

Supplementary Data

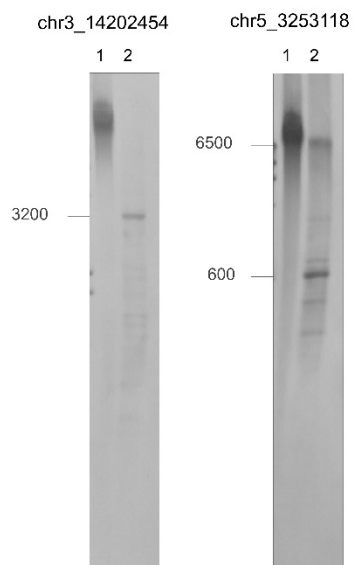
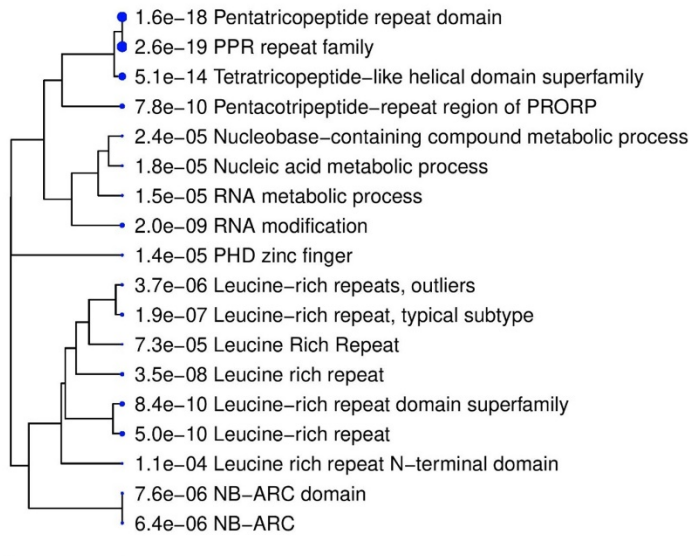


Figure S1: Southern blot detection of eccDNAs

Southern blot of gDNA (1) and EcoRI-digested Phi29-amplified DNA (2) of Col-0 plants, hybridized with eccDNAs chr3_14202454 and chr5_3253118 probes.

a



b

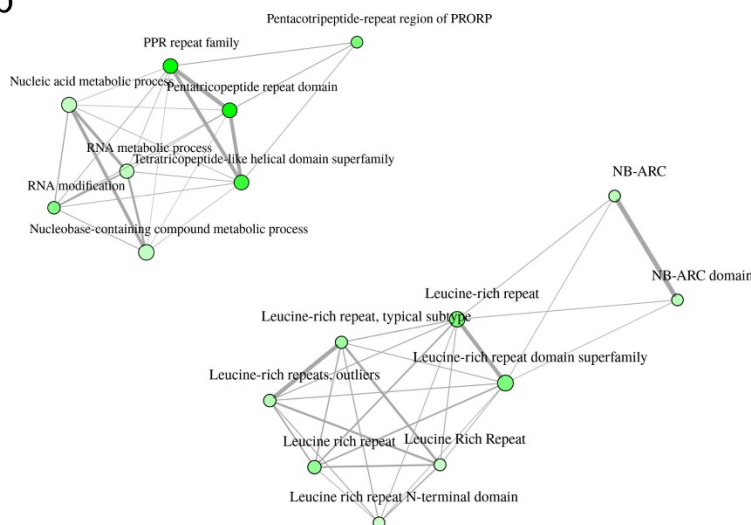


Figure S2: Gene ontology of eccDNA

(a) Hierarchical clustering tree constructed on the eccDNA mapping to the TAIR CDS, summarizing the correlation among significant pathways listed in the gene ontology analysis. Pathways with many shared genes are clustered together. The size of the dots corresponds to the P-values. (b) Network analysis of gene ontology performed on CDS-containing eccDNAs showing the relationship between enriched pathways. Two pathways (nodes) are connected if they share 20% or more genes. The darker colour and bigger size of nodes indicate the more significant p-values and greater number of genes within each set, respectively.

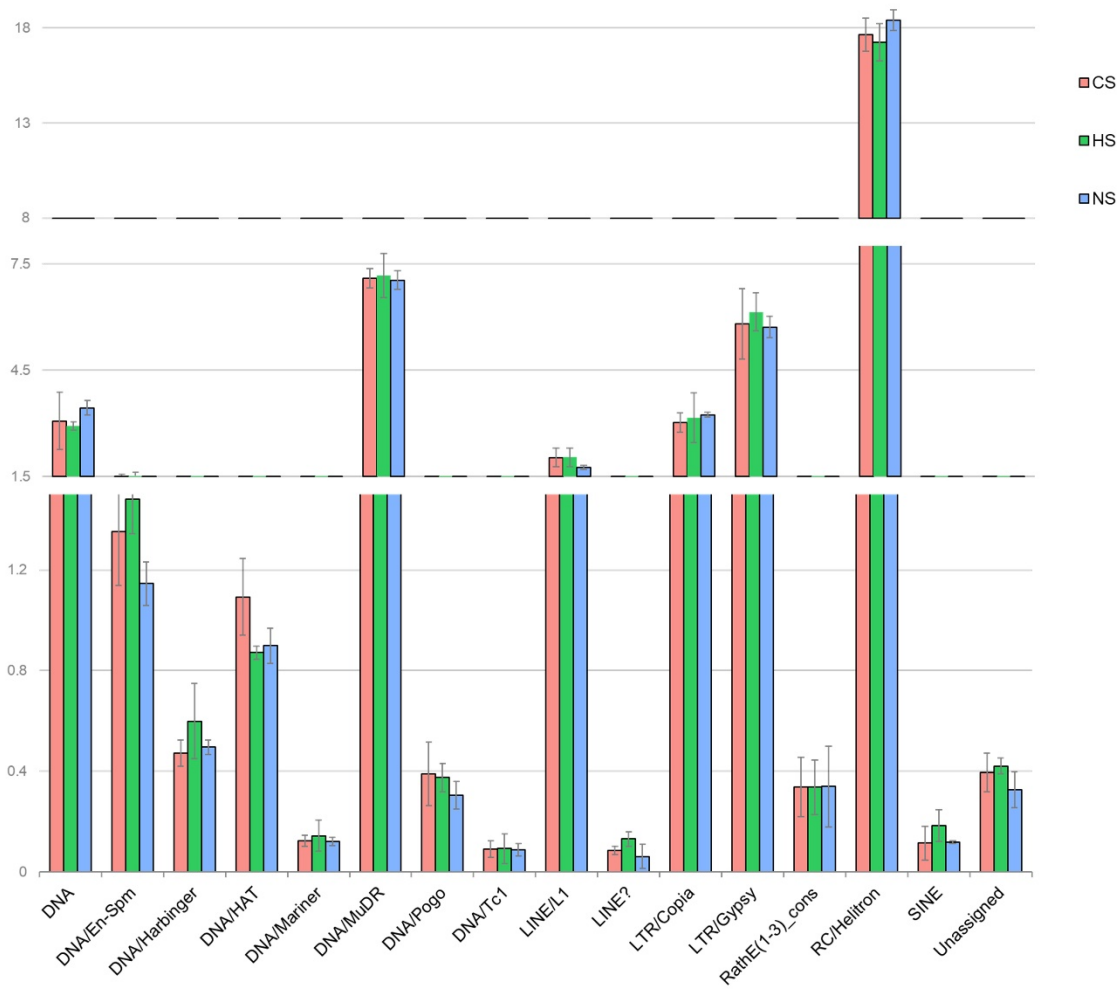


Figure S3: Abundance of eccDNAs in different TE superfamilies.

Arabidopsis TE superfamilies are obtained from TAIR. y-axis represents averaged number of TEs / total eccDNAs in the sample * 100. NS = Col-0 non-stressed, CS = control stress, HS = heat stress (3 replicates for each condition).

