iScience, Volume *26*

Supplemental information

Potential role of N^6-adenine DNA methylation in alternative splicing and endosymbiosis in *Paramecium bursaria*

Bo Pan, Fei Ye, Tao Li, Fan Wei, Alan Warren, Yuanyuan Wang, and Shan Gao

Supplemental figures S1-S9.

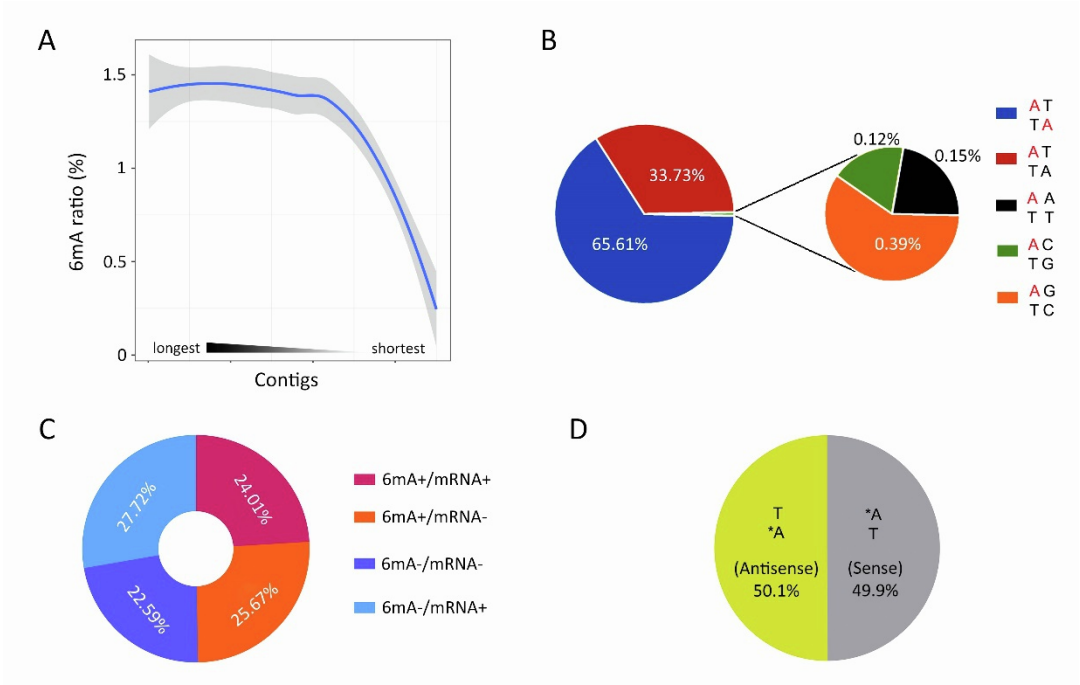

Figure S1. Genomic distribution of 6mA in *Paramecium bursaria*, related to Figure 1-2. (A) 6mA ratio (6mA/A) is plotted across the assembled MAC genome according to the contig length from long to short. (B) Distribution ratios of 6mA (marked by red) on various dinucleotide motifs across the genome. (C) Correlation between 6mA sites on DNA strands and transcription strands. (D) The ratios of 6mA on Watson or Crick strand, calculated by 6mA on a particular strand divided by 6mA on both strands.

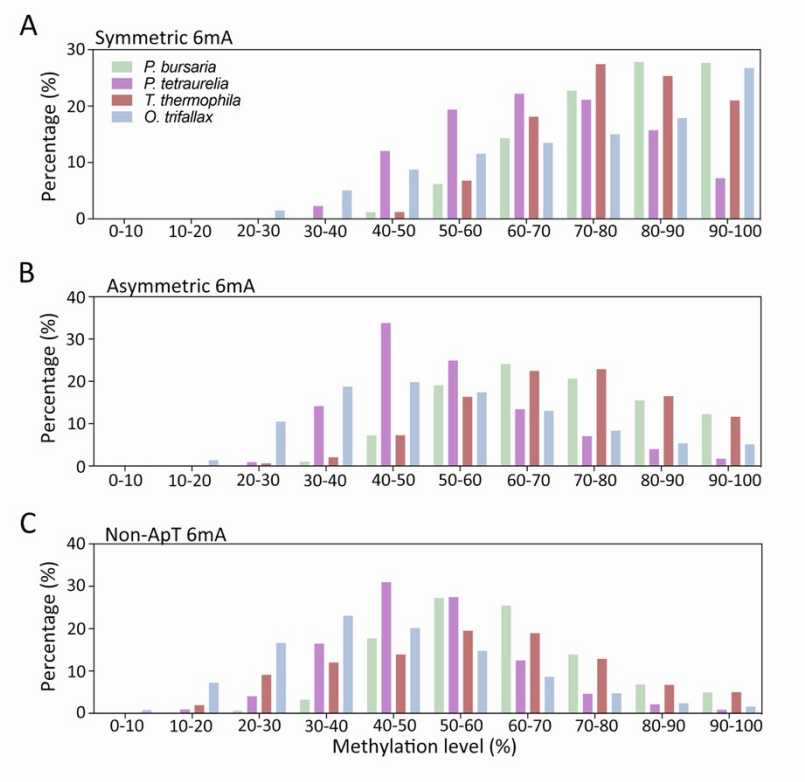

Figure S2. Methylation levels of 6mA in *P. bursaria* (green), *P. tetraurelia* (purple), *T. thermophila* (red) and *O. trifallax* (blue), related to Figure 2. 6mA is divided into symmetric (A), asymmetric (B) and non-ApT (C) 6mA. Methylation levels are ranked from low to high and divided into 10 quantiles. The percentage represents 6mA type/all 6mAs methylation.

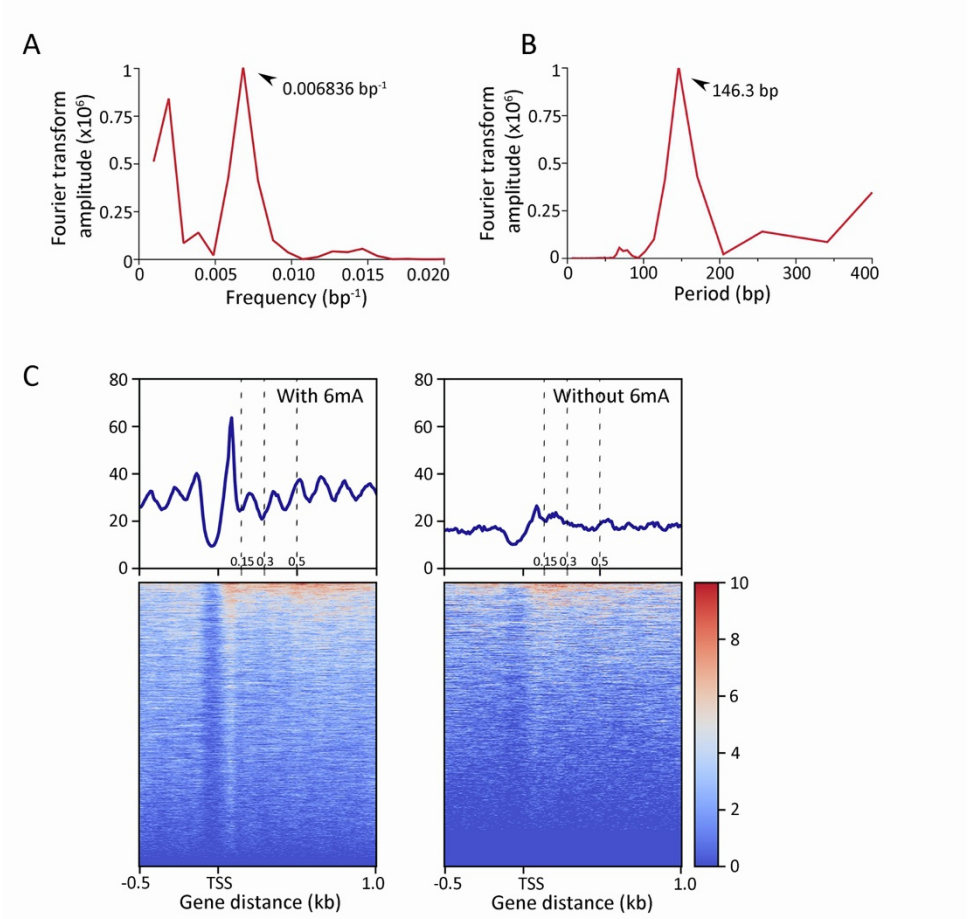

Figure S3. Fourier analysis of the periodic profile of downstream and upstream 6mA peaks, related to Figure 3. (A) Fourier transformation of 6mA distribution peaks. (B) Periods of the corresponding frequency in Fourier transformation. The dominant period length is 146.3 bp. (C) Nucleosomes present a more robust phase relative to TSS in genes with 6mA (left) than genes without 6mA (right). The plot was generated by deeptools with --MNase option. The locations of downstream TSS 150, 300 and 500bp were denoted by dashed lines, respectively.

