

**Supplementary Materials for**  
**An extremely streamlined macronuclear genome in the free-living**  
**protozoa *Fabrea salina***

**This PDF file includes:**

Detailed commands used for genome assembly

Figs. S1 to S8

Tables S1 to S4

## Detailed commands used for genome assembly

The Genome was assembled using the porechop--Flye--Racon--Medaka--Pilon pipeline. The command was as follows:

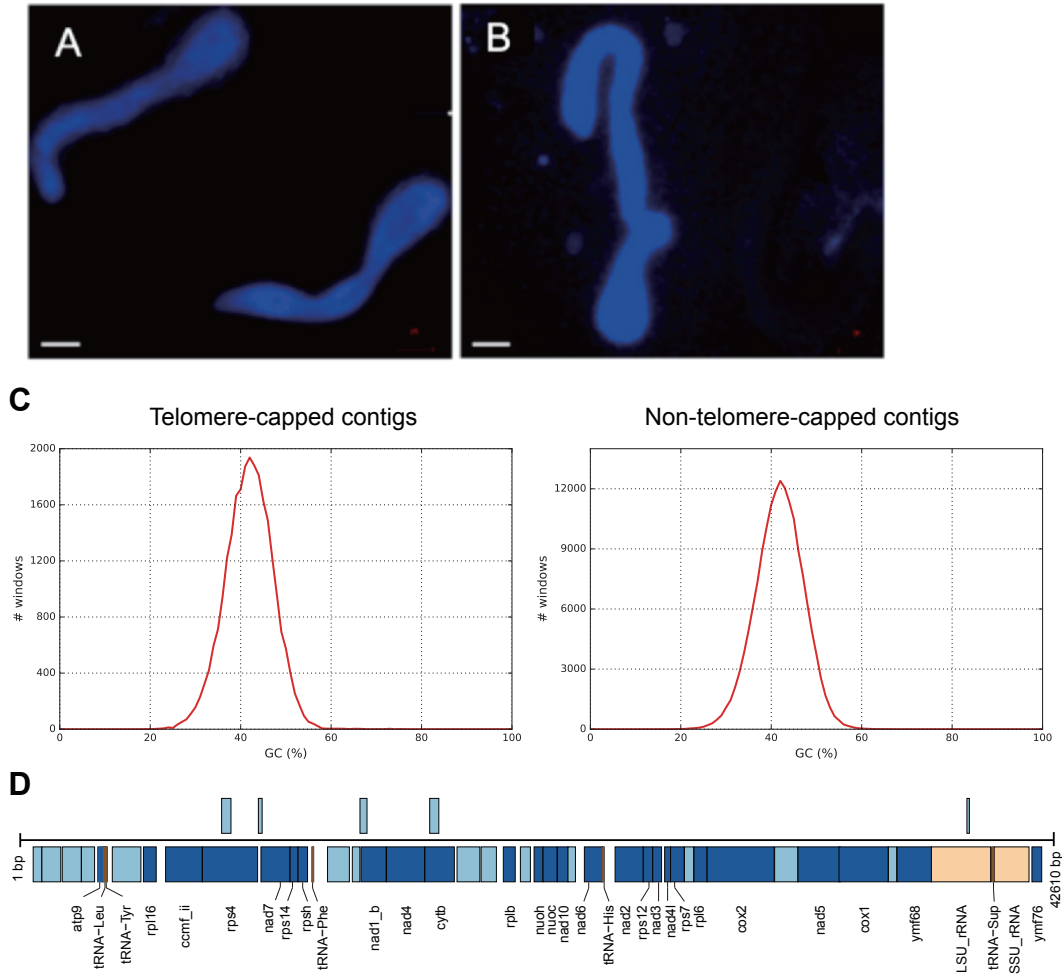
```
# Trimming using Porechop (v0.2.4)
porechop -i ont_reads_raw.fastq.gz -o ont_reads_trimmed.fastq.gz --threads 8
