

Supplemental information

Transposons repressed by H3K27me3 were co-opted as cis-regulatory elements of H3K27me3 controlled protein coding genes during evolution of plants

Tetsuya Hisanaga, Facundo Romani, Shuangyang Wu, Teresa Kowar, Ruth Lintermann, Bhagyshree Jamge, Sean A. Montgomery, Elin Axelsson, Tom Dierschke, John L. Bowman, Takayuki Fujiwara, Shunsuke Hirooka, Shin-ya Miyagishima, Liam Dolan, Daniel Schubert and Frédéric Berger

Figure S1. Annotation of TEs and re-analysis of 5mC presence in *C. merolae*

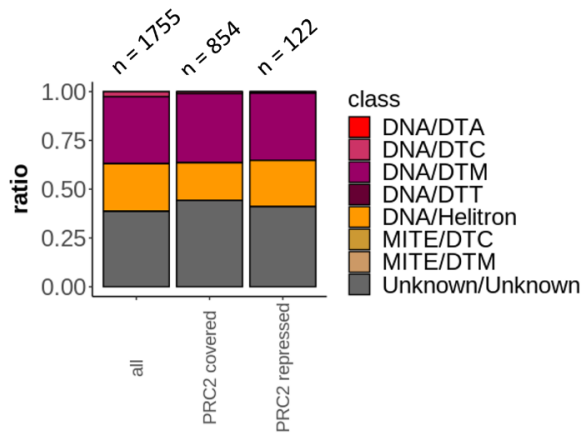
Figure S2. Short TEs in *A. agrestis* and *M. polymorpha*

Figure S3 Generation of PRC2 knock out mutant in *M. polymorpha*

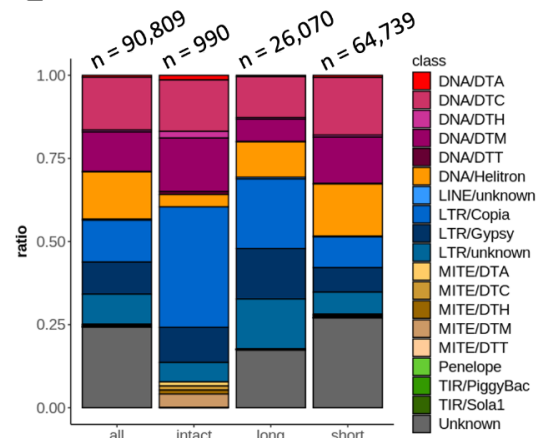
Figure S4 TE clusters in *A. thaliana*

Table S1. oligos used in this study

A



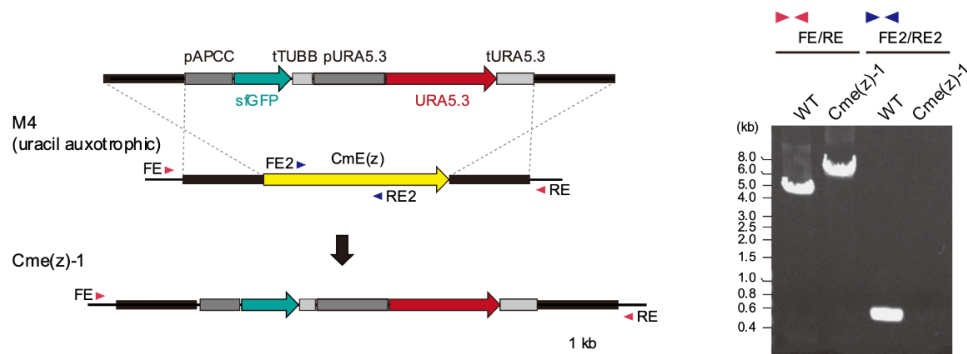
E



B

	mCG (%)	mCHG (%)	mCHH (%)
Nuclear	0.2	0.18	0.19
Chloroplast	0.2	0.19	0.21
Mitochondrion	0.18	0.15	0.2