A

No.	Location	Strand	IPD ratio	Sequence	Methylation level
1	6734	-	2.07	AAAAAAAAGAGGGGGTAGCGACTGTTTATTAAAAACATAAG	43%
2	7883	-	1.72	TTCGATTCTTTAAGGCGGCTAAGTAAATATTATTATTAAAT	28%
3	24227	+	2.39	ATCATTAAACGATTTAAATAAGGTTCATATCTAAAATATGA	49%

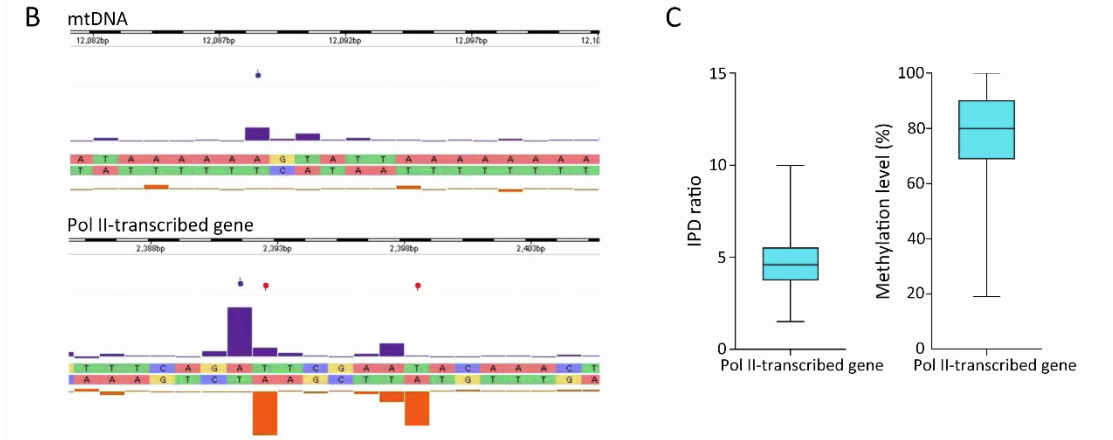

Figure S4. 6mA features in mtDNA of *P. bursaria*, related to Figure 3. (A) Detailed information of three high-confidence 6mA sites in mtDNA. (B) Comparison of representative kinetic signature performed by SMRT sequencing data of vegetative *P. bursaria* cells in mtDNA and Pol II-transcribed genes. Columns indicate the IPD peaks and balloons indicate the detected 6mA signals on both the sense (blue) and antisense (red) strands based on the IPD ratio. (C) The distribution of IPD ratio and 6mA methylation level of Pol II-transcribed genes. Data distributions are shown by boxplots.

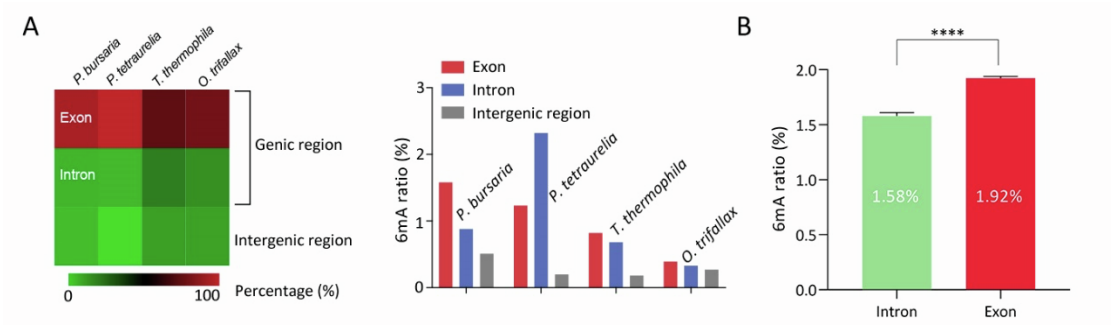

Figure S5. 6mA is preferentially located at exonic regions in *P. bursaria*, related to Figure 4. (A) Left: Heatmap of 6mA distribution on genomic features in *P. bursaria*, *P. tetraurelia*, *T. thermophila* and *O. trifallax*. Right: 6mA ratio (total 6mA sites number/ total A number of corresponding sequence) of exon, intron and intergenic region in three ciliates. (B) 6mA ratio (6mA sites number/ total A number of corresponding sequence) of individual intronic and exonic sequences in *P. bursaria*. Significant differences are labelled by * (Student's *t*-test was performed). $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; ns: not significant, $P > 0.05$.

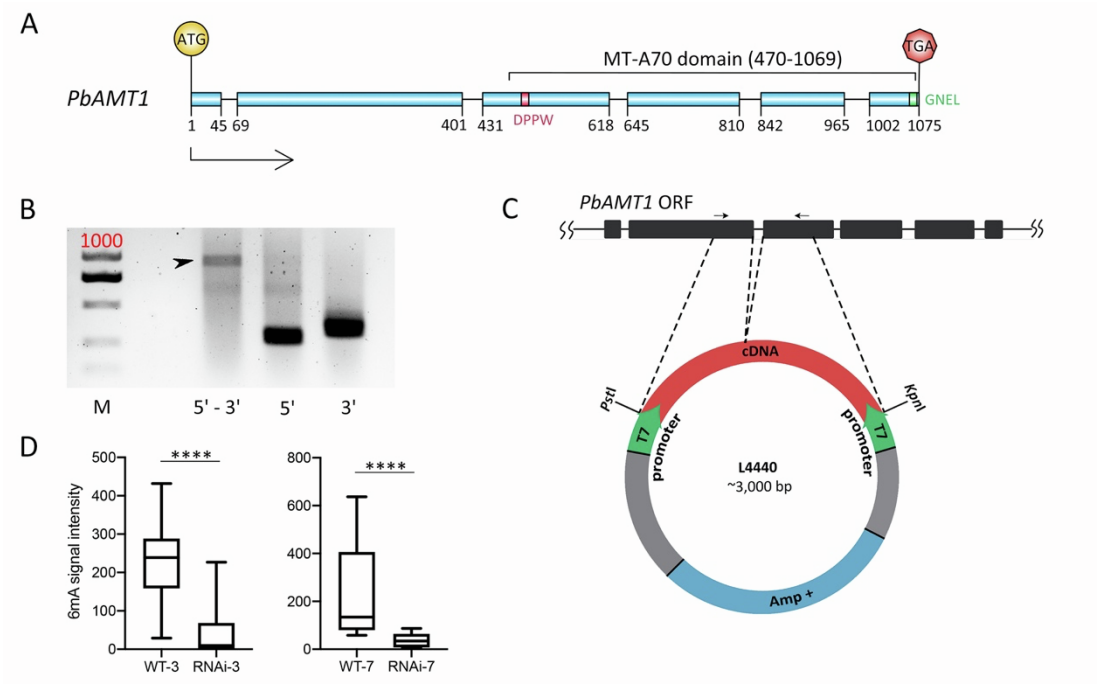

Figure S6. *PbAMT1* gene model detection and RNAi knockdown construct, related to Figure 6. (A) *PbAMT1* gene model is presented. MT-A70 region is delineated and DPPW motif and GNEL motif are highlighted by red and green, respectively. (B) RT-PCR assay using primers flanking *PbAMT1* gene body (lane M: marker; 5'-3': DNA fragment from ATG to TGA; lane 5': ~200 bp DNA fragment downstream of ATG; lane 3': ~400 bp DNA fragment upstream of TGA). (C) Schematic representation of the endogenous *PbAMT1* locus in the genome and the RNAi knockdown construct. (D) Statistical analysis of 6mA IF signal intensity in another replicate of RNAi experiment (3 days and 7days). Cell images were randomly selected and processed by ImageJ. Data are shown as box plots. Student's *t*-test was performed. $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; ns: not significant, $P > 0.05$.