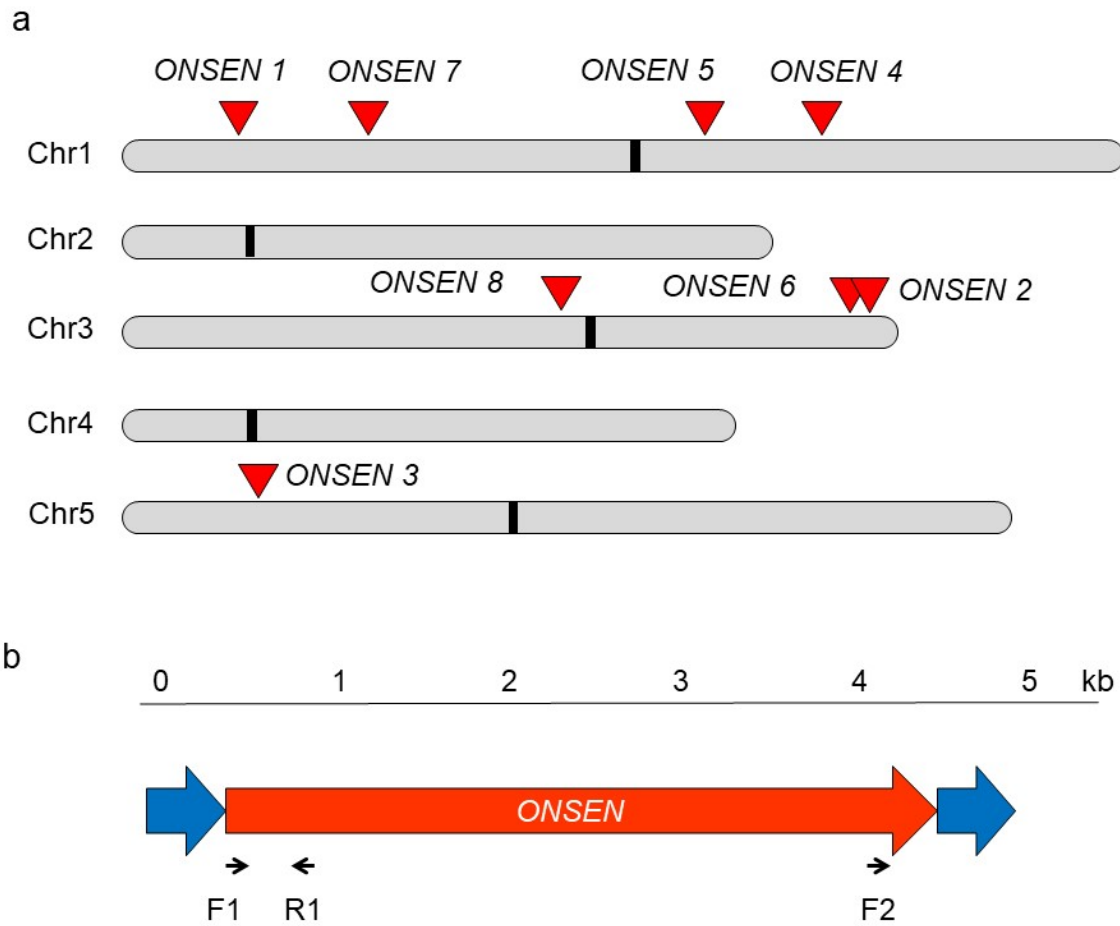


Figure S4. Chromosomal locations and structure of *ONSEN*.

(a) Chromosomal location of *ONSEN* copies in Arabidopsis genome (red triangles); centromeres are marked with black bars. (b) Structure of *ONSEN* with LTRs (blue) and location of primers used for amplification (black arrows).

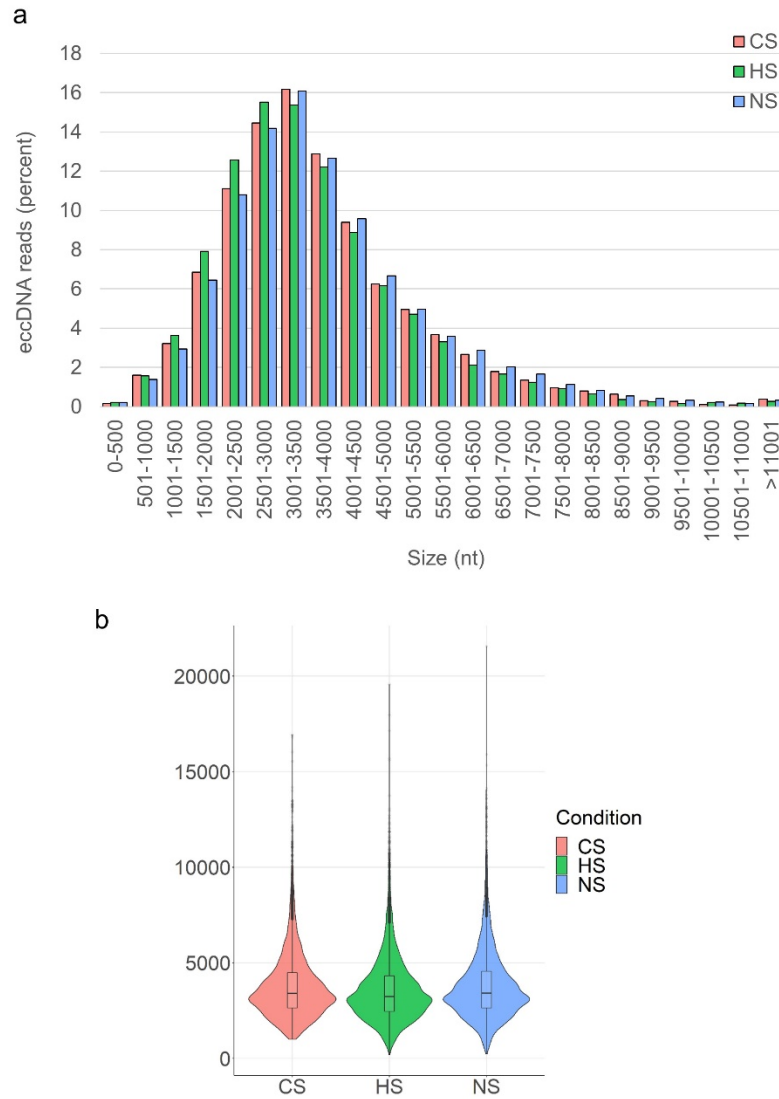


Figure S5: Distribution and size range of eccDNAs in heat and cold stress conditions

(a) Size range-based distribution of eccDNAs from *Arabidopsis*; where y axis indicates percentage abundance of eccDNA reads and x axis indicates the eccDNA size in nucleotides (nt). (b) Size distribution in different conditions displayed as violin plot in which boxplots indicate quantiles and medians. NS = Col-0 non-stressed, CS = control stress, HS = heat stress (3 replicates for each condition).