gunzip ont_reads_trimmed.fastq.gz

# Flye (v2.8.1) assembly
flye --nano-raw ont_reads_trimmed.fastq --out-dir ./flye_out --threads 8

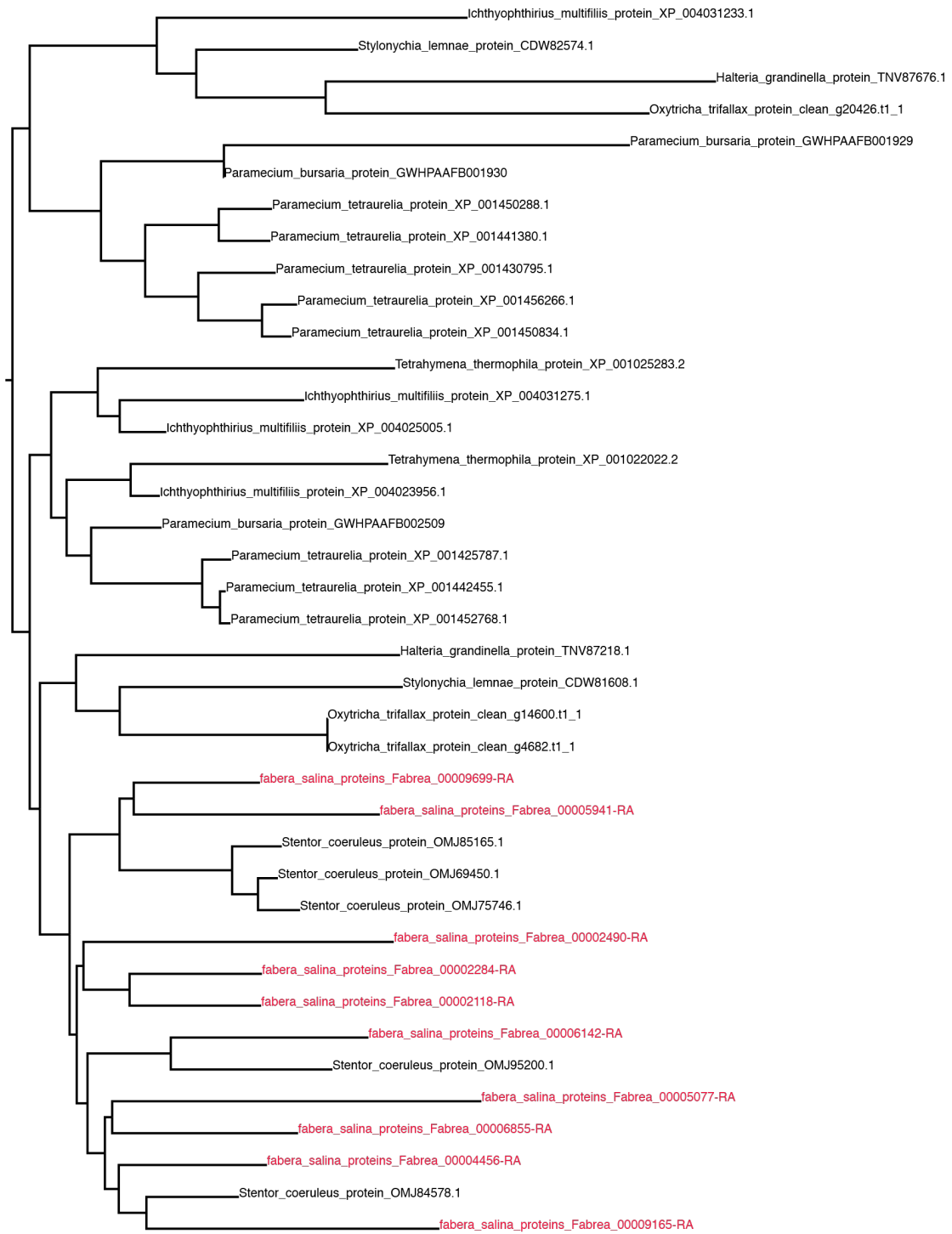
# Polishing using minimap2 (v2.17) and racon (v1.4.3)
mkdir -p racon_out && cd racon_out
minimap2 -t 8 ../flye_out/assembly.fasta ont_reads_trimmed.fastq > flye.paf
racon -t 8 -m 8 -x -6 -g -8 -w 500 -c 1 -b --cudaaligner-batches
1 ../ont_reads_trimmed.fastq flye.paf ../flye_out/assembly.fasta > racon.fasta

# Polishing using medaka (v1.3)
cd ..
medaka_consensus -i ont_reads_trimmed.fastq -d racon.fasta -o ./medaka_out -t 8