C



D

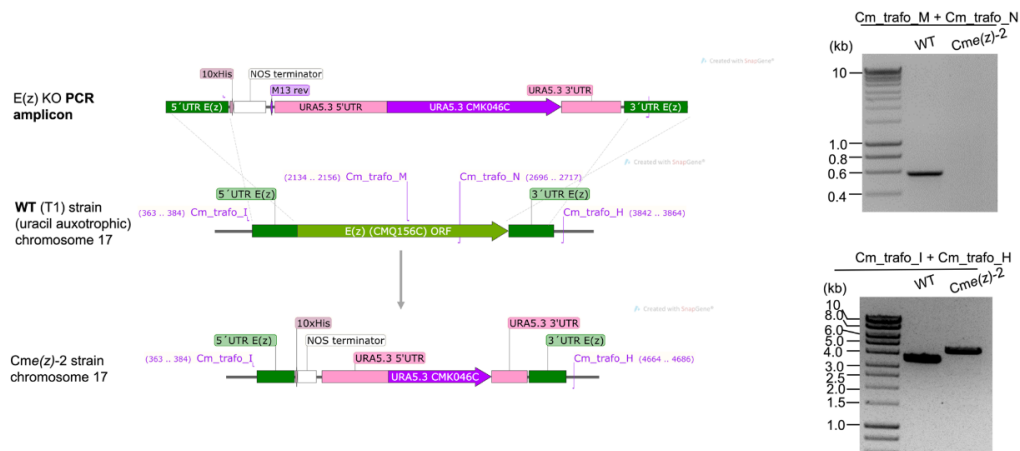


Figure S1. Annotation of TEs and re-analysis of 5mC presence in *C. merolae*

(A) Stacked bar chart indicating proportion of TE families in all TE annotated (all TE), TEs covered by H3K27me3 (PRC2 covered), or TEs repressed by PRC2 (PRC2 repressed). Total numbers of TEs in each category are shown on bars.

(B) Table showing summary of DNA methylation levels in nuclear and organellar genome of *C. merolae*.

(C), (D) Schematic illustration showing details of *Cme(z)-1* (C) and *Cme(z)-2* (D) mutations. Two uracil-auxotrophic WT strains, M4 and T1, respectively, were transformed via homologous recombination resulting in two independent *Cme(z)* mutant lines.

(E) Stacked bar chart indicating proportion of TE families in all TE annotated (all), intact TEs (intact), TEs longer than 500 bp (long) and TEs shorter than 500 bp (short). Total numbers of TEs in each category are shown on bars.

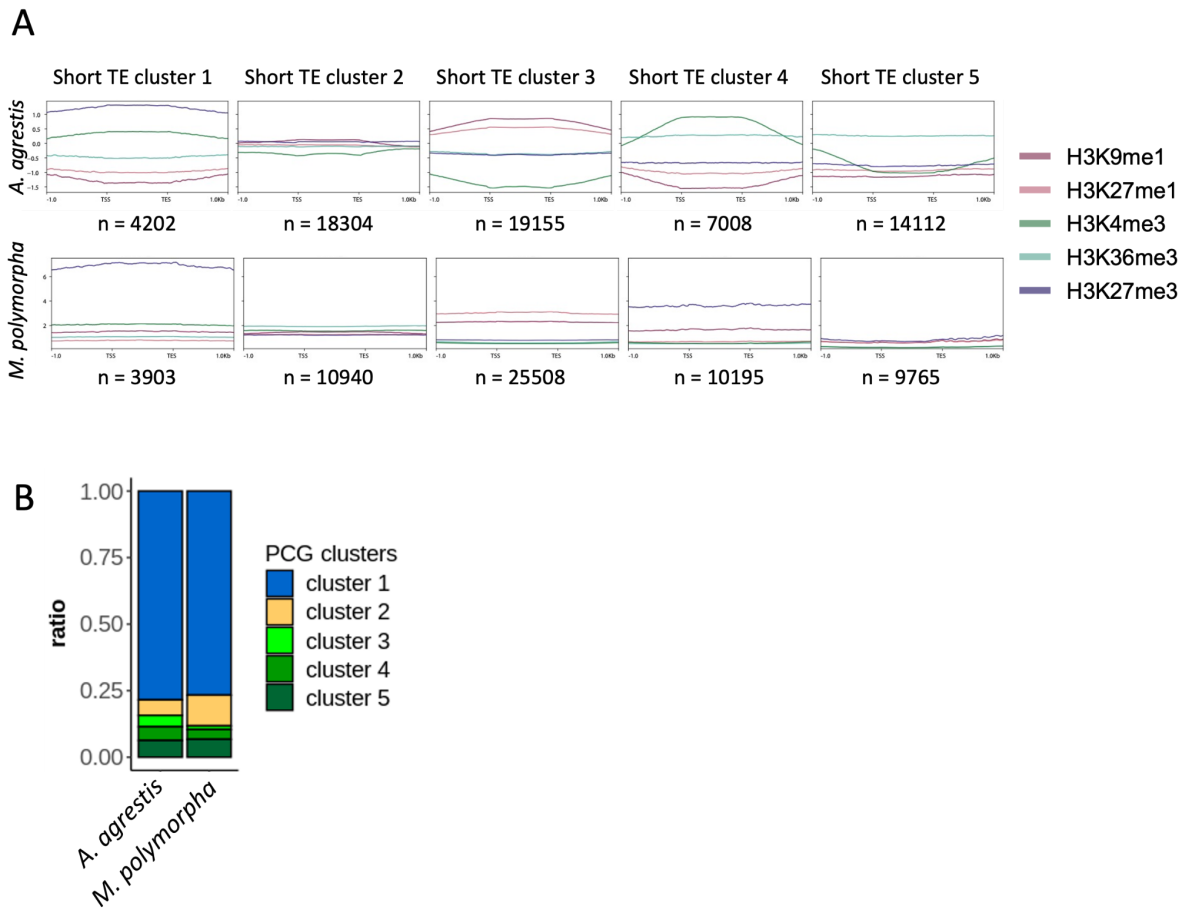
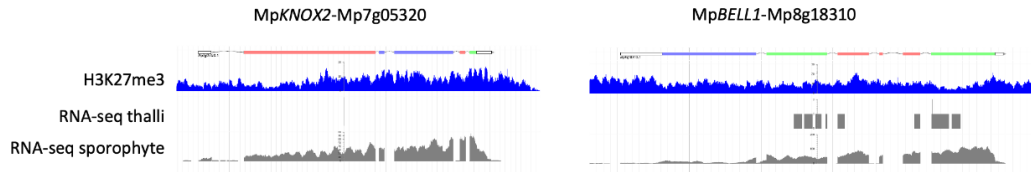