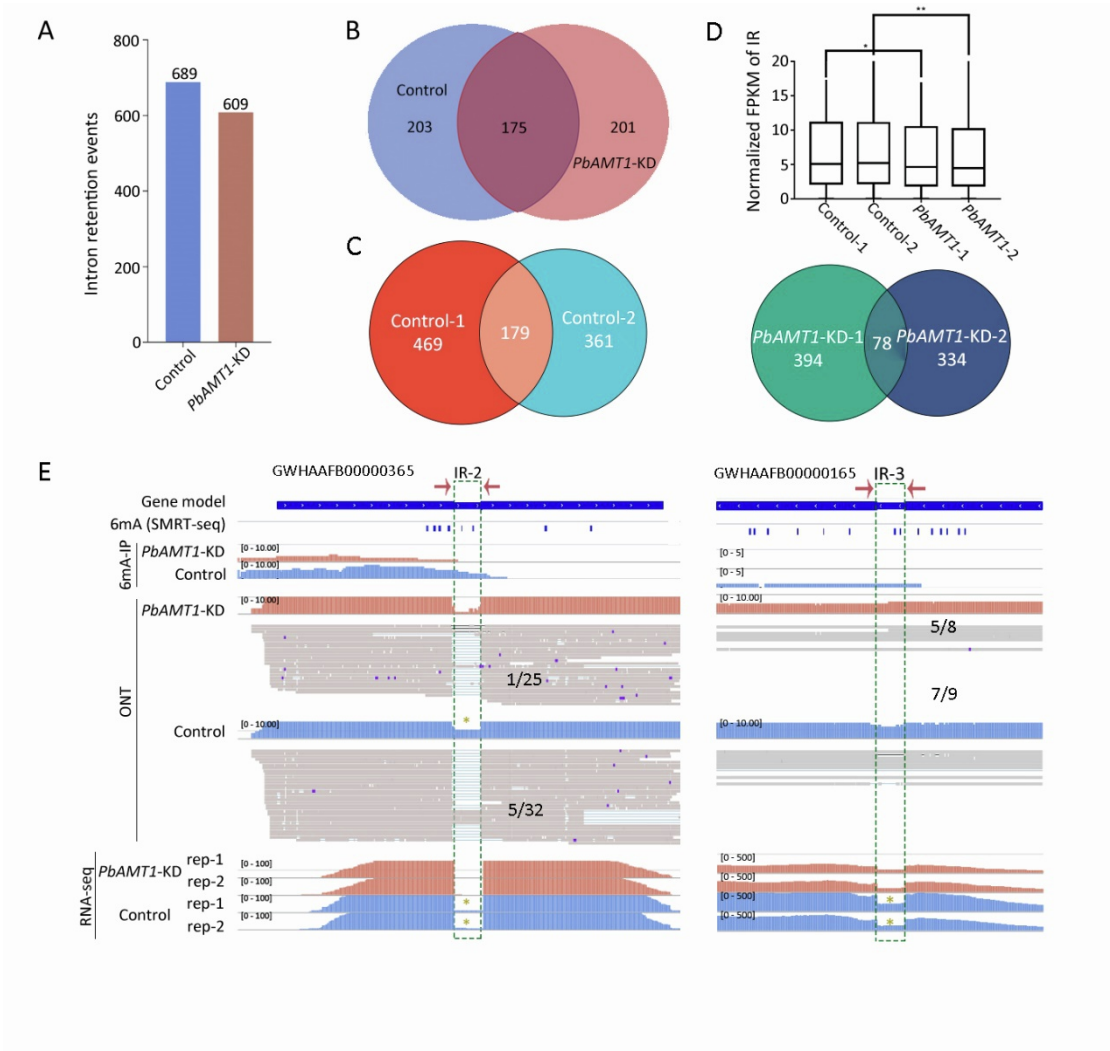

Figure S7. 6mA is associated with intron retention, related to Figure 6. (A) Intron retention events occurred in control and *PbAMT1*-KD cells by ONT cDNA-seq. (B) The overlap of genes harboring intron retention between control and *PbAMT1*-KD cells by ONT cDNA-seq. (C) The overlapped intron retention events between control and *PbAMT1*-KD cells by RNA-seq with two replicates. (D) Intron retention (IR) events are reduced in *PbAMT1*-KD cells. Student's *t*-test was performed. $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; ns: not significant, $P > 0.05$. (E) Alternative splicing events and 6mA distribution on a representative gene locus IR-2 and IR-3 in control and *PbAMT1*-KD cells, respectively. The region of intron retention is highlighted by the dashed green box and the yellow asterisk. All the compared results are set the same data range.