-m r941_min_high_g360 -b 50

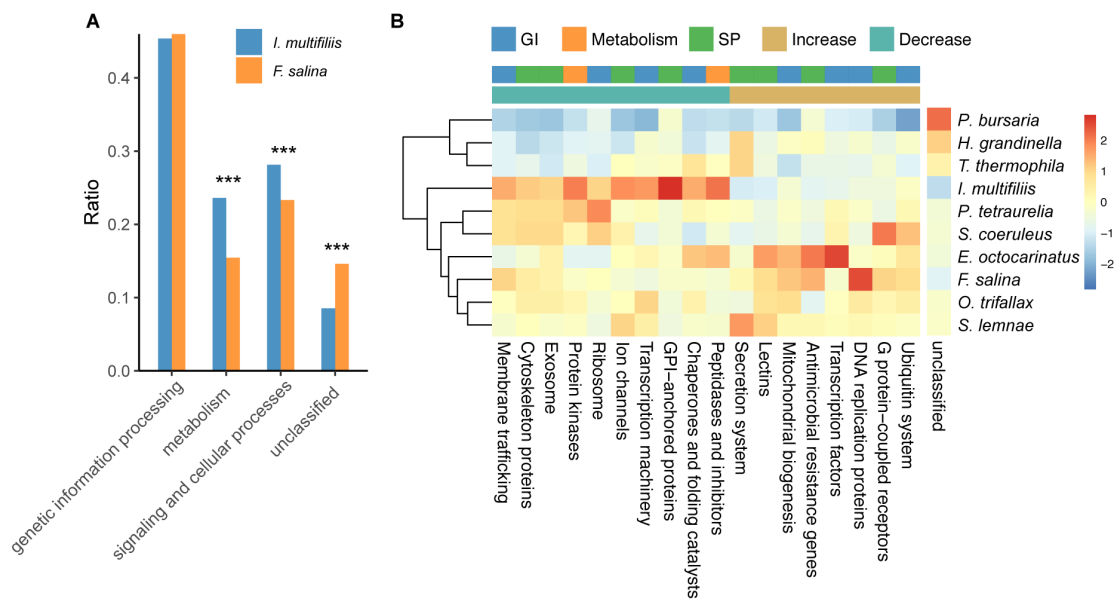
# Polishing using bwa (v0.7.17) and pilon (v1.23)
mkdir -p pilon_out && cd pilon_out
ln -s ../medaka_out/consensus.fasta ./draft.fa
mkdir index
bwa index -p index/draft draft.fa
bwa mem -t 8 index/draft ../short-read1.fq.gz short-read2.fq.gz | samtools sort -@
8 -O BAM -o aligned.sorted.bam
samtools index aligned.sorted.bam
java -Xmx128G -jar pilon-1.23.jar --genome draft.fa --frags aligned.sorted.bam --
threads 8 --fix snps,indels --output pilon_polished --vcf &> pilon.log
```