Figure S2. Short TEs in *A. agrestis* and *M. polymorpha*

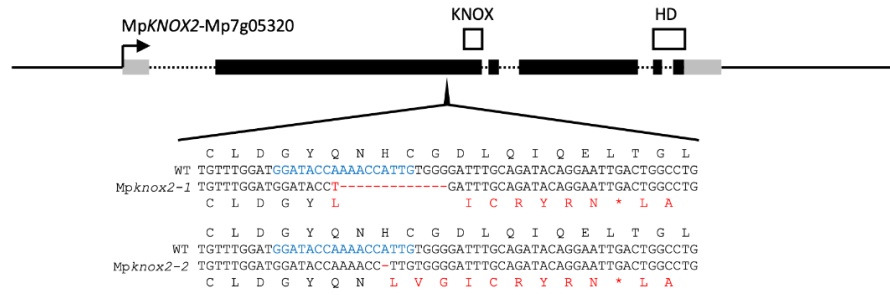
(A) Aggregate profile plots showing log₂ ChIP/H3 enrichment for various chromatin modifications per short TE clusters of *A. agrestis* (top) or *M. polymorpha* (bottom).

(B) Stacked bar chart showing proportion of gene clusters of nearby genes of TEs in short TE cluster 1 in *A. agrestis* (left) and *M. polymorpha* (right).