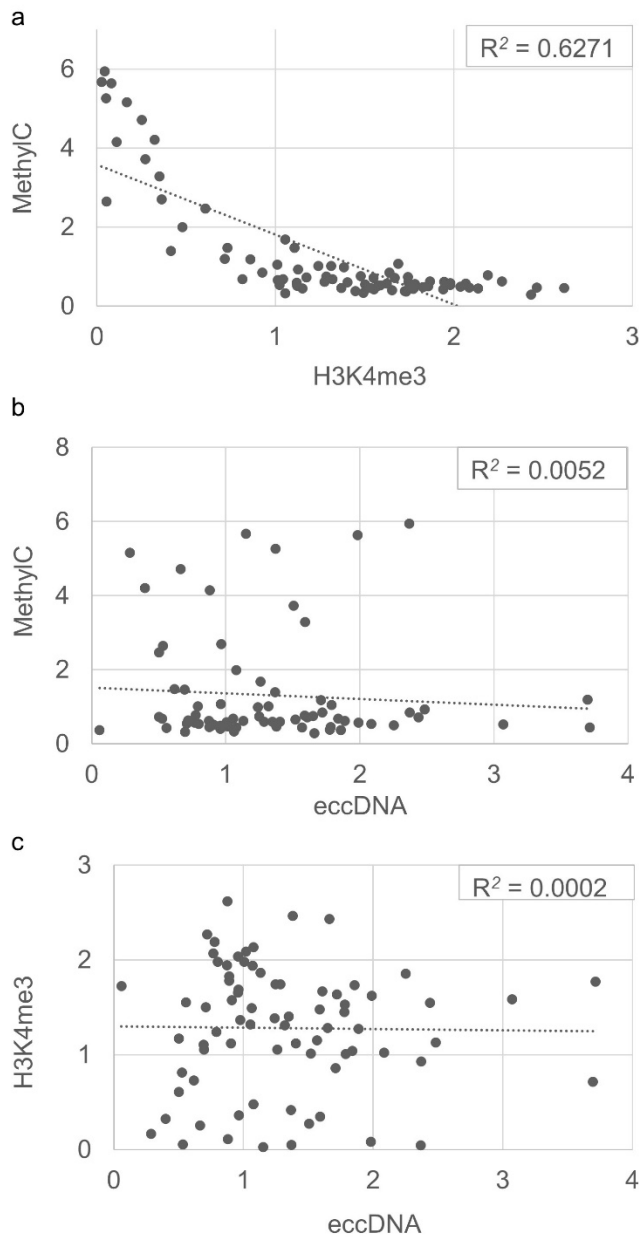


Figure S6: Correlation between eccDNAs and DNA methylation.

(a) Correlation analysis between methylC and H3K4me3. (b) Correlation between eccDNAs and methylC. (c) Correlation between eccDNAs and H3K4me3. R^2 indicates the squared value of correlation coefficient. The units on the x and y axes are percentage H3K4me3, MethylC or eccDNA where each dot represents the respective percentages of the two parameters on each axis. eccDNA refers to the percentage of eccDNAs from the specific genomic locus represented by the dot, where the corresponding value of y axis represents the percentage of MethylC (panel b) or H3K4me3 (panel c).

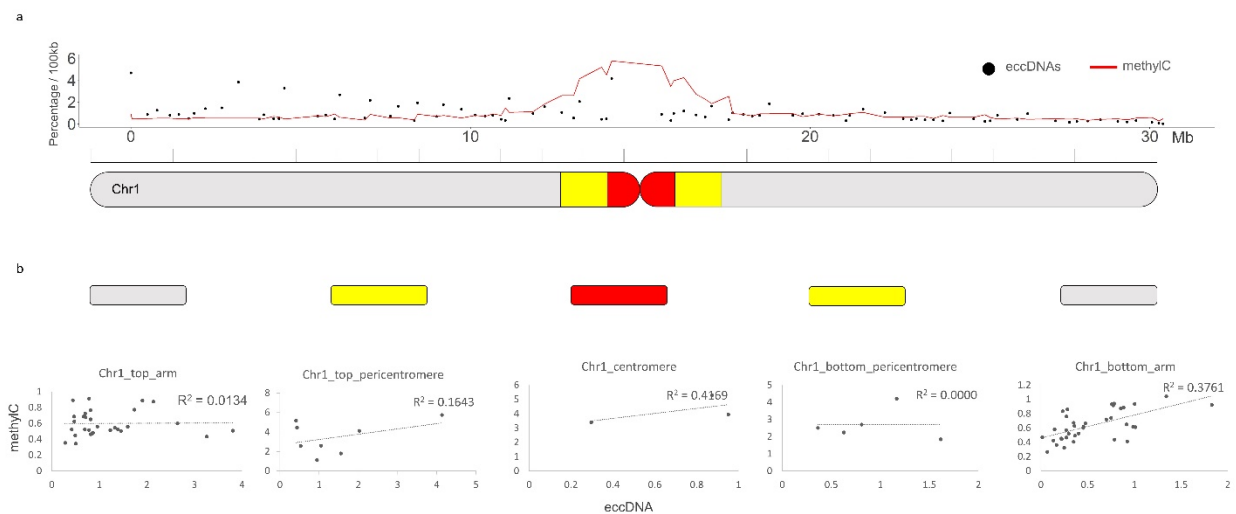


Figure S7: eccDNA distribution compared to the cytosine methylation in Chr1 of Arabidopsis.

(a) eccDNA distribution and C methylation across chromosome 1; x-axis indicates the location on chromosome (Mb), y axis indicates percentage of eccDNAs and methylated Cs in each window (values represent the average of 3 replicates). (b) Correlation among the percentage of eccDNAs and methylated Cs in each region; R^2 indicates the squared value of correlation coefficient.

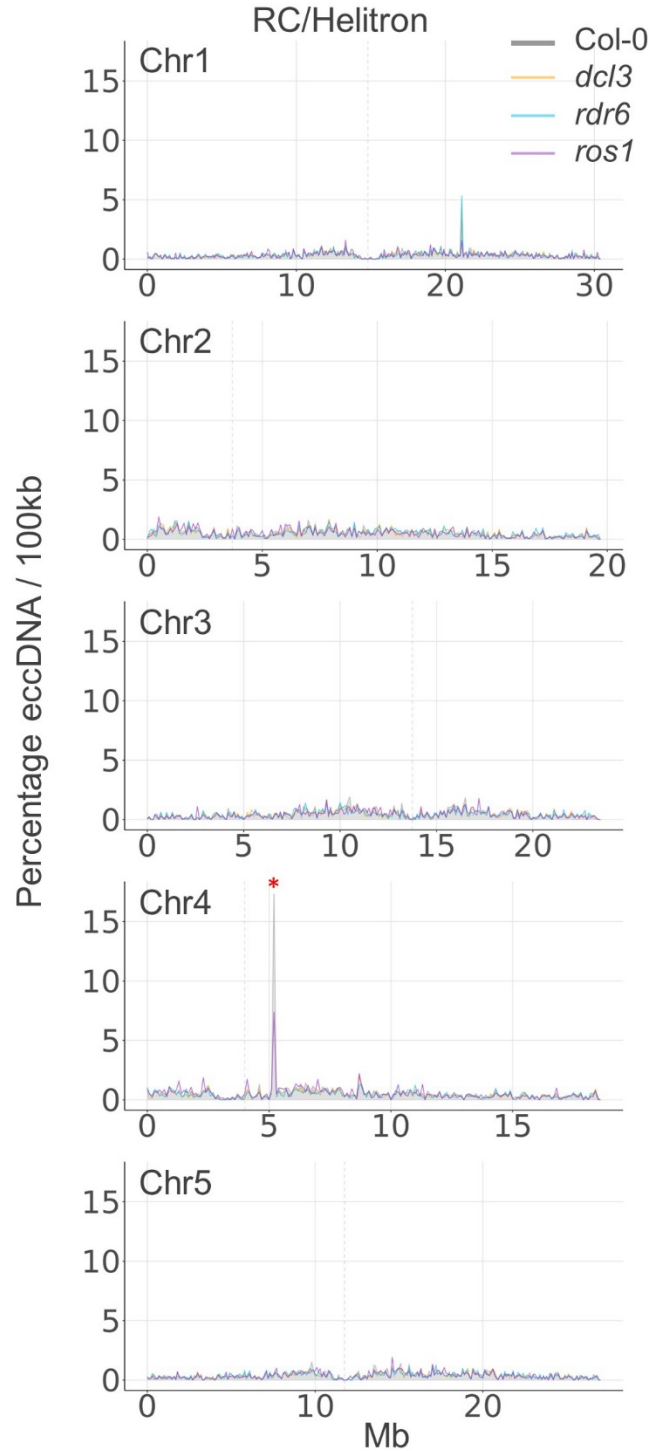


Figure S8. eccDNA profile of Arabidopsis TE superfamily RC/Helitron in Col-0 wt, and mutants *dcl3*, *rdr6* and *ros1*.