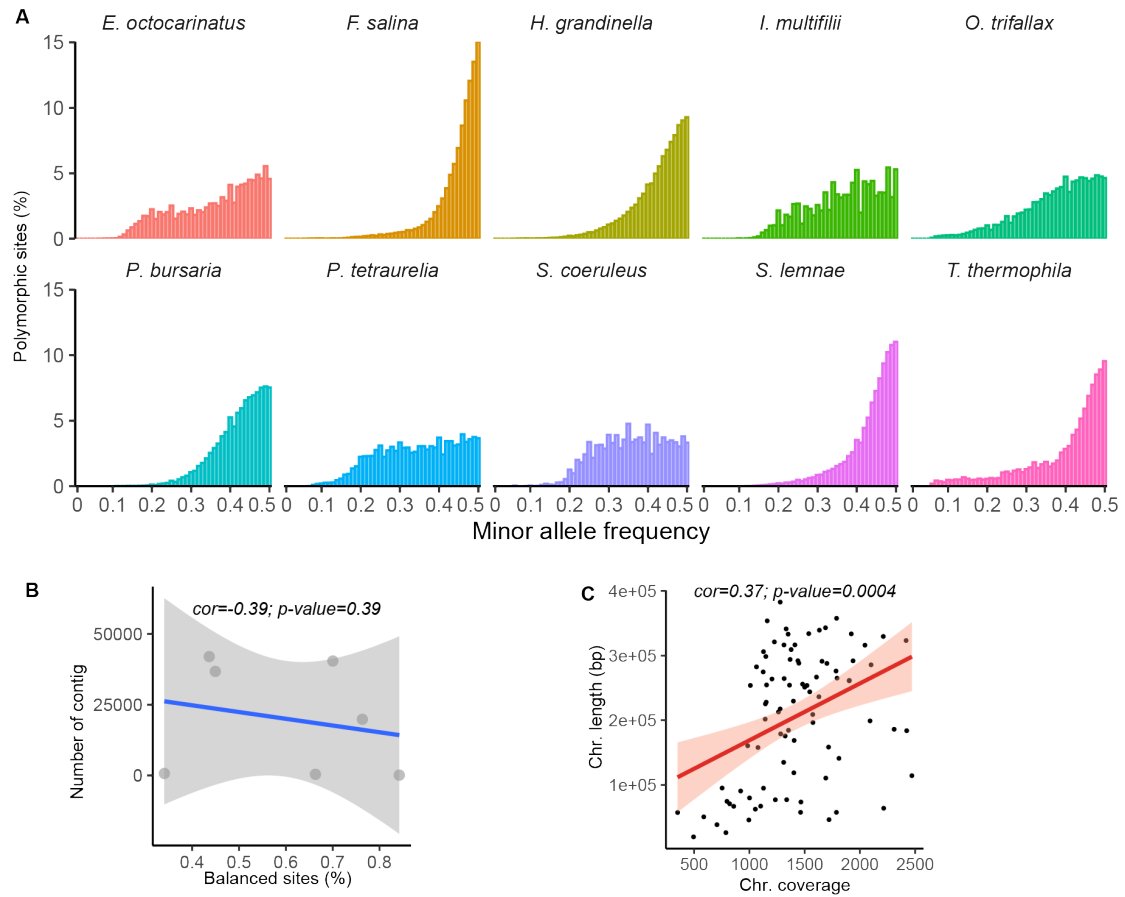
**Figure S1.** Macronuclei with Hoechst 33342 staining (A-B). The scale is 10  $\mu$ m. (C) The GC content distribution of all assembled contigs. (D) Graphical map of the mitochondrial genome. Light blue, dark blue, light brown, and dark brown represent unclassified ORFs, protein-coding genes, tRNA genes and rRNA genes, respectively.