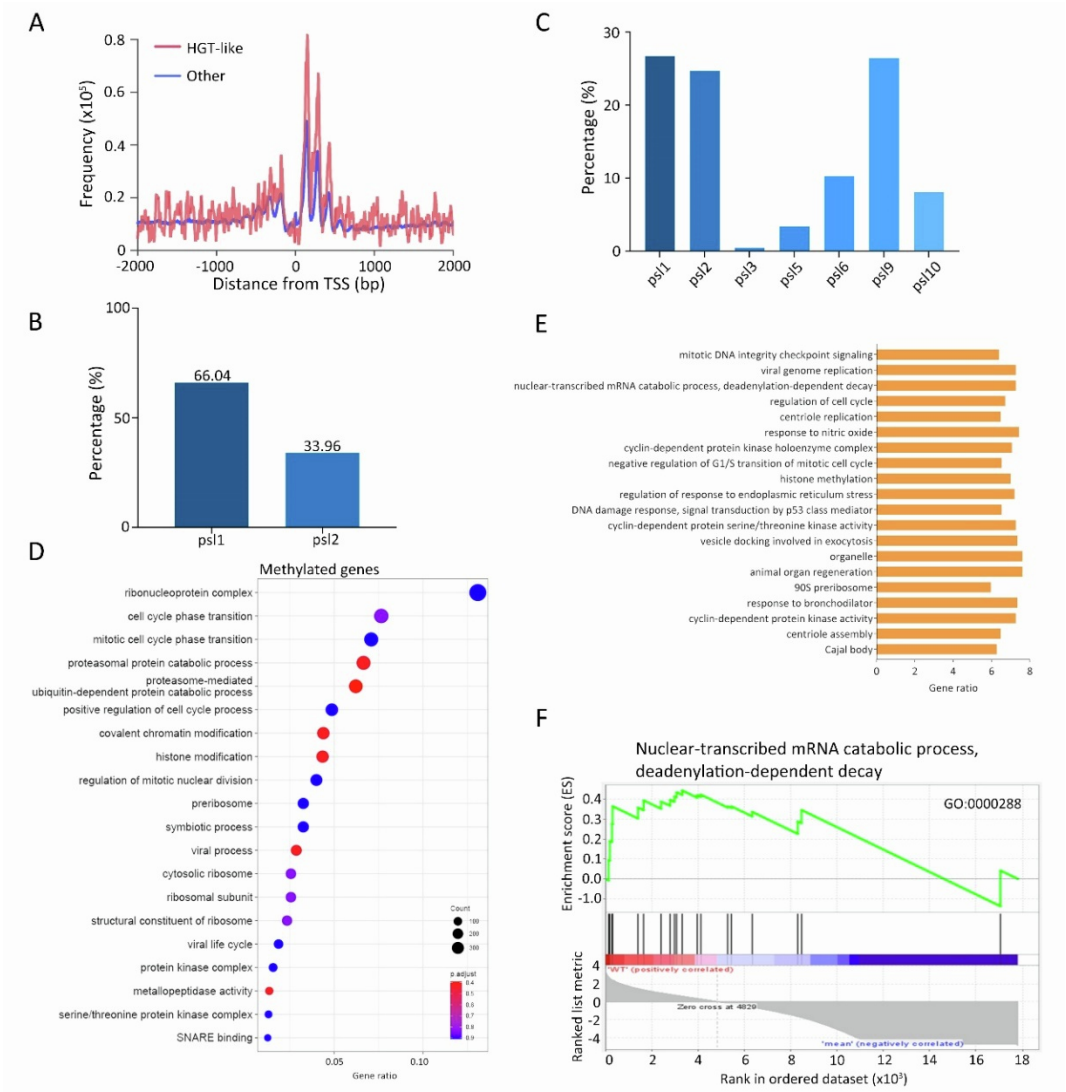

Figure S8. Endosymbiosis-correlated genes harbor lower 6mA levels but putative HGT-like genes from symbiotic algae have higher 6mA levels, related to Figure 7. (A) Distribution profiles of 6mA in HGT-like genes and other genes around transcription start sites (TSS) in *P. bursaria*. Note the higher 6mA level of HGT-like genes than other genes. (B) The percentage of gene number of different HGT-like-gene age groups. (C) The percentage of gene number of different all-gene age groups. (D) Gene ontology enrichment analysis of methylated genes. The R package ggplot2 were performed. The dot size represents the gene number and the color bar shows the *P*-value. (E) Bar plot of enriched pathways of the top 20-ranked high 6mA-level genes. ~67% genes of *P. bursaria* are methylated, which were spread among the various pathways. (F) The GSEA analysis of genes bearing 6mA enriched in the nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay pathways.

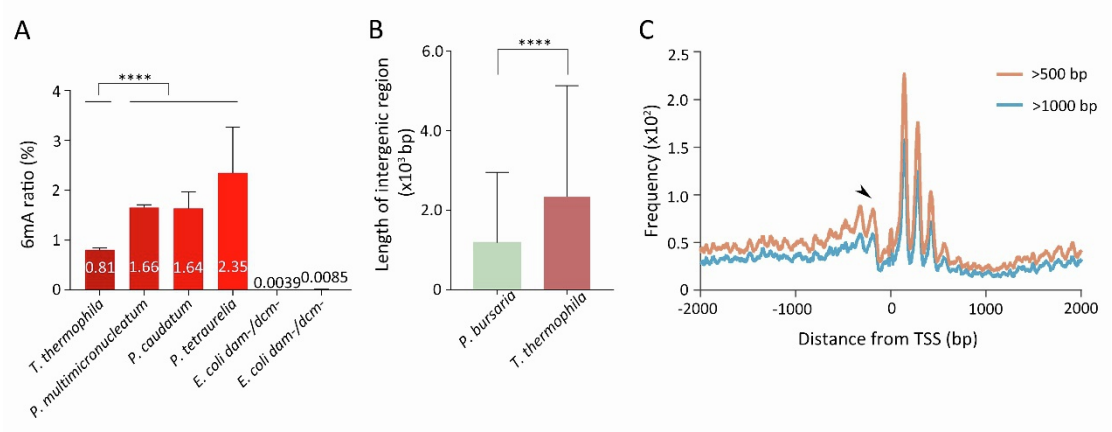

Figure S9. Unique features of 6mA in genus *Paramecium*, related to Figure 7. (A) Mass spectrometry analysis of 6mA, performed on three biological replicates for each species. Data are presented as bar plots. The strains used to detect are *T. thermophila* SB210 and *P. tetraurelia* 51. Other species of *Paramecium* genus which were detected by LC-MS are collected in wild environment by our lab mates. Student's *t*-test was performed. $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; ns: not significant, $P > 0.05$. (B) The distribution of length of intergenic region in *P. bursaria* (green) and *T. thermophila* (red). The y-axis represents the distance between genes. Student's *t*-test was performed. $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; ns: not significant, $P > 0.05$. (C) Distribution profiles of 6mA in the long-distanced genes (> 500 bp or > 1,000 bp) in *P. bursaria*.