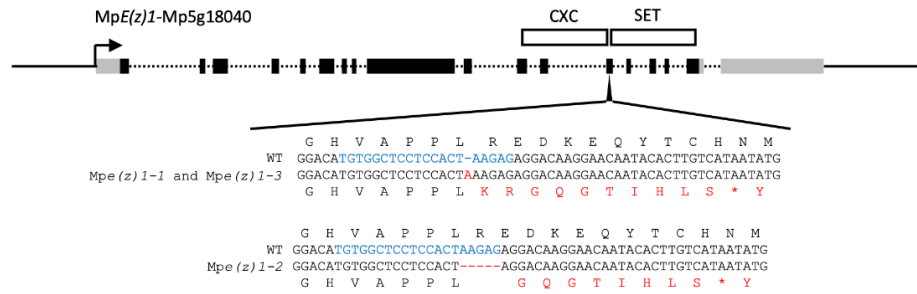
A



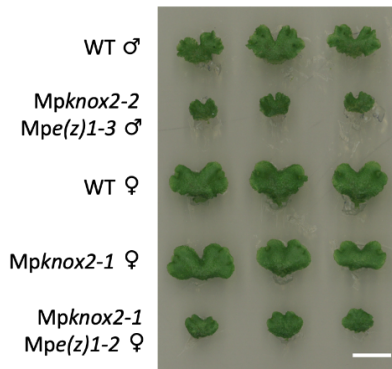
B



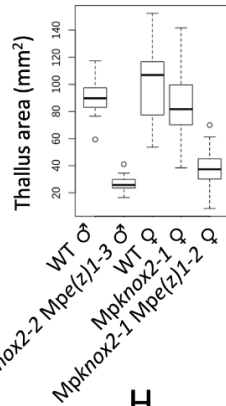
C



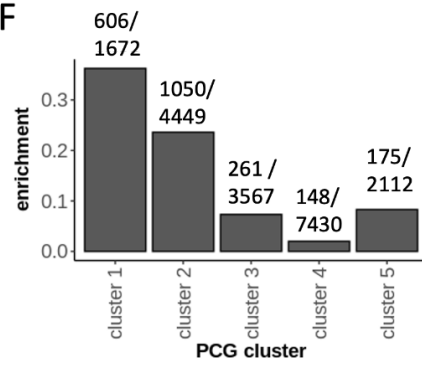
D



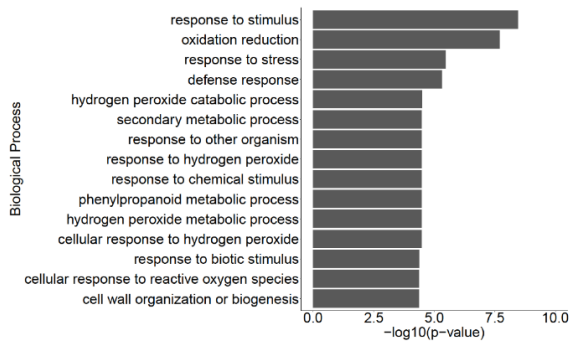
E



F



G



H

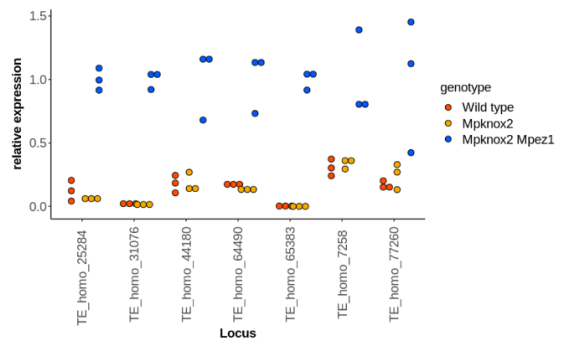


Figure S3 Generation of PRC2 knock out mutant in *M. polymorpha*

(A) Screenshots of genome browser showing H3K27me3 coverage in thalli (blue) and expression levels in thalli and sporophyte (grey) of Mp*KNOX2* (left) and Mp*BELLI*.

(B), (C) Gene structure of Mp*KNOX2* (B) and Mp*E(z)I* (C). Black lines, grey boxes, black boxes and dotted lines indicate intergenic regions, UTRs, exons and introns, respectively. Regions coding conserved motifs and domains are indicated by white boxes. In sequence alignment, guide RNA target sites are shown in blue and mismatches between wild type and mutants are shown in red.

(D) Images of 14 day old vegetative thalli in each genotype. Three independent plants are shown.

(E) Boxplot showing thallus area of each genotype. n>30

(F) Bar plot indicating enrichment of up-regulated PCGs in each PCG cluster, calculated as the number of PCGs exhibiting increased expression in each cluster divided by the total number of PCGs in each cluster.

(G) GOterm enrichment analysis using PCGs directly repressed by PRC2. Top 15 biological processes are indicated.

(H) Real time RT-PCR analysis confirming increased expression of several intact TE loci in Mp*knox2* Mp*e(z)I*. Expression levels of each TE are normalized against the expression level of the Mp*EFI* gene. In each TE, expression levels are normalized using the expression level in the Mp*knox2* Mp*e(z)I* mutant. Measurements of three biological replicates are plotted.