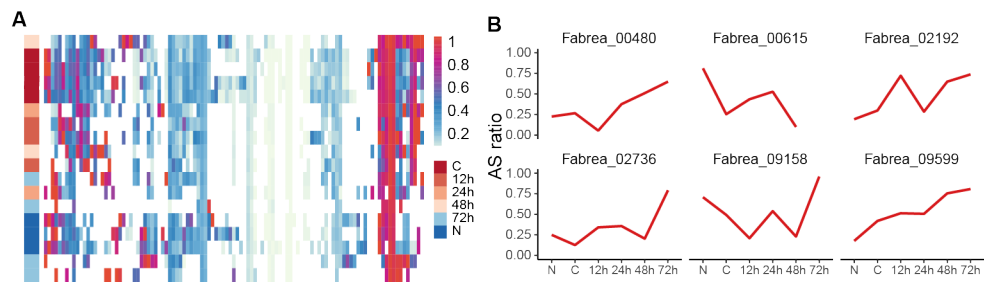
**Figure S2. Phylogenetic tree of the orthogroup (GO:0010766 and GO:1903288).**



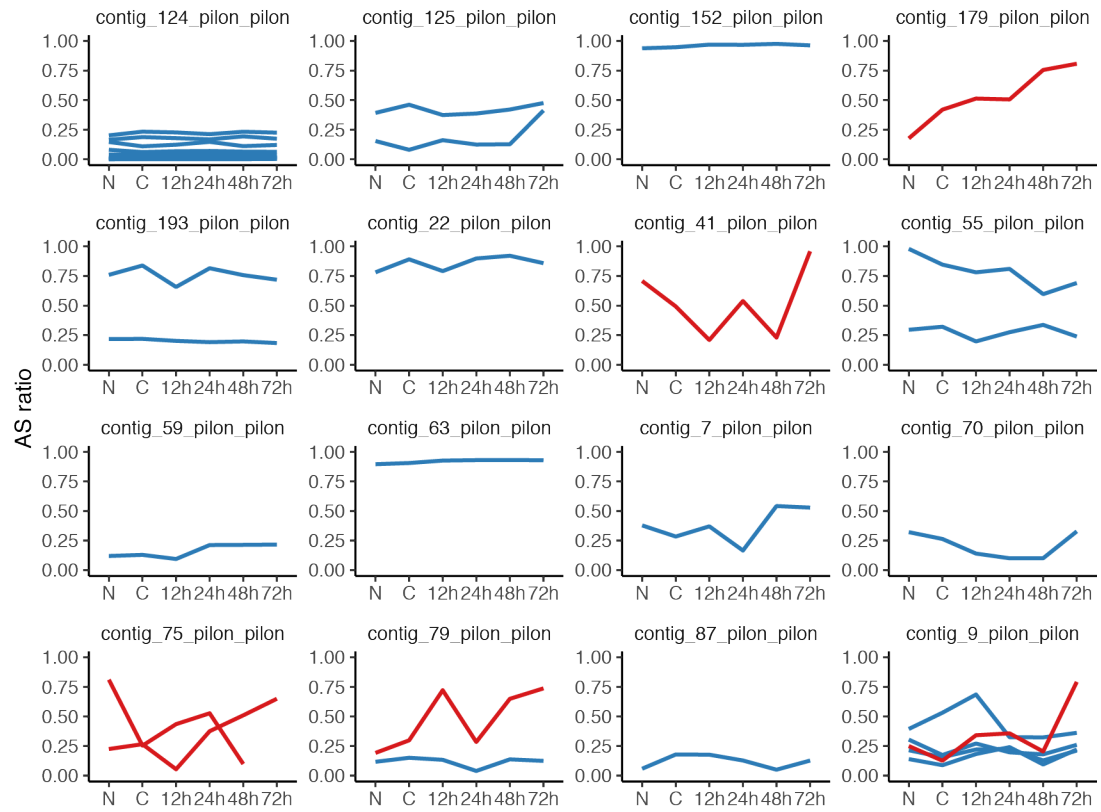
**Figure S3. Enhanced energy utilization and gene replication compared to a non-free-living ciliate.** (A) The proportion of gene annotation to genetic information processing (GI), signaling and cellular processes (SC) and metabolism for *I. multifiliis* and *F. salina*. The p value was calculated with Fisher exact test. (B) All KO terms with significant differences between *I. multifiliis* and *F. salina*.



**Figure S4. Characteristics of polymorphic sites among ciliates.** (A) Distribution of minor-allele frequency (MAF) of polymorphic sites of ten ciliates. The bin size is 0.01. (B) The spearman correlation between the balanced sites proportion and the contig number. Each dot denotes a species. (C) Spearman correlation between contig length and the coverage. Each dot denotes a contig in *F. salina*.