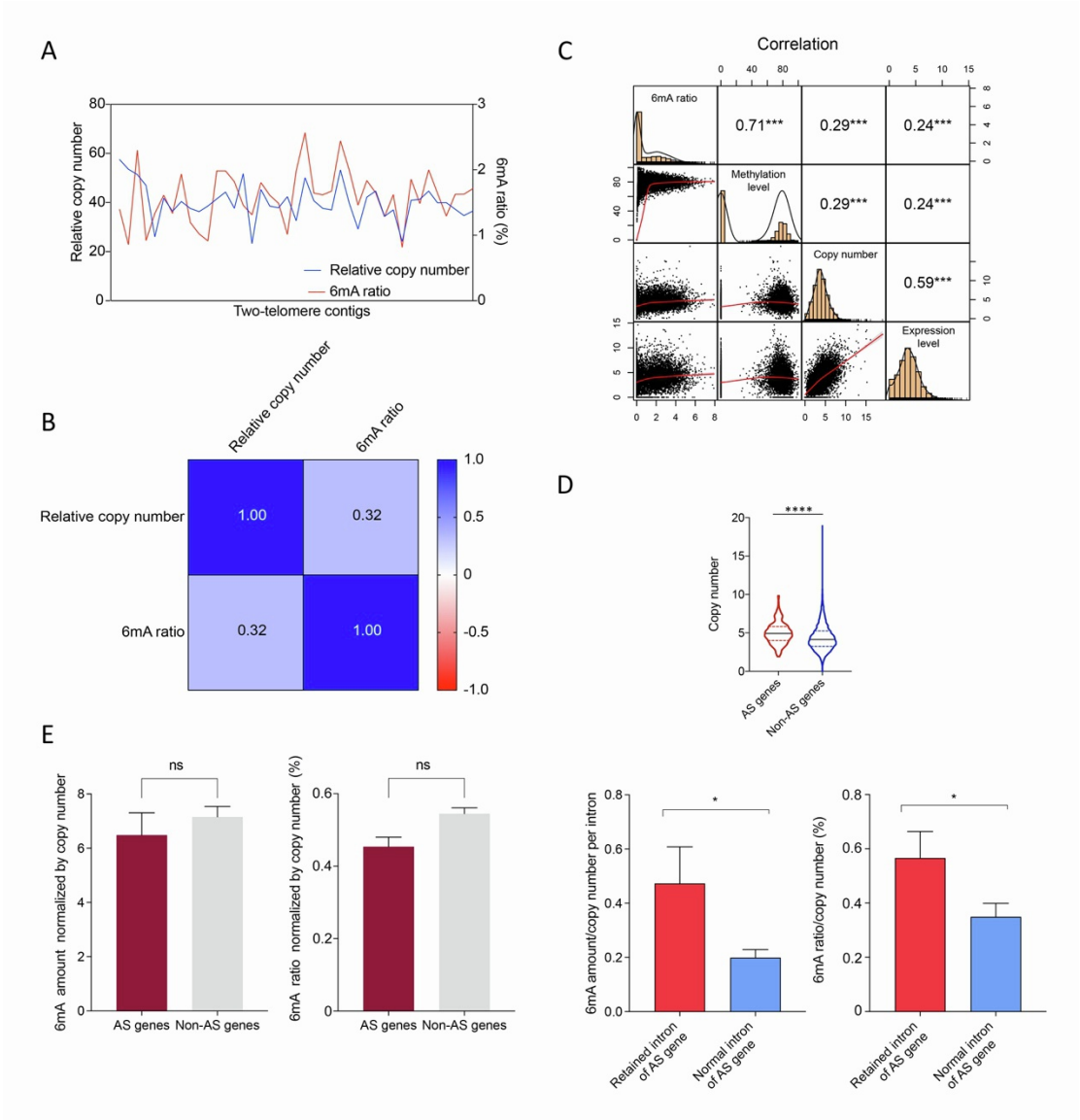

Figure S10. The relationship between 6mA and copy number, related to Figure 7. (A) Different chromosomes (two-telomere contigs) have different copy numbers and 6mA ratio. (B) A positive correlation between the copy number of two-telomere contigs/chromosomes and their 6mA ratio (r=0.32). (C) Correlation analysis among 6mA ratio, methylation level, copy number and expression level (FPKM) for genes longer than 1kb. (D) The copy number variances between AS and non-AS genes. Student's t-test was performed. $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; ns: not significant, $P > 0.05$. (E) The differences of 6mA normalized by gene copy number in AS and non-AS genes, retained intron and normal intron of AS genes, respectively. Student's t-test was performed. $^{****}P < 0.0001$; $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; ns: not significant, $P > 0.05$.

Table S1. Detailed information of 6mA in *P. bursaria*, related to Figure 1-2.

	Sites number	Percentage (%)
6mA sites	266,215	-
Symmetric	210,928 (105,464*2)	79.23
Asymmetric	54,230	20.37
Non-ApT	1,057	0.40
Intergenic	20,765	7.80
Genic	245,450	92.20
Exon	221,534	83.22
Intron	23,916	8.98
6mA ratio (6mA/A)	-	1.28

SMRT-seq	Coverage	Cut-off
P. bursaria	123×	30×
P. tetraurelia	400×	100×
T. thermophila	144×	36×
O. trifallax	278×	70×

Table S2. Methylated adenines distributed on four forms of ApN dinucleotides in three ciliate species, related to Figure 2.

	ApA	ApT	ApG	ApC	Species
Site numbers on genome	7,674,350	7,614,120	3,194,252	2,289,665	*P. bursaria*
Number of methylated adenines	239	265,158	627	191	
percentage (%)	0.003	**3.48**	0.02	0.01	
Site numbers on genome	14,550,842	18,640,072	7,633,241	5,515,156	*P. tetraurelia*
Number of methylated adenines	432	545,844	1,222	362	
percentage (%)	0.003	**2.93**	0.02	0.01	
Site numbers on genome	35,375,906	27,831,470	10,116,164	6,973,798	*T. thermophila*
Number of methylated adenines	27,251	382,863	15,840	10,322	
percentage (%)	0.08	**1.38**	0.16	0.15	
Site numbers on genome	17,265,715	14,632,462	8,342,515	5,769,414	*O. trifallax*
Number of methylated adenines	2,190	166,145	2,941	1,182	
percentage (%)	0.01	**1.14**	0.04	0.02	

Table S3. Comparison of 6mA in different categories in three ciliates, related to Figure 2.

	P. bursaria	*P. tetraurelia*	*T. thermophila*	*O. trifallax*
Symmetric (%)	79.23	80.86	61.06	81.27
Asymmetric (%)	20.37	18.77	26.70	14.71
Non-ApT (%)	0.40	0.37	12.24	4.02

Table S4. Comparison of 6mA methylation levels in *P. bursaria*, related to Figure 2.

Quantiles (Methylation level)	Sites number	Percentage (%)
0-10%	0	0
10-20%	2	0.001
20-30%	65	0.02
30-40%	726	0.27
40-50%	6,595	2.48
50-60%	23,700	8.90
60-70%	43,504	16.34
70-80%	59,314	22.28
80-90%	67,201	25.24
90-100%	65,108	24.46

Table S5. Pol I- and Pol III-transcribed genes of *P. bursaria*, related to Figure 3.

Contig	Source	Begin	End	Type	RNA polymerases
GWHAAFB00000380	RNAmmer	11822	15403	28S rRNA	Pol I
GWHAAFB00000380	RNAmmer	2856	6437	28S rRNA	Pol I
GWHAAFB00000380	RNAmmer	20451	24028	28S rRNA	Pol I
GWHAAFB00000072	RNAmmer	156482	156596	5S rRNA	Pol III
GWHAAFB00000074	RNAmmer	35575	35689	5S rRNA	Pol III
GWHAAFB00000072	RNAmmer	158302	158416	5S rRNA	Pol III
GWHAAFB00000074	RNAmmer	33780	33894	5S rRNA	Pol III
GWHAAFB00000072	RNAmmer	160099	160213	5S rRNA	Pol III
GWHAAFB00000052	RNAmmer	90916	91030	5S rRNA	Pol III
GWHAAFB00000074	RNAmmer	30903	31017	5S rRNA	Pol III
GWHAAFB00000072	RNAmmer	153565	153679	5S rRNA	Pol III
GWHAAFB00000380	RNAmmer	984	2709	18S rRNA	Pol I
GWHAAFB00000380	RNAmmer	18579	20304	18S rRNA	Pol I
GWHAAFB00000380	RNAmmer	9950	11675	18S rRNA	Pol I
GWHAAFB00000001	tRNAscan	1861	1941	tRNA	Pol III
GWHAAFB00000001	tRNAscan	40347	40276	tRNA	Pol III
GWHAAFB00000001	tRNAscan	25403	25333	tRNA	Pol III
GWHAAFB00000001	tRNAscan	5632	5552	tRNA	Pol III
GWHAAFB00000002	tRNAscan	241346	241274	tRNA	Pol III
GWHAAFB00000003	tRNAscan	165448	165377	tRNA	Pol III
GWHAAFB00000003	tRNAscan	164748	164678	tRNA	Pol III
GWHAAFB00000003	tRNAscan	164447	164376	tRNA	Pol III
GWHAAFB00000011	tRNAscan	64240	64168	tRNA	Pol III
GWHAAFB00000011	tRNAscan	63972	63900	tRNA	Pol III
GWHAAFB00000011	tRNAscan	63704	63632	tRNA	Pol III
GWHAAFB00000012	tRNAscan	87431	87503	tRNA	Pol III
GWHAAFB00000012	tRNAscan	115565	115647	tRNA	Pol III
GWHAAFB00000015	tRNAscan	16719	16637	tRNA	Pol III
GWHAAFB00000017	tRNAscan	44377	44297	tRNA	Pol III
GWHAAFB00000027	tRNAscan	7134	7206	tRNA	Pol III
GWHAAFB00000027	tRNAscan	7542	7613	tRNA	Pol III
GWHAAFB00000039	tRNAscan	5947	5867	tRNA	Pol III
GWHAAFB00000045	tRNAscan	7923	7853	tRNA	Pol III
GWHAAFB00000050	tRNAscan	97471	97391	tRNA	Pol III
GWHAAFB00000050	tRNAscan	97310	97239	tRNA	Pol III
GWHAAFB00000050	tRNAscan	96945	96864	tRNA	Pol III
GWHAAFB00000057	tRNAscan	118002	117931	tRNA	Pol III
GWHAAFB00000057	tRNAscan	117838	117766	tRNA	Pol III
GWHAAFB00000057	tRNAscan	117546	117475	tRNA	Pol III
GWHAAFB00000059	tRNAscan	24240	24320	tRNA	Pol III