Col-0, *dcl3*, *rdr6* and *ros1* leaf tissues were processed on Cider-Seq pipeline (three replicates for each plant type). RC/Helitron-derived eccDNA reads are depicted on Arabidopsis genome; y-axis in each panel indicates percentage eccDNA reads mapped per 100kb bins on Arabidopsis chromosomes. Genomic region on Chr4, represented in Figure 3c, has been highlighted with red asterisk.

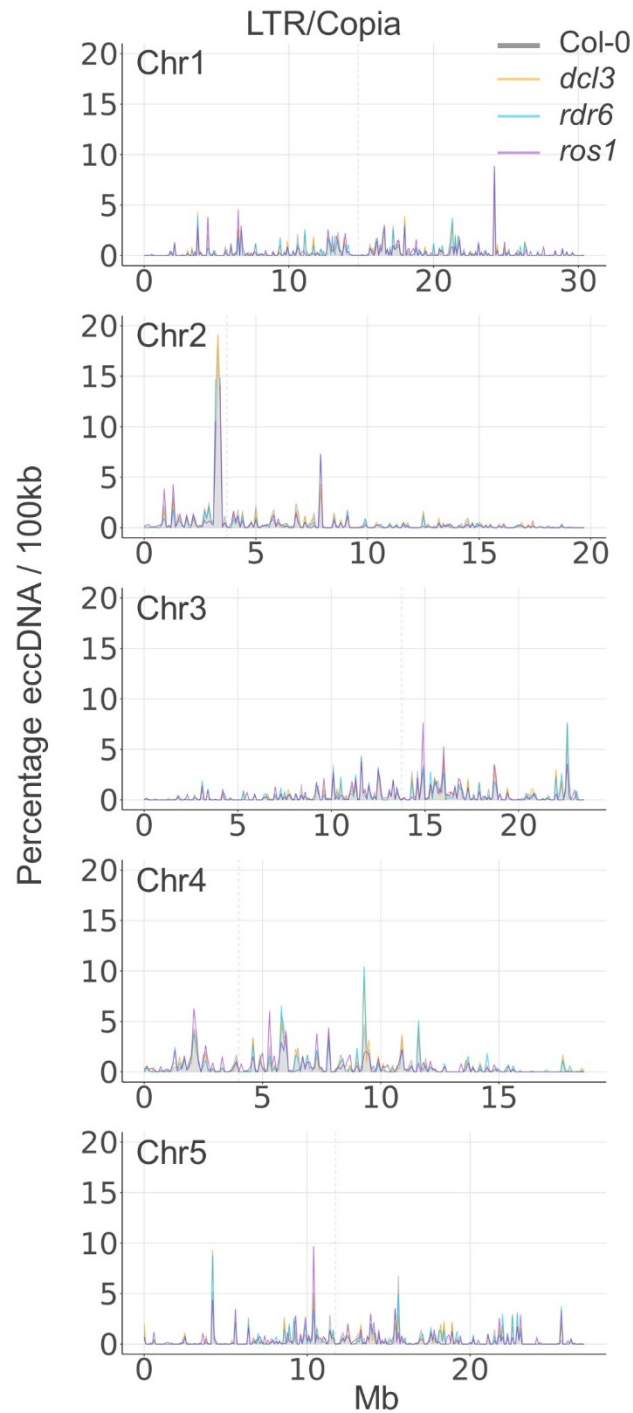


Figure S9. eccDNA profile of Arabidopsis TE superfamily LTR/Copia in Col-0 wt, and mutants *dcl3*, *rdr6* and *ros1*.

Col-0, *dcl3*, *rdr6* and *ros1* leaf tissues were processed on Cider-Seq pipeline (three replicates for each plant type). LTR/Copia-derived eccDNA reads are depicted on Arabidopsis genome; y-axis in each panel indicates percentage eccDNA reads mapped per 100kb bins on Arabidopsis chromosomes.

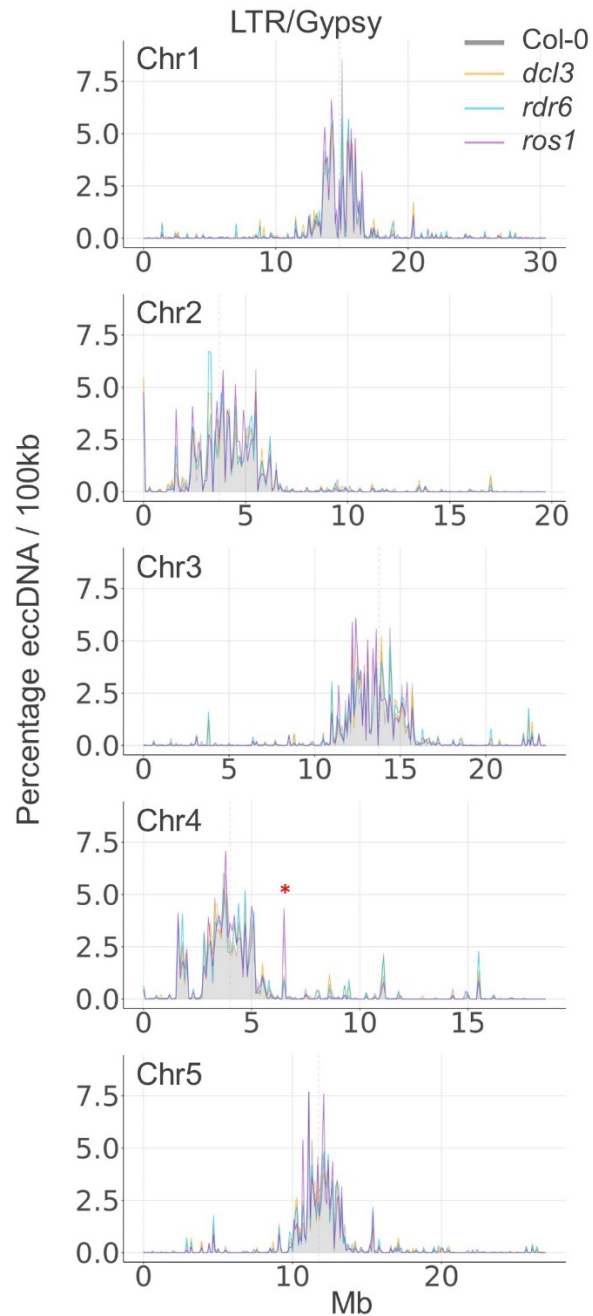


Figure S10. eccDNA profiles of Arabidopsis TE superfamily LTR/Gypsy in Col-0 wt, and mutants *dcl3*, *rdr6* and *ros1*.

Col-0, *dcl3*, *rdr6* and *ros1* leaf tissues were processed on Cider-Seq pipeline (three replicates for each plant type). LTR/Gypsy-derived eccDNA reads are depicted on Arabidopsis genome; y-axis in each panel indicates percentage eccDNA reads mapped per 100kb bins on Arabidopsis chromosomes. Genomic region on Chr4, represented in Figure 3b, has been highlighted with red asterisk.