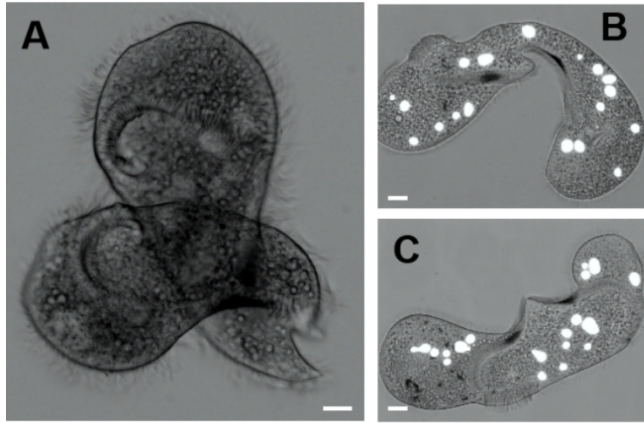


**Figure S5. Internally eliminated sequences (IESs) in ciliate genomes.** (A) Heatmap of the AS ratio for AEIs during conjugation. C represents the undivided conjugation pairs. N represents vegetative cells. Twelve hours, 24 hours, 48 hours, and 72 hours represent samples at 12, 24, 48, and 72 hours after conjugation separation, respectively. (B) The six highly variable AEIs ( $sd > 2$ ) show different directions of variation.

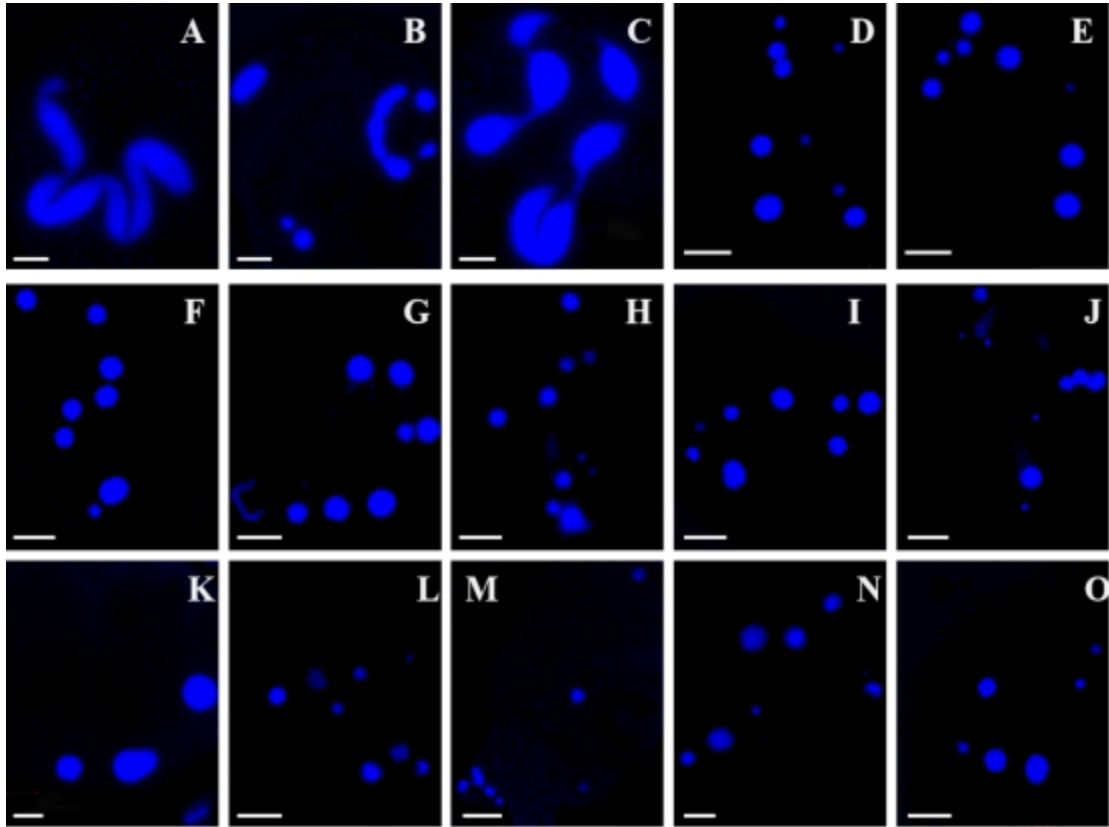


**Figure S6. The expression changes of 33 AS-IESs during conjugation in *F. salina*.**





**Figure S7. Conjugating pair of *F. salina* in different conjugation stages.**



**Figure S8. Nuclear phase during conjugation.** (A-E) Fluorescence photomicrograph of *F. salina* with Hoechst33342 staining before mate pair separation. (F-O) *Fabrea salina* at 12h (F-H), 24h (I-J), 48h (K-L) and 72h (M-O) after mate pair separation. Scale bar represents 10  $\mu\text{m}$ .