GWHAAFB00000059	tRNAscan	24411	24491	tRNA	Pol III
GWHAAFB00000060	tRNAscan	53584	53655	tRNA	Pol III
GWHAAFB00000060	tRNAscan	53736	53806	tRNA	Pol III
GWHAAFB00000060	tRNAscan	54077	54147	tRNA	Pol III
GWHAAFB00000072	tRNAscan	44775	44847	tRNA	Pol III
GWHAAFB00000072	tRNAscan	45044	45116	tRNA	Pol III
GWHAAFB00000077	tRNAscan	14056	13985	tRNA	Pol III
GWHAAFB00000081	tRNAscan	87782	87711	tRNA	Pol III
GWHAAFB00000086	tRNAscan	108650	108732	tRNA	Pol III
GWHAAFB00000091	tRNAscan	58875	58945	tRNA	Pol III
GWHAAFB00000096	tRNAscan	29092	29008	tRNA	Pol III
GWHAAFB00000096	tRNAscan	28850	28766	tRNA	Pol III
GWHAAFB00000096	tRNAscan	28442	28371	tRNA	Pol III
GWHAAFB00000096	tRNAscan	28118	28045	tRNA	Pol III
GWHAAFB00000120	tRNAscan	15552	15623	tRNA	Pol III
GWHAAFB00000120	tRNAscan	15814	15886	tRNA	Pol III
GWHAAFB00000124	tRNAscan	99067	98996	tRNA	Pol III
GWHAAFB00000126	tRNAscan	89709	89638	tRNA	Pol III
GWHAAFB00000128	tRNAscan	74294	74366	tRNA	Pol III
GWHAAFB00000130	tRNAscan	14792	14865	tRNA	Pol III
GWHAAFB00000136	tRNAscan	46608	46535	tRNA	Pol III
GWHAAFB00000147	tRNAscan	17925	17997	tRNA	Pol III
GWHAAFB00000147	tRNAscan	21255	21327	tRNA	Pol III
GWHAAFB00000152	tRNAscan	55529	55458	tRNA	Pol III
GWHAAFB00000152	tRNAscan	55328	55257	tRNA	Pol III
GWHAAFB00000152	tRNAscan	55164	55094	tRNA	Pol III
GWHAAFB00000152	tRNAscan	23605	23534	tRNA	Pol III
GWHAAFB00000152	tRNAscan	23346	23274	tRNA	Pol III
GWHAAFB00000155	tRNAscan	98579	98658	tRNA	Pol III
GWHAAFB00000158	tRNAscan	90146	90073	tRNA	Pol III
GWHAAFB00000158	tRNAscan	89961	89891	tRNA	Pol III
GWHAAFB00000183	tRNAscan	16159	16230	tRNA	Pol III
GWHAAFB00000183	tRNAscan	16361	16432	tRNA	Pol III
GWHAAFB00000183	tRNAscan	16525	16596	tRNA	Pol III
GWHAAFB00000187	tRNAscan	44338	44411	tRNA	Pol III
GWHAAFB00000190	tRNAscan	46780	46707	tRNA	Pol III
GWHAAFB00000205	tRNAscan	62948	62877	tRNA	Pol III
GWHAAFB00000217	tRNAscan	22894	22964	tRNA	Pol III
GWHAAFB00000223	tRNAscan	30029	30102	tRNA	Pol III
GWHAAFB00000223	tRNAscan	10822	10750	tRNA	Pol III
GWHAAFB00000233	tRNAscan	2660	2731	tRNA	Pol III
GWHAAFB00000233	tRNAscan	2829	2899	tRNA	Pol III
GWHAAFB00000233	tRNAscan	25125	25054	tRNA	Pol III

GWHAAFB00000238	tRNAscan	15613	15540	tRNA	Pol III
GWHAAFB00000244	tRNAscan	6607	6537	tRNA	Pol III
GWHAAFB00000252	tRNAscan	24168	24096	tRNA	Pol III
GWHAAFB00000257	tRNAscan	50653	50582	tRNA	Pol III
GWHAAFB00000275	tRNAscan	1776	1703	tRNA	Pol III
GWHAAFB00000301	tRNAscan	14317	14237	tRNA	Pol III
GWHAAFB00000315	tRNAscan	16928	16845	tRNA	Pol III
GWHAAFB00000315	tRNAscan	16673	16600	tRNA	Pol III
GWHAAFB00000315	tRNAscan	16472	16399	tRNA	Pol III
GWHAAFB00000324	tRNAscan	18670	18600	tRNA	Pol III
GWHAAFB00000329	tRNAscan	7897	7825	tRNA	Pol III
GWHAAFB00000346	tRNAscan	38412	38325	tRNA	Pol III
GWHAAFB00000348	tRNAscan	31827	31756	tRNA	Pol III
GWHAAFB00000393	tRNAscan	53606	53686	tRNA	Pol III
GWHAAFB00000393	tRNAscan	107983	108055	tRNA	Pol III

Table S6. Comparison of 6mA distribution on different genomic features in three ciliates, related to Figure 4.

	P. bursaria	*P. tetraurelia*	*T. thermophila*	*O. trifallax*
Intergenic	7.80%	3.09%	13.45%	12.87%
Genic	92.20%	96.91%	86.55%	87.13%
Exon	83.22%	88.64%	65.98%	70.63%
Intron	8.98%	8.27%	20.58%	16.50%

Table S7. Sequence of *PbAMT1* in *P. bursaria*, related to Figure 5.

>Protein sequence

MISKQKGNKKSKIIEDEDQIQKKVKLDDDQSVSQILEEEDIDSEEIEGLDVSEEEKEPIE
DNNYIFIDKLPNQETALLKFQEQVRNRIKFYKQRYLQECYENQTKKSFVNDDAIPICA
DVRYLDFQKLHDSQLQIAGRLFDVIMMDPPWQLSSSQPSRGVAIAYQSLSDDMISKM
PIQSLQDDGIILIWTINAKYKVTCKLMENWGYKIVDEIVWVKKTVNGKIAKGHGFYL
QHAKENCLVGVKGNLDWERCKKSVASDVIFSERRGQSQKPEEIYKYVEELIPNGHYL
EIFGRRNNIRKNWVTIGNEL

>Gene (introns were labelled by orange)

ATGATCTCCAAACAAAAGGGAAACAAAAAATCGAAAATAATAGAGGTATTGAGCA
TGAATATAGGAAGGATGAGGATTAGATATAAAAAAAAGTAAAGCTAGACGATGACT
AAAGTGTATCCTAAATTCTTGAAGAGGAGGATATCGACTCTGAAGAAATTGAAGG
ATTGGATGTATCGGAGGAAGAAAAAGAGCCGATCGAAGACAATAATTATATCTTTA
TAGACAAATTACCTAATCAAGAGACAGCCTTATTAAAATTTCAGGAATAGGTCAGG
AATAGAATTAAGTTTTACAAACAACGATATCTCTAGGAGTGCTATGAGAATTAGAC
GAAAAAAAGTTTTGTAAATGATGATGCTATACCAATTTGTGCTGATGTTAGATATCT
TGACTTTTAAGTATTTTAATTAAAAAAAAAATATTCAAGAAATTACATGATTCACAA
CTTCAAATAGCTGGAAGGCTTTTTGATGTTATTATGATGGACCCTCCATGGTAACTC
TCAAGTTCATAACCATCTAGAGGTGTTGCTATTGCTTATTAATCTTTATCAGACGATA
TGATAAGTAAAATGCCCATTTAATCATTATAGGATGATGGAATCATTCTCATATGGTA
ATATTATATTCATAATTCAATAGGACTATCAATGCAAAATATAAGGTAACTTGCAAGC
TAATGGAAAATTGGGGATATAAAATTGTTGATGAAATTGTGTGGGTTAAGAAGACT
GTGAATGGTAAAATAGCAAAAGGACATGGATTTTATTTATAACATGCTAAAGAAAA
TTGCTTAGTAGGCGTTAAGGTATGATCAATAGAAAATTTTAAAATTTTAGGGCAATT
TGGATTGGGAAAGATGCAAAAAATCGGTAGCGAGTGATGTTATCTTCAGCGAACG
AAGAGGATAATCATAGAAACCTGAAGAAATTTATAAATATGTAGAAGAACTAATTC
CAAATGGTATTTTATATTTTTTAAAAAAACCTCTTATTTTAGGGCATTATTTGGAAAT
ATTCGGAAGAAGAAATAATATCAGGAAAAATTGGGTTACAATAGGAAATGAATTGT
GA