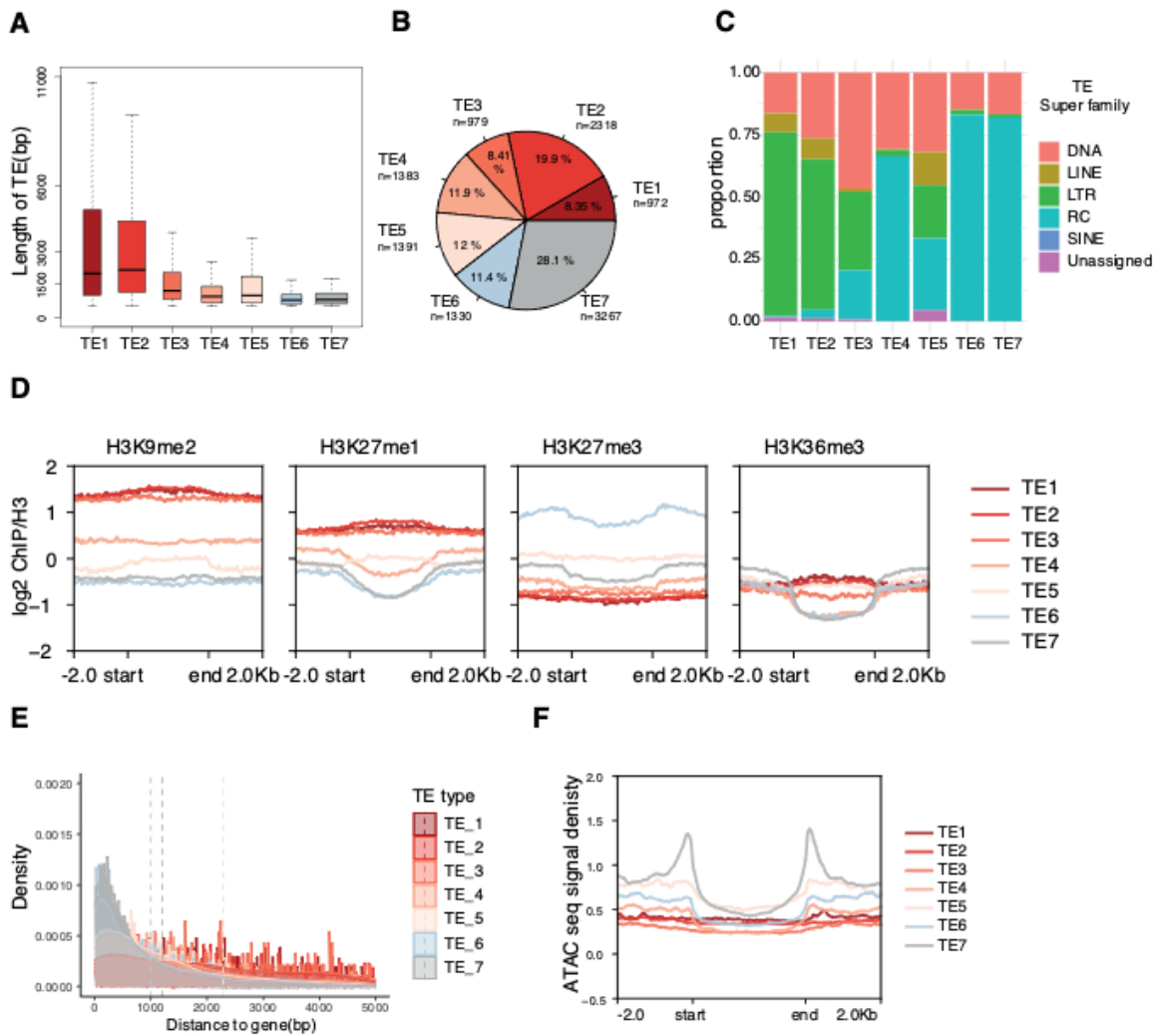


Figure S4 TE clusters in *A. thaliana*

- (A) Boxplot showing length distribution across different groups of TEs defined in ³².
- (B) Pieplot showing the genomic abundance of each group of TEs (TE>500 bp).
- (C) Barplot comparing the proportions of TEs within each group of TEs w.r.t TE super families.
- (D) Aggregate profile plots showing log₂ ChIP/H3 enrichment for various chromatin modifications on each group of TEs.
- (E) Density plot and histogram showing the distance between TEs and their closest PCG.
- (F) Accessibility for each group of TEs.