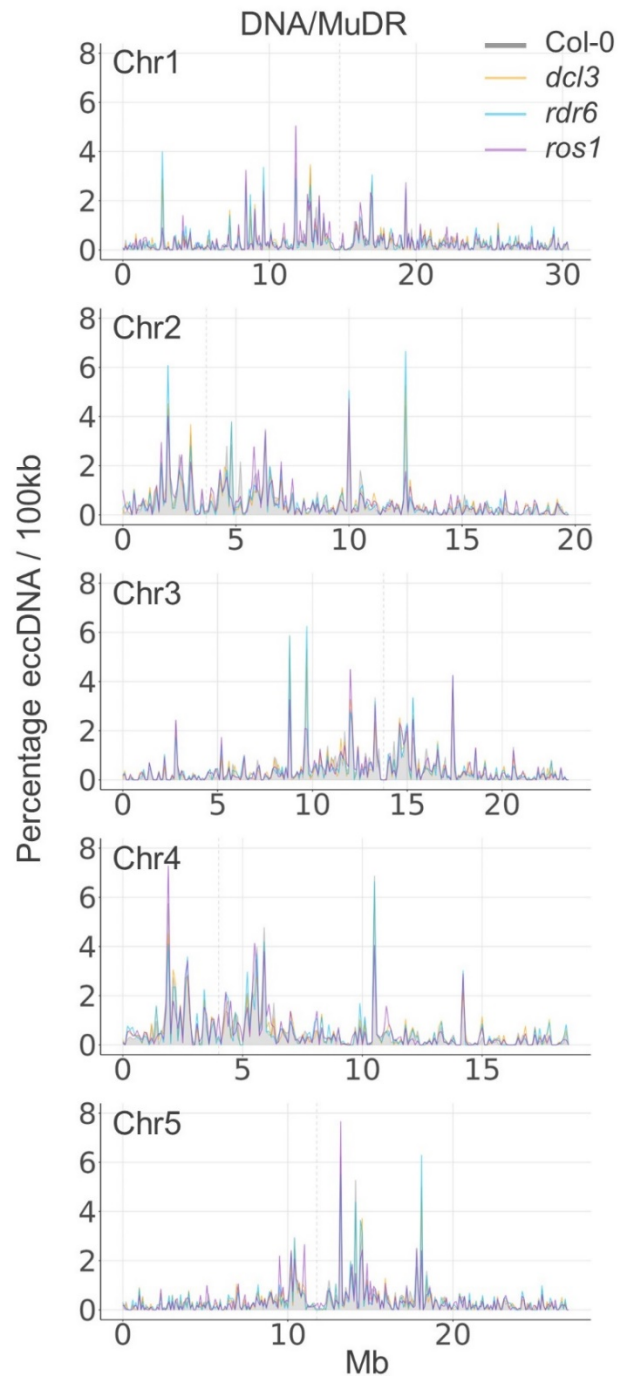


Figure S11. eccDNA profiles of Arabidopsis TE superfamily DNA/MuDR in Col-0 wt, and mutants *dcl3*, *rdr6* and *ros1*.

Col-0, *dcl3*, *rdr6* and *ros1* leaf tissues were processed on Cider-Seq pipeline (three replicates for each plant type). DNA/MuDR-derived eccDNA reads are depicted on Arabidopsis genome; y-axis in each panel indicates percentage eccDNA reads mapped per 100kb bins on Arabidopsis chromosomes.

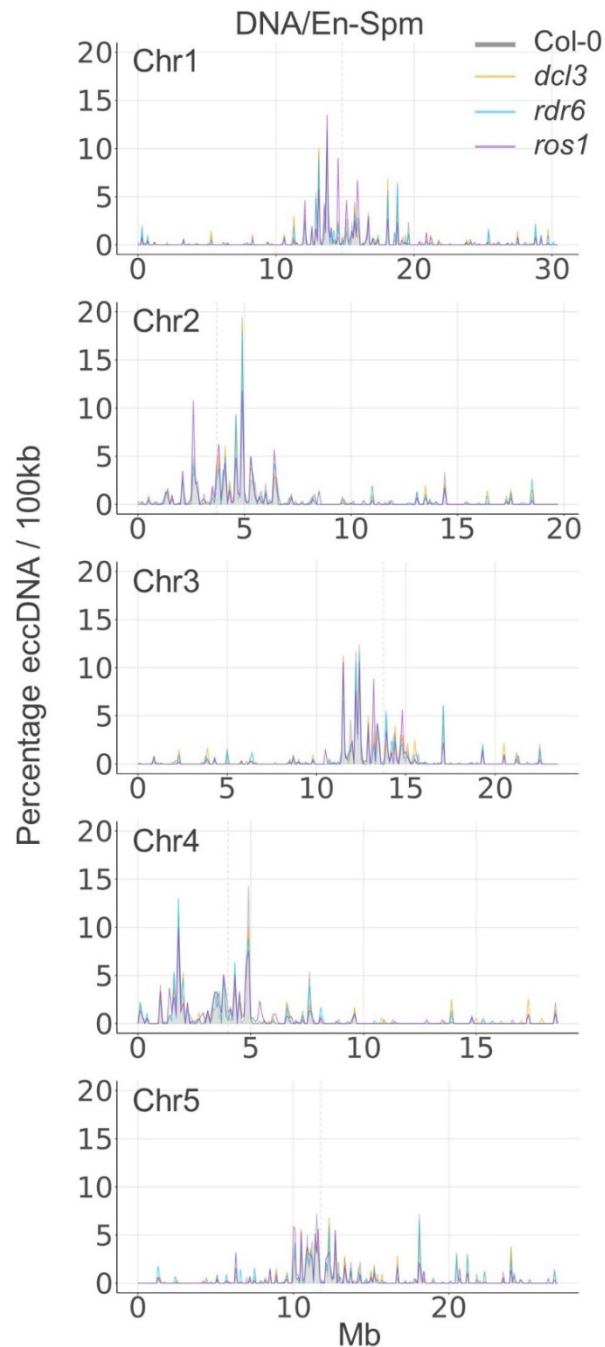


Figure S12. eccDNA profiles of Arabidopsis TE superfamily DNA/En-Spm in Col-0 wt, and mutants *dcl3*, *rdr6* and *ros1*.

Col-0, *dcl3*, *rdr6* and *ros1* leaf tissues were processed on Cider-Seq pipeline (three replicates for each plant type). DNA/En-Spm-derived eccDNA reads are depicted on Arabidopsis genome; y-axis in each panel indicates percentage eccDNA reads mapped per 100kb bins on Arabidopsis chromosomes.

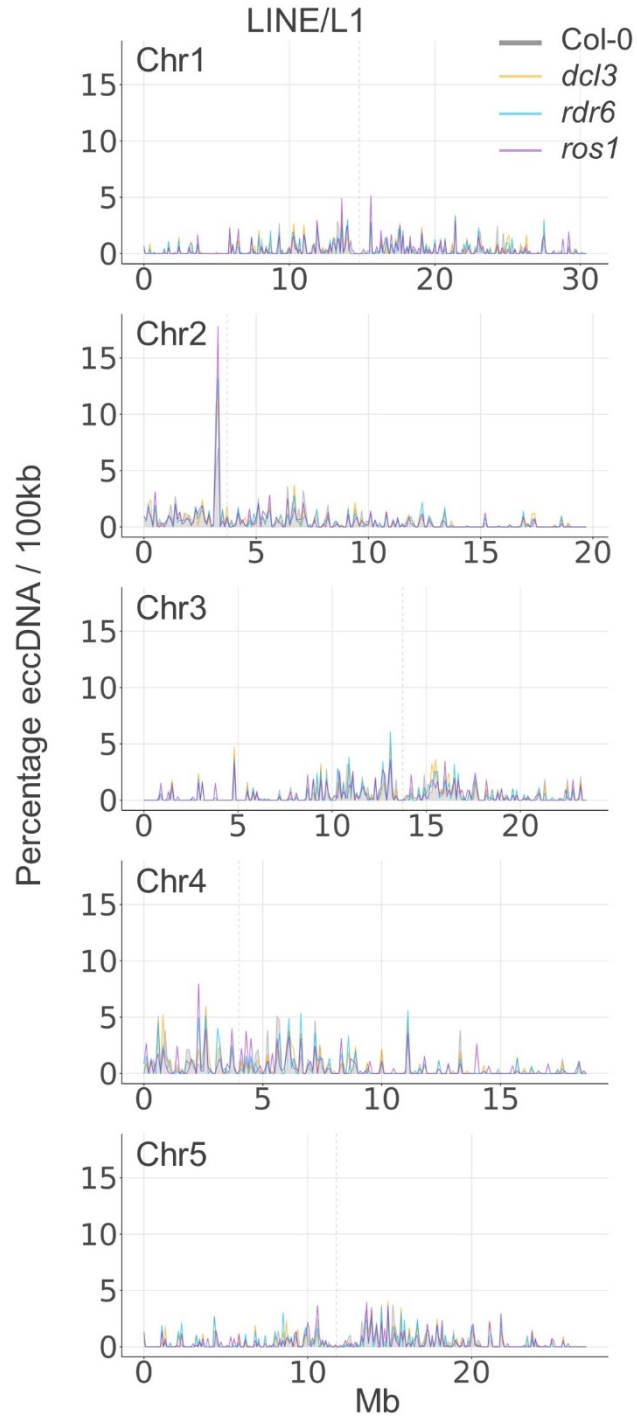


Figure S13. eccDNA profiles of Arabidopsis TE superfamily LINE/L1 in Col-0 wt, and mutants *dcl3*, *rdr6* and *ros1*.