>CDS

ATGATCTCCAAACAAAAGGGAAACAAAAAATCGAAAATAATAGAGGATGAGGATT
AGATATAAAAAAAAGTAAAGCTAGACGATGACTAAAGTGTATCCTAAATTCTTGAA
GAGGAGGATATCGACTCTGAAGAAATTGAAGGATTGGATGTATCGGAGGAAGAAA
AAGAGCCGATCGAAGACAATAATTATATCTTTATAGACAAATTACCTAATCAAGAGA
CAGCCTTATTAAAATTTCAGGAATAGGTCAGGAATAGAATTAAGTTTTACAAACAA
CGATATCTCTAGGAGTGCTATGAGAATTAGACGAAAAAAAGTTTTGTAAATGATGA
TGCTATACCAATTTGTGCTGATGTTAGATATCTTGACTTTTAAAAATTACATGATTCA
CAACTTCAAATAGCTGGAAGGCTTTTTGATGTTATTATGATGGACCCTCCATGGTAA
CTCTCAAGTTCATAACCATCTAGAGGTGTTGCTATTGCTTATTAATCTTTATCAGACG
ATATGATAAGTAAAATGCCCATTTAATCATTATAGGATGATGGAATCATTCTCATATG

GACTATCAATGCAAAATATAAGGTAACTTGCAAGCTAATGGAAAATTGGGGATATA
AAATTGTTGATGAAATTGTGTGGGTTAAGAAGACTGTGAATGGTAAAATAGCAAA
AGGACATGGATTTTATTTATAACATGCTAAAGAAAATTGCTTAGTAGGCGTTAAGGG
CAATTTGGATTGGGAAAGATGCAAAAAATCGGTAGCGAGTGATGTTATCTTCAGCG
AACGAAGAGGATAATCATAGAAACCTGAAGAAATTTATAAATATGTAGAAGAACTA
ATTCCAAATGGGCATTATTTGGAAATATTCGGAAGAAGAAATAATATCAGGAAAAA
TTGGGTTACAATAGGAAATGAATTGTGA

Table S8. Sequence of PbAMTs 2-6 in *P. bursaria*, related to Figure 5.

>PbAMT2

MPLVLEDQVSVQLQYQKDNQIRELLMYHQQNQKQKINFLCQQQYQTSQYSMPYFR
AIPIASDFETPPLLLAYRSTLFPIKTLTYISVFSFSVIQCSIFSNVYRFETLFIHNALKYHQ
TYRQQQLLLGKIQRRYLVFYVYQVFPSTLSYIMYHRPCILQSLQIYFQYFRNFKYYLPS
KNVDFPTDFSPCLKKKSFLKHYTQKHDPKFEVIVFGFHFXXXXXXXXXXXSVQKEN
VYAFEILQRHQNKKNKFFMSEADQQLRKKENMKTVFNQIDLDNLVFQKRERKQVGQ
QTQIPQPVVKKTVVSKPKSTVKVQKAPQTERVQNEEPQYKQPVVMESESSEFDPKER
SKKLGGKKFRRTHHKFKENDSSDEEEYKKHLILKRVNKIFKKSYEQKQVLHTGSEQD
ILVRNLFKNLYEKIFQDLKIIQENSTEEIQGDNEKRILPEECTDMETRRNLKLLGVDNA
EYYSKLTDVKTYINCDIRYFNLDFLVEKVGSFDGGQQNDSSFMFTNSKFQLDYNTMS
NQEIMDIKIEKLSKKGFLFLWILNTQITIASEMVNKWGYEIVDQIVWVKLNAQGNNIY
LSTGYYFMHSYEMCLVGYKCPQGDHVEYHSKVSNNIIFSPVRNKSQKPLELYEIIELM
MPGAKRLEIFARNHNLRHGWFSIGNQLGETYQKWHIQINCNSCPTSLQPGIARFKSKR
VANFDICQKCYEEKLQQGEFTENDFFHLANKADEEVLHQYHQCNICEQEPIWGTRFE
CVTCENYDLCEACFDHTLTQEEQQHKDHDFRAIELPIFAQGIPCHDAKCNSCFQKPIL
GVKIAFSLVTQPNQPVKDINLIIEWNLSMSHKTXXXXXXXXXXXXXXXXQGCEIEVS
GGIYKCENCFGFYFCDQCHALKKDDWKCYVATSHKNYHTFIKIQ

>PbAMT3

MNQKLPDLCMNFINMNERPQNFVKKYKPEERFEDYPKAKELIRLKSELIKKRNHPPA
YLKEDLRSFDLTKLGKFDVILIDPPWAEYTRRFLQANMQVKEHQQSWSLDELKSLQIE
KLADIPSFLFLWCGSEHLDDGRELFKHWGYKRCEDIVWLKTNRDISKQNQYVAGHD
YGDNVLRRVKEHCLVGLKGDVKRASDQHFIHANIDTDVLISEEEAMGSTKKPEELYD
VIERFCLGRRRLELFGEIHNIREGWYQIFQLKRLTIGSQLRDSRFSLQEYQSYFVQDRF
NEEKPFVGPKFIESTPEIENLRPKSPSKEQQQQQQQFSFSFGQM

>PbAMT4

MKYKNKKERSRSREQEKLDPLKYPPINQDQFRAVVIDYLMSDQFDYPISGNELYHCIN
SAFLSRQNTPYLRNLQIEPILKQLADTSAIQLSKLAVGTQLQDFILSISLEKLKQLNIQY
SKDRPKKAPTVEEWINEEIRQIAYGQTFQRIDHLLDEARNQKQNVSLLPKIAAQKKR
MLENIGNISLKYEDNVVQLALERKRAALLKTENYDEEEARRKLNERGLVICEYCSRD
RCRNVSCQKTHFRKIVKANTQTKLGNCSNLDQCPEQETCKYIHYILDGNDHDWQKRI
KRSIPPQWINCDLRIFDFRVLGKFDVIMADPPWDIHMNLPYGTLKDKEMKALRVDLL
QDDGMIFLWVTGRAMELGRECLSLWGYRRVEELVWIKVNQMHRIIRTGRTGHWLNH
SKEHCLVGIKGNPSIIKALDCDVIVSEVRETSRKPDEIYGLIQRMCPNGKKIELFGRPHN
CRENWITLGNQLPGVFIKDETILSRFAEAYPQVDISENSMNTNQKQMSNQENLNVIYN
NHIGKKE

>PbAMT6

MVLEYRQNPKYQPQIAELAPEQYKYAIQQFSIVKQANPLEAQSPDIQALFINIKWNHE
QGATIQEFVKIHHHHQQQMVIPESMMSNGILFIWCDKDNIMEIIDHLDSQGFNYVENF
TMVILSREKILTMNDKTKKITDFFTKKEKKEQKLNDKQELQLSDLPDLDPSTVFWNQ
DQQYFRKAKRTLLMFRKISKQNLELRHQRTGDIFFDVVDDPQGISDIGLEYVYKMIET
LLPKAKYSNGEQLKMMELFADPKSTGRLGWVQVIKE

Table S9. The adjusted *P*-value of Tukey's multiple comparisons test of gene methylation level among the gene age groups, related to Figure 7.

	psl1	psl2	psl3	psl5	psl6	psl9	psl10
psl1	-						
psl2	****	-					
psl3	0.6163	0.9869	-				
psl5	0.0071**	0.9093	>0.9999	-			
psl6	****	****	0.9975	0.3099	-		
psl9	****	****	0.9556	0.0108*	0.6351	-	
psl10	****	****	0.0641	****	****	****	-

PSL: Phylostratigraphic level. **** $P < 0.0001$; *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$.