**Table S1. Genome completeness was assessed using BUSCO.**

Species	Completeness	Fraction of full-length single-copy genes	Fraction of full-length duplicate genes	Fraction of fragmented genes
<i>Fabera_salina</i>	0.836	0.772	0.064	0.023
<i>Euplotes_octocarinatus</i>	0.866	0.778	0.088	0.029
<i>Halteria_grandinella</i>	0.556	0.503	0.053	0.07
<i>Ichthyophthirius_multifiliis</i>	0.825	0.772	0.053	0.053
<i>Oxytricha_trifallax</i>	0.942	0.17	0.772	0.006
<i>Paramecium_bursaria</i>	0.216	0.175	0.041	0.281
<i>Paramecium_tetraurelia</i>	0.988	0.298	0.69	0.006
<i>Stentor_coeruleus</i>	0.924	0.462	0.462	0.023
<i>Stylonychia_lemnae</i>	0.971	0.936	0.035	0.012
<i>Tetrahymena_thermophila</i>	0.994	0.982	0.012	0

**Table S2. Codon usage in *F. salina* inferred by Codetta.**

<b>Codon</b>	<b>Inference</b>	<b>Codon</b>	<b>Inference</b>
TTT	F	ATT	I
TTC	F	ATC	I
TTA	L	ATA	I
TTG	L	ATG	M
TCT	S	ACT	T
TCC	S	ACC	T
TCA	S	ACA	T
TCG	S	ACG	T
TAT	Y	AAT	N
TAC	Y	AAC	N
TAA	Stop codon	AAA	K
TAG	Stop codon	AAG	K
TGT	C	AGT	S
TGC	C	AGC	S
TGA	Stop codon	AGA	R
TGG	W	AGG	R
CTT	L	GTT	V
CTC	L	GTC	V
CTA	L	GTA	V
CTG	L	GTG	V
CCT	P	GCT	A
CCC	P	GCC	A
CCA	P	GCA	A
CCG	P	GCG	A
CAT	H	GAT	D
CAC	H	GAC	D
CAA	Q	GAA	E
CAG	Q	GAG	E
CGT	R	GGT	G
CGC	R	GGC	G
CGA	R	GGA	G
CGG	R	GGG	G

**Table S3. Gene duplication events in *F. salina* and *S. coeruleus***

<b>Species</b>	<b>Singleton</b>	<b>Dispersed</b>	<b>Proximal</b>	<b>Tandem</b>	<b>WGD or segmental</b>
<i>F. salina</i>	2651	6663	151	226	227
<i>S. coeruleus</i>	14372	18659	258	1013	8917

**Table S4. The genomic sequencing data from nine published ciliates.**

<b>Species</b>	<b>Accession ID</b>	<b>Web site</b>
<i>Stentor coeruleus</i>	SRR5025913	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
<i>Euplotes octocarinatus</i>	SRR2467954;SRR2474294	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
<i>Ichthyophthirius multifiliis</i>	SRR088818;SRR088824;SRR088827;SRR089376;SRR089404;SRR1346113	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
<i>Oxytricha trifallax</i>	SRR578166	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
<i>Paramecium bursaria</i>	CRR078362	<a href="https://ngdc.cncb.ac.cn/">https://ngdc.cncb.ac.cn/</a>
<i>Paramecium tetraurelia</i>	SRR652988;SRR652989	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
<i>Stylonychia lemnae</i>	ERR469296;ERR469297;ERR469298;ERR469299	<a href="https://www.ebi.ac.uk/ena/browser/home">https://www.ebi.ac.uk/ena/browser/home</a>
<i>Tetrahymena thermophila</i>	SRR11906179	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>
<i>Halteria grandinella</i>	SRR8270648	<a href="https://www.ncbi.nlm.nih.gov">https://www.ncbi.nlm.nih.gov</a>