Col-0, *dcl3*, *rdr6* and *ros1* leaf tissues were processed on Cider-Seq pipeline (three replicates for each plant type). LINE/L1-derived eccDNA reads are depicted on Arabidopsis genome; y-axis in each panel indicates percentage eccDNA reads mapped per 100kb bins on Arabidopsis chromosomes.

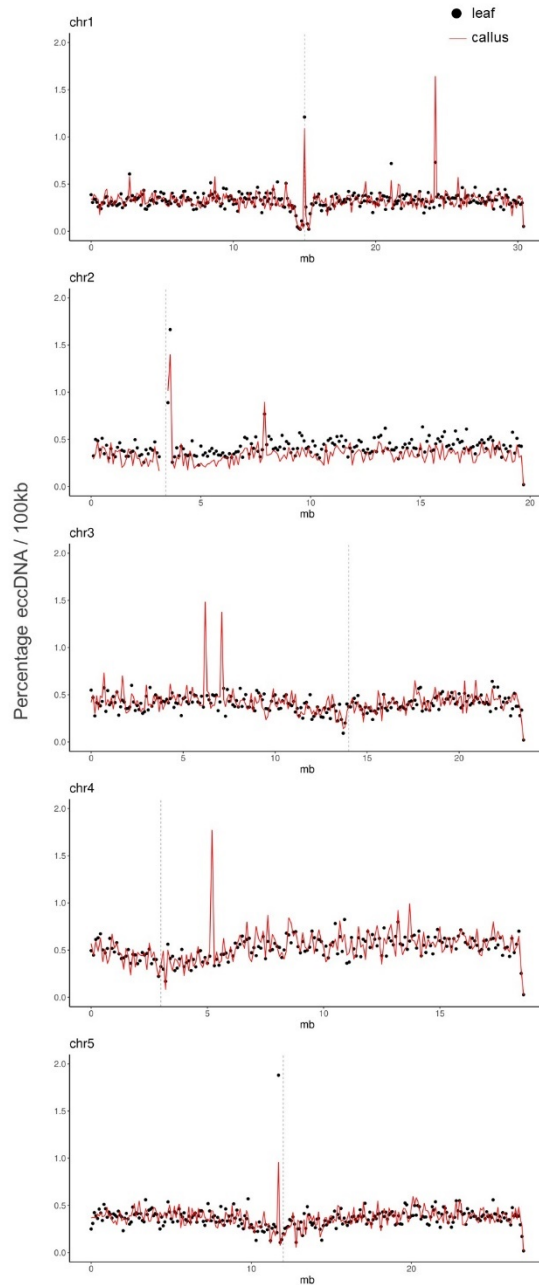
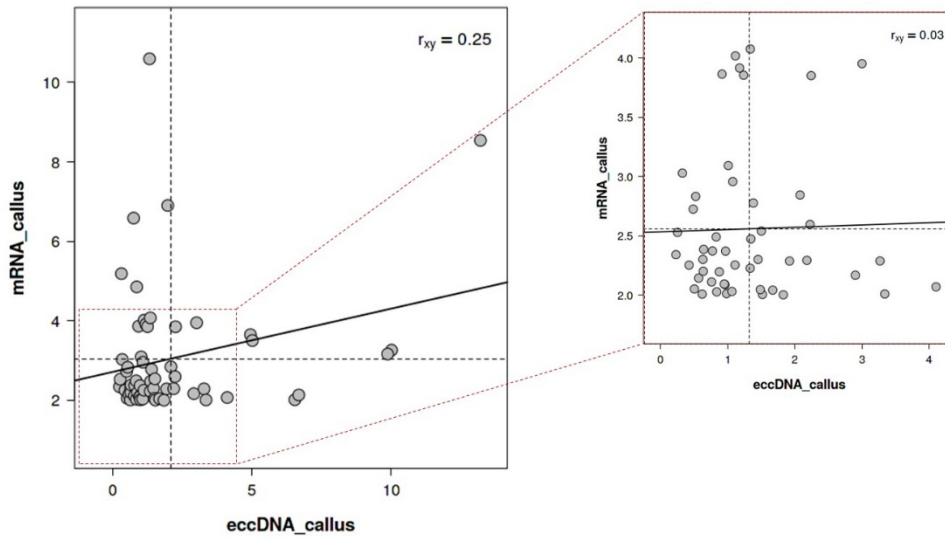


Figure S14: eccDNA profiles of Arabidopsis calli and leaf tissues.

DNA extracted from calli induced from the Col-0 leaves were processed on Cider-Seq pipeline (three replicates for each condition). eccDNA reads from callus and leaf were mapped on Arabidopsis genome; y-axis in each panel indicates percentage eccDNA reads mapped per 100kb bins on Arabidopsis chromosomes.

a



b

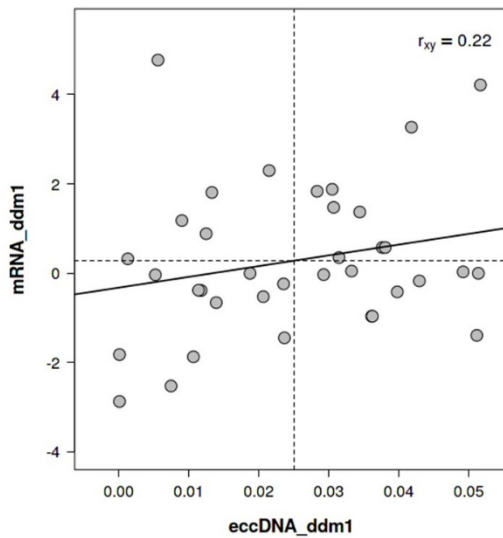


Figure S15: Correlation between eccDNAs and transcript fold change in calli and meristem from *ddm1* mutant.

(a) Correlation analysis between transcript fold change calli/leaf and eccDNA fold change calli/leaf. Pearson's $R = 0.25$. The right panel indicates the same analysis on the clustered area. (b) Correlation analysis between transcript fold change meristem/non-meristem cells and eccDNA fold change meristem/non-meristem cells in Arabidopsis *ddm1*. Pearson's $R = 0.22$.