Table S10. The adjusted *P*-value of Tukey's multiple comparisons test of expression level among the gene age groups, related to Figure 7.

	psl1	psl2	psl3	psl5	psl6	psl9	psl10
psl1	-						
psl2	****	-					
psl3	0.8334	0.9991	-				
psl5	0.0263*	0.9676	>0.9999	-			
psl6	0.7657	0.4156	0.9613	0.3843	-		
psl9	****	0.0654	>0.9999	0.9994	0.0006***	-	
psl10	****	0.1784	>0.9999	0.9925	0.0042**	0.9991	-

PSL: Phylostratigraphic level. **** $P < 0.0001$; *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$.

Table S11. The comparison of gene methylation level and expression level in endosymbiosis-related genes and other genes, related to Figure 7.

	Gene	Gene methylation level	6mA amount	6mA ratio (%)	Copy number	6mA amount/copy number	6mA ratio/copy number (%)	RPKM	Phylostratum level
Cation channel family protein	g2947	0	0	0	3.05	0.00	0.00	8.09	1
Cation channel family protein	g5937	0	0	0	1.89	0.00	0.00	0.79	2
Cation channel family protein	g5938	0	0	0	1.88	0.00	0.00	1.28	1
Cation channel family protein	g5940	7.6	9	2.30	2.02	4.46	1.14	3.47	1
Cation channel family protein	g6962	5	6	1.19	3.31	1.81	0.36	13.22	9
Cation channel family protein	g6963	0	0	0	4.11	0.00	0.00	2.88	1
Cation channel family protein	g6964	0	0	0	5.94	0.00	0.00	0.99	1
globin	g1045	8.87	14	5.28	7.67	1.83	0.69	483.46	1
MATE family	g5951	0	0	0	4.87	0.00	0.00	203.88	1
MATE family	g17276	3.07	4	0.44	10.20	0.39	0.04	42.13	1
MATE family	g17549	13.93	18	3.73	5.20	3.46	0.72	22.98	1
mean of other genes	-	16.75	13.79	1.93	4.54	7.14	0.99	12.43	-

Table S12. Comparison of gene annotation in *Paramecium bursaria*, related to Figure 1.

	This study	He et al., 2019
Gene number	17,825	17,266
Overlapped gene	-	4,472
Number of genes longer than 1kb	8,220	7,654

Table S13. The NCBI GI number of MT-A70 family proteins used for phylogenetic tree construction, related to Figure 5.

		AMT1	AMT2/5	AMT3	AMT4	AMT6/7	DAMT1
Rhizaria	*Reticulomyxa filosa*	569355952				569429466	
Stramenopiles	*Blastocystis hominis*	300122443				300121574	
Apicomplexa	*Plasmodium falciparum*			906521926	910273602		
Ciliates	*Oxytricha trifallax*	403359546 403376498	403348598			403331225 403366155	
	Tetrahymena thermophila	586728217	586735011 586734668	89304141	586734236	586743342 586727895	
	Paramecium tetraurelia	145476411 145544559	74830621 145473723 74830631 74830636	145499669 145492063	74830646 145534770	145487402 145546436	
Algae	*Chlamydomonas reinhardtii*	159474530 1335048048		1335054311	159466562 1335048210	159480906 1335043186	
Plants	*Arabidopsis thaliana*			4539001	73620975		18394726
	Physcomitrella patens			1373907462	1333491572		1373923975
Ameoba	*Acanthamoeba castellanii*	440792815	440802961			440791284 440794060	
Basal fungi	*Anaeromyces robustus*	1183359562				1183358173	
	Absidia repens	1183482424				1183497549 1183498450	1183491450
True fungi	*Saccharomyces cerevisiae*			1023945599 (KAR4)	1322815 (IME4)		
Animals	*Drosophila melanogaster*			74869774	33301422		24647514
	Caenorhabditis elegans						74962031
	Danio rerio			82186021	45501238		189522093
	Homo sapiens			24308265	21361827		84040265

Table S14. Primers used for RT-PCR and qRT-PCR verification on the *PbAMT1* gene, related to Figure 5-6.

Primers	Sequences (5' to 3')
PbAMT1_f3669	CCGAATATTTCCAAATAATGCCC
PbAMT1_r3890	GGCGTCAAGGTATGATCAATAG
PbAMT1_r4079	CATATGGTAATATTATATTC
PbAMT1_f4112	GGGCATTTTACTTATCATATCG
PbAMT1_r4392	CAACGATATCTCTAGGAGTGC
PbAMT1_f42017	GCACTCCTAGAGATATCGTTG
PbAMT1_r41758	ATAGGAAGGATGAGGATTAG
PbAMT1_f42770	CAATTCATTTCCTATTGTAACCC
PbAMT1_f42231	GCAACACCTCTAGATGGTTATG
PbAMT1_r42363	GGTAACTTGCAAGCTAATGG
PbAMT1_5prime	ATGATCTCCAAACAAAAGGG
PbAMT1_3prime	TCACAATTCATTTCCTATTG
GAPDH_f181	CCATGGGGAGATTCAGGAG
GAPDH_r411	GCATGAAGCATTACTGATTATG
PbAMT1_f3797	GAAGATAACATCACTCGCTACCG
PbAMT1_r4009	GGAAAATTGGGGATATAAAATTG

Table S15. Primers used for the qPCR verification coupled with *Dpn*I/*Dpn*II digestion, related to Figure 3.

Primers	Sequences (5' to 3')	Site name	RNA polymerases
Contig6:85,913f	CATATTAATGCATTGAATCC	H1	Pol II
Contig6:86,112r	TAATGAGAAGTTTTGTGAGAAGG		
Contig8:43,720f	TGGCCGTCAGCACTCCAAGC	H2	Pol II
Contig8:43,919r	AGAAATAAACCAAGACCCATTG		
Contig11:149,430f	GCGGGAAGTGGCTAAGGC	H4	Pol II
Contig11:149,629r	CTCAACAATTTTATTCTTTC		
Contig2:42,783f	GATGGCTAATTATTAAG	I1	Pol II
Contig2:42,982r	AATAAATATGATACTTAAAGG		
Contig3:85,336f	AATACATTCTATCCTTCTGG	I2	Pol II
Contig3:85,535r	ATTTATTATATTAATGTTAT		
Contig10:2,631f	TTATCTAATATCTGTCAATTCG	I3	Pol II
Contig10:2,830r	TCATTTGTAACAAAAGAG		
Contig5:84,965f	ATCGAATAGACTAAAATCTGC	N1	Pol II
Contig5:85,164r	CATTAGCTATGAAAATTGTTG		
Contig6:46,670f	GAAAAATGAGTTAAAAC	N6	Pol II
Contig6:46,869r	TAGTATCGTACCATCGTTAC		
Contig11:85,251f	GGAACTCATTTTAATTCAGG	N7	Pol II
Contig11:85,450r	GAGCTGCTACACTATTATCTC		
Contig380:3,095f	GAAGTAGTCACATTTAATGTG	N3	Pol I
Contig380:3,294r	ATTAACATTTTGAGCTGTTGCC		
Contig72:156,482f	TCGGCCATACTAAACC	N4	Pol III

Table S16. Detailed information of alternative splicing analysis and 6mA-IP, related to Figure 6.

Sites	ONT Intron retention (WT/RNAi)	RNA-seq-1 Intron retention (WT/RNAi)	RNA-seq-2 Intron retention (WT/RNAi)	6mA sites' number
IR-1	Y/N (4/29 vs. 0/19)	Y/Y	Y/Y	8
IR-2	Y/Y (5/32 vs. 1/25)	Y/Y	Y/Y	2
IR-3	Y/Y (7/9 vs. 5/8)	Y/Y	Y/Y	3

(WT/RNAi)	WT*	WT-1	WT-2	RNAi-1	RNAi-2
Total events' number	825	887	736	631	550
AE	65	225	182	146	127
ES	8	14	14	13	11
IR	752 (ONT: WT 689 vs. RNAi 609)	648	540	472	412

WT*: RNA-seq data from He et al., 2019.