Helitron*** **families in** ***C. elegans***

Family name	No. of intact element	Size range, bp	Original name(s) of known *Helitron(s)*	Putative autonomous elements	Gene fragment(s)
Hike	206	182–4,802	HELITRON2_CE, HELITRONY2_CE, HELITRONY3_CE, HELITRONY4_CE	+	
Hibe	38	866–8,965	HELITRON1_CE, HELITRONY1_CE	+	
Hino	30	1,313–1,605	NDNAX3_CE		
Hite	7	1,709–2,164	NDNAX1_CE		
Total	281	182–8,965			

Table S3. ***Helitron*** **families in** ***M. truncatula***

Family name	No. of intact element	Size range, bp	Original name(s) of known *Helitron(s)*	Putative autonomous elements	Gene fragment(s)
Hip	99	1,281–5,357	HELITRON1N_MT	+	+
Hem	53	375–3,249			
Heno	33	2,445–5,717	HELMET	+	+
Heto	33	3,055–5,643			+
Hefo	3	8,279–8,394	HELMET2	+	
Heke	2	450–472			
Hebo	2	817			
Hecu	2	1,910–1,913			
Heku	2	3,043			
Hecho	1	14,261	HELITRON1MT	+	
Total	230	375–14,261			

Table S4. *Helitron* families in *O. sativa* ssp. *japonica* cultivar Nipponbare

Family name	No. of intact element	Size range, bp	Original name(s) of known *Helitron(s)*	Putative autonomous elements	Gene fragment(s)
Hup	280	230–15,654		+	+
Hip	196	242–4,984			+
Hair	116	180–2,577			+
Hole	18	2,357–4,705			
Hoy	10	350–10,152	HELITRON7_OS	+	
Hocku	6	1,058–6,263			+
Hova	5	3,776–13,092	HELITRON2_OS, HELITRON3_OS	+	
Holie	4	705–1,911			
Horia	4	2,635–3,198			
Howi	3	1,896–1,955			
Hota	2	2,085–2,133			+
Hoku	2	1,429–1,695			
Honu	2	210–219			
Hok	2	187–189			
Hani	2	1,140–1,143			
Hoju	2	2,180–2,233			
Holda	2	15,063–15,147	HELITRON1_OS, HELITRON5_OS, HELITRON8_OS	+	
Holku	2	1,891–1,977			
Homno	2	3,532–3,790			+
Holta	2	1,172–1,422			
Holvo	2	6,117–6,186			+
Howa	2	1,503–1,506			
Hofo	2	911–912			
Total	668	180–15,147			

Table S5. ***Helitron*** **families in** ***S. bicolor***

Family name	No. of intact element	Size range, bp	Original name(s) of known *Helitron(s)*	Putative autonomous elements	Gene fragment(s)
Hip	289	289–21,899		+	+
Hair	153	238–532			
Hole	119	1,248–4,499			+
Hok	22	356–3,005			
Hoy	9	986–11,062			
Huyo	6	268–384			
Hubi	2	148–149			
Husi	2	3,423–3,426			
Huca	2	172–174			
Huphi	2	1,718–1,903			
Huga	2	2,435–2,568			
Total	608	148–21,899			

Table S6. Gene fragments acquired by *Helitrons*

Name	Size	Gene fragments	Expect value	Identity, %
Hug_AT2_1	2,182	Putative protein	e-35	94
		Protein in rice	e-14	46
Hake_AT1_1	3,389	Transferase	e-24	93
Hug_AT4_75	747	Hypothetical protein (yeast)	e-14	51
Hup_OS3_1	3,410	RNA polymerase beta chain	e-42	84
Hota_OS1_1	2,083	Unknown	e-40	94
Hocku_OS1_1	6,261	FAD-binding domain-containing protein-like	e-79	81
		Unknown	e-16	90
Hup_OS1_3	2,467	Calmodulin	e-17	80
Hup_OS1_56	2,830	RNA polymerase beta subunit	e-9	51
Hup_OS1_65	2,359	Calmodulin	e-8	43
Hup_OS1_157	2,687	ARM repeat fold domain	e-41	80
Hair_OS8_1	2,372	Unknown	e-10	57
Hair_OS9_1	1,927	Unknown	e-30	89
		Unknown	e-30	87
Hair_OS10_1	373	Unknown	e-10	73
Hip_OS5_1	4,313	Unknown	e-11	43
Hip_OS1_1	1,404	Root cap protein 1-like	e-31	82
Hip_OS7_1	4,908	Unknown	e-22	78
		Unknown	e-11	53
Hip_OS7_4	4,140	Unknown	e-36	92
		Unknown	e-18	97
Homno_OS1_1	3,530	Unknown	e-71	96
Holvo_OS1_1	6,184	Putative RING zinc finger protein	e-35	65
Hip_MT1_1	5,016	Nucleic acid-binding, OB-fold	e-14	85
Hip_MT1_13	4,364	Nucleic acid-binding, OB-fold	e-18	91
Hip_MT1_43	3,913	Nucleic acid-binding, OB-fold	e-10	50
Hip_MT2_19	3,768	Polynucleotidyl transferase, Ribonuclease H fold	e-23	58
Heno_MT1_1	4,140	Nucleic acid-binding, OB-fold	e-111	93
Heto_MT1_1	3,202	Nucleic acid-binding, OB-fold	e-51	68
Heto_MT1_5	3,187	Nucleic acid-binding, OB-fold	e-52	69
Heto_MT1_15	5,428	Nucleic acid-binding, OB-fold	e-29	59
Hip_MT3_1	4,228	Nucleic acid-binding, OB-fold	e-10	86
Hole_SB1_19	2,261	Putative glucosyl transferase	e-147	87
Hole_SB1_23	1,610	Glycerol-3-phosphate dehydrogenase	e-17	82
Hole_SB1_28	3,451	Putative proteasome 26S non-ATPase subunit 2	e-10	40
Hole_SB1_40	2,610	AP-3 complex delta subunit-like protein	0	68
Hole_SB1_46	4,125	Tetratricopeptide repeat (TPR)-containing protein	e-19	50
Hole_SB1_53	2,795	PUTATIVE ubiquitin carboxyl terminal hydrolase	e-13	81
Hole_SB1_55	2,936	Putative RNA-binding protein	e-39	93
Hole_SB1_57	3,159	Hypothetical protein in rice	e-15	58
Hole_SB1_59	2,699	DNA directed RNA polymerase	e-13	72
Hole_SB1_68	4,499	UMP synthase	e-97	96
Hip_SB5_1	4,072	Protein phosphatase 2C-related	e-122	68
Hip_SB2_43	1,763	*F-box domain containing protein*	e-21	68
Hip_SB2_44	1,721	Hypothetical protein in rice	e-56	77
Hip_SB2_54	1,675	Unknown in maize	e-29	76
Hip_SB2_56	2,493	Unknown in maize	e-19	90
Hip_SB3_1	4,400	Unknown in maize	e-19	83
Hip_SB3_4	5,982	Unknown in maize	e-12	52
Hip_SB3_9	2,408	Cycloartenol synthase	e-79	86