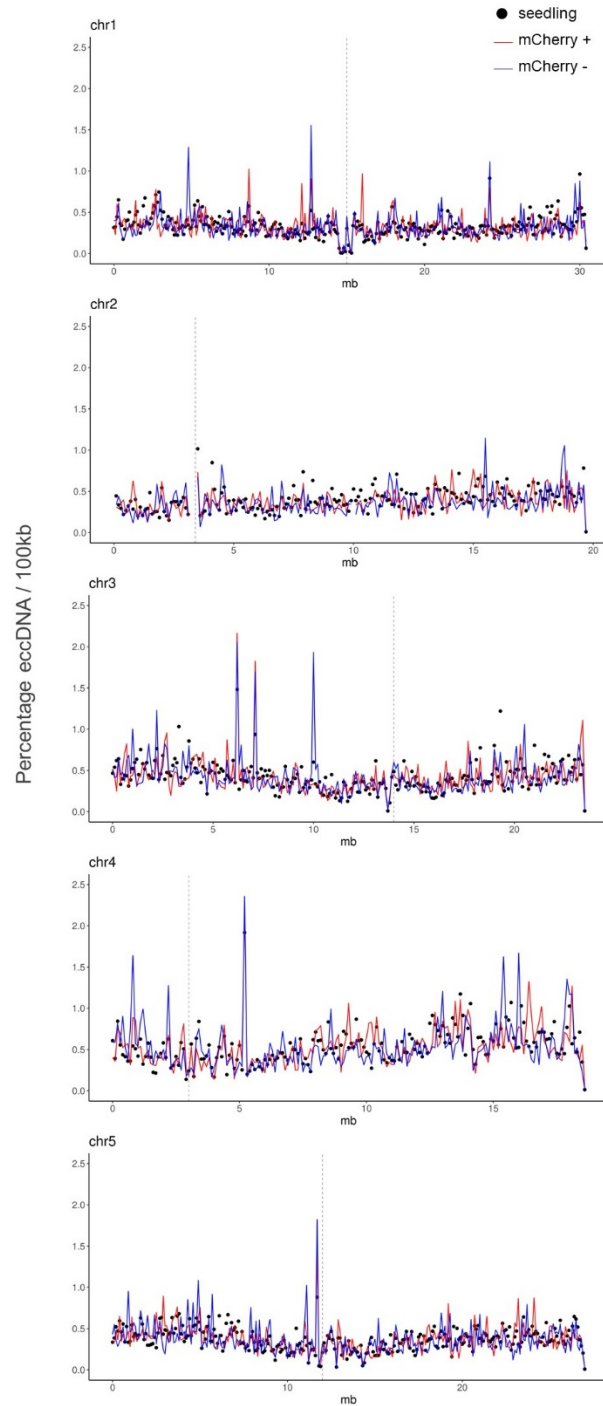


Figure S16: eccDNA profiles of *ddm1* shoot apical meristem-enriched and non-enriched samples.

Shoot apical meristem (SAM) cells were sorted using fluorescence-activated nuclear sorting (FANS) and processed, with respective controls from *ddm1* seedlings, on Cider-Seq pipeline for eccDNA amplification (three replicates for each cell type). EccDNA reads from SAM (mCherry +), surrounding cells (mCherry -) and *ddm1* seedlings were mapped on Arabidopsis genome; y-axis in each panel indicates percentage eccDNA reads mapped per 100kb bins on Arabidopsis chromosomes.

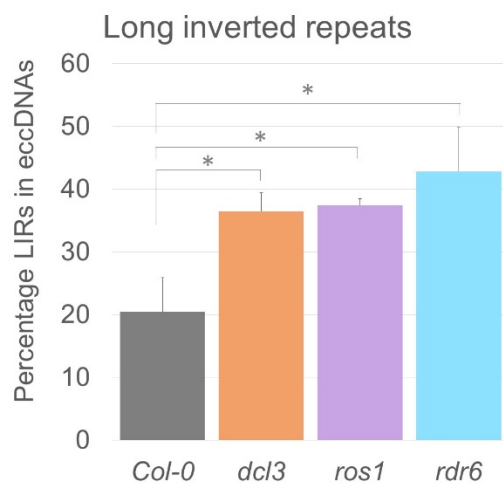


Figure S17: Long inverted repeats in the eccDNAs.

Long inverted repeats (LIRs) were identified in the eccDNAs of Col-0 and mutants *dcl3*, *ros1* and *rdr6* (3 replicates each) by mapping the eccDNAs to a comprehensive database of Arabidopsis LIRs, LIRBase. Asterisks indicate p-value < 0.05.

Table S1. PacBio sequencing of eccDNA; data, read statistics and clustering.

Demultiplexed data from Sequel II is presented in Mb with number of reads in corresponding samples; demultiplexing performed with LIMA. Clustering is performed with cd-hit where unique clusters represent the clusters with single reads. CCS = circular consensus sequencing, NS = Col-0 non-stressed, CS = control stress, HS = heat stress.

Sample	Condition	CCS reads	eccDNA			Clustering	
		No.	No.	bases	Mean length	No. of clusters	Unique clusters
cs1	control stress-1	11727	4451	1.4E+07	3165.7	3845	3555
cs2	control stress-2	12482	4765	1.6E+07	3297	3932	3553
cs3	control stress-3	14569	4956	1.8E+07	3599.7	4011	3670
hs1	heat stress-1	15906	6445	2.1E+07	3184.6	5164	4634
hs2	heat stress-2	9180	4067	1.4E+07	3437	3506	3226
hs3	heat stress-3	9984	3206	1.1E+07	3360	2774	2580
wt1	No stress Col0-1	32887	398	1672072	4201.2	297	270
wt2	No stress Col0-2	111946	3376	5003134	1482	2311	2151
wt3	No stress Col0-3	18967	415	1471814	3546.5	274	235
plus5k-1	mCherry + 1	43525	942	3057522	3245.8	757	37
plus5k-2	mCherry + 2	35391	638	2260398	3309.5	515	36
plus5k-3	mCherry + 3	32121	794	2224509	2801.6	484	451
min10k1	mCherry - 1	98615	2282	6970826	3054.7	1639	1505
min10k2	mCherry - 2	53996	1441	4584957	3181.8	879	816
min10k3	mCherry - 3	32568	593	2044158	3447.1	310	277
ddm1 -1	<i>ddm1</i> -1	78437	1555	5217786	3355.5	1287	1197
ddm1 -2	<i>ddm1</i> -2	48134	936	3027933	3235	826	789
ddm1 -3	<i>ddm1</i> -3	58508	1725	4624944	2681.1	1243	1133
cal1	callus-1	25674	426	1640948	3852	332	302
cal2	callus-2	50259	844	3248366	3848.8	536	472
cal3	callus-3	33212	674	2708945	3474.3	442	403
ros1 -1	<i>ros1</i> -1	68186	804	2814933	3501.2	441	394
ros1 -2	<i>ros1</i> -2	81264	1082	4528322	4185.1	757	668
ros1 -3	<i>ros1</i> -3	74668	827	3198883	3868.1	487	417
rdr6 -1	<i>rdr6</i> -1	70251	807	3718641	4608	556	468
rdr6 -2	<i>rdr6</i> -2	63685	781	3290653	4213.4	484	421
rdr6 -3	<i>rdr6</i> -3	84428	1244	6921003	5563.5	932	824
dcl3 -1	<i>dcl3</i> -1	65855	847	4638491	5476.4	619	542
dcl3 -2	<i>dcl3</i> -2	78366	1136	6199007	5456.9	869	755
dcl3 -3	<i>dcl3</i> -3	63497	1307	6807898	5208.8	767	670

* CCS was performed with 0.99 % accuracy

** Based on >1 rounds of deconcat

Table S2: Primers used in this study.

Primer	Sequence
ONSEN_F1	TAATGTTCCCTTCCAAGTCCC
ONSEN_R1	GCTTGTAATGACCCAAGAAGT
ONSEN_F2	AAGTCGGCAATAGCTTTGGCGAAGA
chr3_14202454-F	CGAAGGTGCATAGTGAGAAGAGTAA
chr3_14202454-R	GTTCCCTTGGCTGTGGTTTCGCTG
chr5_3253118-F	GGTAATATCATGCGAACAAG
chr5_3253118-FR	GTGGCGGTATAGTCGTCTTG

Table S3. eccDNA characterization and number of full-length TEs in eccDNAs.