Table S1. oligos used in this study

name	seq5->3	description
E(z)_KO_F1	cggtagccggggatcCGTTCACCCGACATATGGTTCG	To generate Cmez-1 mutant
E(z)_KO_R1	cgactctagaggatcCTCGCTGTTCTATTCGATAGAGG	To generate Cmez-1 mutant
E(z)-KO_F2	agtcagctgctagggCAAGATCCGTGAAGAGAGACGC	To generate Cmez-1 mutant
E(z)_KO_R2	ttcgcccctcagttcATCAATTTTGTGGGGCGCTTTTC	To generate Cmez-1 mutant
URA_F	GAACTGAGGGGCGAACGC	To generate Cmez-1 mutant
URA_R	CCCTAGCAGCTGACTGTATCTC	To generate Cmez-1 mutant
pUC19_F	GCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTT CCC	To generate Cmez-1 mutant
pUC19_R	TTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCG G	To generate Cmez-1 mutant
5'UTR_E(z)_for	CTTATCTCAATATTTGGCCATCGTTTCGGTATCTT	To generate Cmez-2 mutant
5'UTR_E(z)_rev	AACAATCACCAATTTGATCAATTTTGTGGGGCGCTT	To generate Cmez-2 mutant
3'UTR_E(z)_for	AGTTGAAGTATGTTACAAGATCCGTGAAGAGAGACG	To generate Cmez-2 mutant
3'UTR_E(z)_rev	ATGTTAAGTGGATTACGCAAAGAAAGGAGAAAAATGG AA	To generate Cmez-2 mutant
Cm_trafo_H	TCTTCGTCGTCGTCAATCCAGCG	To generate Cmez-2 mutant
Cm_trafo_I	TCGCAGAGACTCGTTGGGGAAT	To generate Cmez-2 mutant
Cm_trafo_M	TTTGTCGTATCCGGCAAGTTTGC	To generate Cmez-2 mutant
Cm_trafo_N	CGCAAAGCAGGGACATTTCTTG	To generate Cmez-2 mutant
Cm_trafo_A	ggccatcgttcgggatctt	To generate Cmez-2 mutant
Cm_trafo_B	gcaaagaaaggagaaaaatggaatcttgg	To generate Cmez-2 mutant
TH637	ctcgGGATACCAAAACCATTG	gRNA targetting MpKNOX2
TH638	aaacCAATGGTTTTGGTATCC	gRNA targetting MpKNOX2
MpEz1-gRNA-1-Fw	ctcgTGTGGCTCCTCCACTAAGAG	gRNA targetting MpEz(1)
MpEz1-gRNA-1-Rv	aaacCTCTTAGTGGAGGAGCCACA	gRNA targetting MpEz(1)
MpEz1-gRNA-4-Fw	ctcgGTAACAACATGACTATCT	gRNA targetting MpEz(1)
MpEz1-gRNA-4-Rv	aaacAGATAGTCATGTTTGTTTAC	gRNA targetting MpEz(1)
TH650	CCTACGACTGCACCTCCCGAGT	for Mpez1 genotyping
TH651	AAAACGCACCCACCCAGCTAC	for Mpez1 genotyping
TH652	GTGCAGCAAGTGGAGGCAGGAG	for Mpknox2 genotyping
TH653	CATCAGGCTGGGGGCATTTGGG	for Mpknox2 genotyping

TH749	CTCTTCGTCCAGTGGGAGGC	TE_homo_64490-Fw
TH750	AGTTCTGGCAGTATTTAAGGAAGGG	TE_homo_64490-Rv
TH751	GTGAGCCGACAGAGTTGAACTCTCC	TE_homo_77260-Fw
TH752	CACTTGGAACGCAAGAGGATTGTGC	TE_homo_77260-Rv
TH761	CCGCTGAAGGGTGAGACAAACAAGG	TE_homo_25284-Fw
TH762	CTCTAGCACCCGAAACTCTGTGTCC	TE_homo_25284-Rv
TH696	CTCCGTGTTGCTTATAAGGTAGCTC	TE_homo_31076-Fw
TH697	CCCATGAGGCACCACTATTAAATTTTG	TE_homo_31076-Rv
TH698	TACAAGCTTGGCCTCGGAATAAATC	TE_homo_44180-Fw
TH699	AGCCATGCATTGTTTGGAAATTTCC	TE_homo_44180-Rv
TH708	GCCTTTCCGTCCCGAGAACAGCTC	TE_homo_65383-Fw
TH709	CGCCATGCACCGAGCTCTGAATTTG	TE_homo_65383-Rv
TH702	GGCTAAGAACTTCAGGAAGTACTC	TE_homo_7258-Fw
TH703	GCATCGCTGGAAGATGAGCG	TE_homo_7258-Rv
MpEF1a-qPCR-Fw	AAGCCGTCGAAAAGAAGGAG	Mp3g23400
MpEF1a-qPCR-Rv	TTCAGGATCGTCCGTTATCC	Mp3g23400