Sample	Condition	CCS reads	eccDNA	TE eccDNA	Full TE in eccDNA	
		No.	No.	No.	No.	percent
cs1	control stress-1	11727	4451	1659	861	51.898
cs2	control stress-2	12482	4765	1691	884	52.276
cs3	control stress-3	14569	4956	1869	1006	53.825
hs1	heat stress-1	15906	6445	2267	1236	54.521
hs2	heat stress-2	9180	4067	1579	832	52.691
hs3	heat stress-3	9984	3206	1178	641	54.414
wt1	No stress Col0-1	32887	398	141	88	62.411
wt2	No stress Col0-2	111946	3376	688	231	33.575
wt3	No stress Col0-3	18967	415	112	65	58.035
plus5k-1	mCherry + 1	43525	942	291	140	48.11
plus5k-2	mCherry + 2	35391	638	211	102	48.341
plus5k-3	mCherry + 3	32121	794	204	92	45.098
min10k1	mCherry - 1	98615	2282	669	305	45.590
min10k2	mCherry - 2	53996	1441	422	185	43.838
min10k3	mCherry - 3	32568	593	160	77	48.125
ddm1 -1	<i>ddm1</i> -1	78437	1555	487	267	54.825
ddm1 -2	<i>ddm1</i> -2	48134	936	326	161	49.386
ddm1 -3	<i>ddm1</i> -3	58508	1725	442	202	45.701
cal1	callus-1	25674	426	153	85	55.555
cal2	callus-2	50259	844	258	139	53.876
cal3	callus-3	33212	674	202	107	52.970
ros1 -1	<i>ros1</i> -1	68186	804	198	110	55.555
ros1 -2	<i>ros1</i> -2	81264	1082	339	185	54.572
ros1 -3	<i>ros1</i> -3	74668	827	231	118	51.082
rdr6 -1	<i>rdr6</i> -1	70251	807	271	171	63.099
rdr6 -2	<i>rdr6</i> -2	63685	781	220	122	55.454
rdr6 -3	<i>rdr6</i> -3	84428	1244	527	338	64.136
dcl3 -1	<i>dcl3</i> -1	65855	847	339	203	59.882
dcl3 -2	<i>dcl3</i> -2	78366	1136	516	311	60.271
dcl3 -3	<i>dcl3</i> -3	63497	1307	386	250	64.766

Table S4: TE-derived eccDNAs from respective TE superfamilies in Arabidopsis col-0 and mutants.

TE superfamily / Plant	Col -0 1	Col -0 2	Col -0 3	dcl 3 1	dcl 3 2	dcl 3 3	rdr 6 1	rdr 6 2	rdr 6 3	ros 1 1	ros 1 2	ros 1 3	dd m1 1	dd m1 2	dd m1 3
RC/Helitron	60	266	61	142	219	159	105	86	233	94	132	102	183	118	167
LTR/Gypsy	20	117	13	72	98	70	58	57	81	39	71	47	77	43	70
DNA/MuDR	22	127	12	49	69	56	37	32	88	33	48	29	83	68	79
LTR/Copia	13	56	7	29	39	41	18	16	45	12	34	22	47	24	33
DNA	11	30	7	13	26	28	10	8	22	12	13	13	24	13	15
LINE/L1	4	29	6	13	14	22	12	6	28	7	17	9	20	16	27
DNA/En-Spm	6	31	5	14	25	12	14	10	18	7	11	18	20	18	18
DNA/HAT	4	12	5	8	21	10	5	6	13	3	7	6	9	6	12
Unassigned	1	9	2	2	3	3	7	2	10	5	6	1	6	5	5
DNA/Harbing er	1	4	1	3	5	5	5	4	5	2	3	1	7	7	6
DNA/Pogo	2	11		2	5	2	1	1	5		4	3	3	4	4
LINE?		5		1	2				1		4	1	1	1	1
DNA/Tc1		1			2	1		1	3			2	2	2	1
RathE1_cons				1	1	1	1		2			1	3		2
DNA/Mariner	1	1	1	1	2	1	1		1			1			
RathE3_cons		2			1	2					1		1		2
RathE2_cons	1	2			1			1						1	
SINE		2					1		1				1		

Table S5: TEs with most abundant eccDNA reads.

TAIR id	TE copy No.									TE family	TE Super family
	CS1	CS2	CS3	HS1	HS2	HS3	WT1	WT2	WT3		
AT4TE14890	31	43	80	83	41	29	48	74	41	DNA	ATREP18
AT2TE26350	10	12	26	15	9	5	20	28	24	DNA	ATREP19
AT1TE38090	19	22	50	59	23	19	41	48	49	DNA/MuDR	VANDAL2N1
AT2TE35755	26	27	15	35	11	6	17	22	26	LTR/Copia	ATCOPIA28
AT2TE16235	21	29	19	39	8	18	19	29	19	LTR/Gypsy	ATLANTYS3
AT4TE76870	17	30	27	29	19	11	16	30	30	RC/Helitron	ATREP10
AT2TE44930	13	21	37	23	11	15	26	39	27	RC/Helitron	ATREP10B
AT5TE11315	32	20	38	39	16	12	49	28	25	RC/Helitron	ATREP10D
AT1TE69930	13	14	14	20	3	11	27	26	25	RC/Helitron	ATREP4
AT3TE77985	6	10	21	27	9	8	17	21	15	RC/Helitron	HELITRONY1B

Table S6: TE-derived eccDNAs from respective TE families in Arabidopsis col-0 and mutants.

TE family / Plant	Col -0 1	Col -0 2	Col -0 3	dcl 3 1	dcl 3 2	dcl 3 3	rdr 6 1	rdr 6 2	rdr 6 3	ros 1 1	ros 1 2	ros 1 3	dd m1 1	dd m1 2	dd m1 3
ATREP3	5	32	11	22	22	11	20	15	27	6	20	13	23	18	20
ATREP4	5	31	7	14	22	14	7	7	25	19	11	13	11	17	16
ATREP10	8	33	4	17	18	19	7	9	18	11	8	5	26	7	26
HELITRO NY3	7	22	4	10	20	12	7	9	16	9	19	8	15	5	7
ATREP10 B	1	19	4	10	17	21	6	8	22	8	9	10	6	5	11
ATREP15	6	9	5	11	13	10	8	2	14	7	9	5	14	11	16
ATHILA2	4	13	1	11	14	12	9	12	11	6	8	2	6	11	6
ATREP11	4	19	3	10	11	9	7	4	9	6	6	4	10	9	11
ATREP19	5	6	6	6	12	14	3	4	10	6	5	7	8	3	8
HELITRO NY1D	2	16	3	6	9	5	3	1	14	5	3	2	14	7	8
ATREP5	1	13		5	14	9	5	5	11	4	8		7	6	10
ATHILA6 A	1	13	5	7	14	5	9	3	7	5	4	4	9	2	10
ATGP1	2	7	1	9	11	8	4	5	6	1	8	1	5	4	9
ATREP1	1	5	3	2	9	11	8	4	6	1	5	5	7	4	8
ATHILA3	1	7	1	4	9	7	5	7	11	1	4	3	3	2	6

Table S7: EccDNA mapped on Arabidopsis genome:

EccDNAs from Col-0, heat-stress, control-stress, meristem-enriched, meristem non enriched, callus tissues and mutants dcl3, rdr6, ros1 and ddm1 have been mapped on Arabidopsis genome (tair10). Mapping was done using BLAST and various summary statistics are provided such as alignment length, start and end sites, percentage identities etc.

Table S8: EccDNA mapped on Arabidopsis CDS:

EccDNAs from Col-0, heat-stress, control-stress, meristem-enriched, meristem non enriched, callus tissues and mutants dcl3, rdr6, ros1 and ddm1 have been mapped on Arabidopsis CDS. A BLAST database was generated using tair9 CDS dataset and mapping was done using BLAST. Various summary statistics are provided such as alignment length, start and end sites, percentage identities etc.

Table S9: EccDNA mapped on Arabidopsis transposable elements:

EccDNAs from Col-0, heat-stress, control-stress, meristem-enriched, meristem non enriched, callus tissues and mutants dcl3, rdr6, ros1 and ddm1 have been mapped on Arabidopsis transposable elements (TEs). A BLAST database was generated using tair9 TE dataset and mapping was done using BLAST. Various summary statistics are provided such as alignment length, start and end sites, percentage identities etc.