MOLECULAR ECOLOGY
RESOURCES

Supplemental Information for:

Genome analyses of the new model protist Euplotes vannus focusing on genome rearrangement and resistance to environmental stressors

Xiao Chen, Yaohan Jiang, Feng Gao, Weibo Zheng, Timothy J. Krock, Naomi A. Stover, Chao Lu, Laura A. Katz, Weibo Song

Table of Contents:

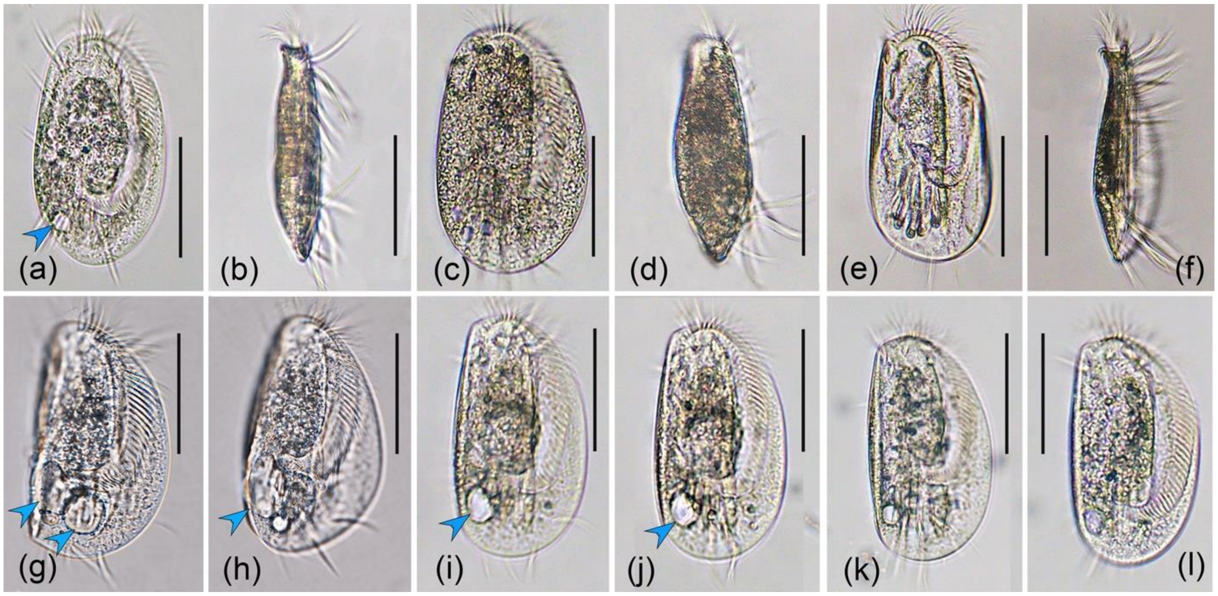

Figure S1. *Euplotes vannus* from life under different conditions. (a, b) Ventral and lateral views of the cell under normal conditions (20 ^{o}C and 30 psu). Arrowhead indicates the contractile vacuole. (c, d) Ventral and lateral views of the cell under low salinity (20 ^{o}C and 10 psu). (e, f) Ventral and lateral views of the cell under high salinity (20 ^{o}C and 60 psu). (g, h) Ventral views of the cell under low temperature (4 ^{o}C and 30 psu). Arrowheads indicate the bubbles. (i, j) Ventral views of the cell under high temperature (35 ^{o}C and 30 psu). Arrowhead indicates the usual contractile vacuole. (k, l) Ventral views of the cell in presence of free ammonia (100 mg/L NH_4Cl, pH 8.2, 20 ^{o}C and 30 psu).

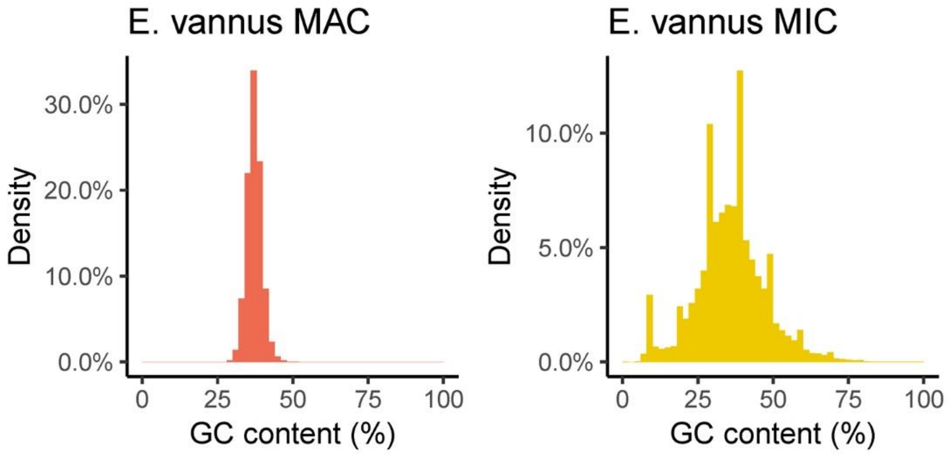

Figure S2. GC content distribution of MAC and MIC contigs of *Euplotes vannus.*

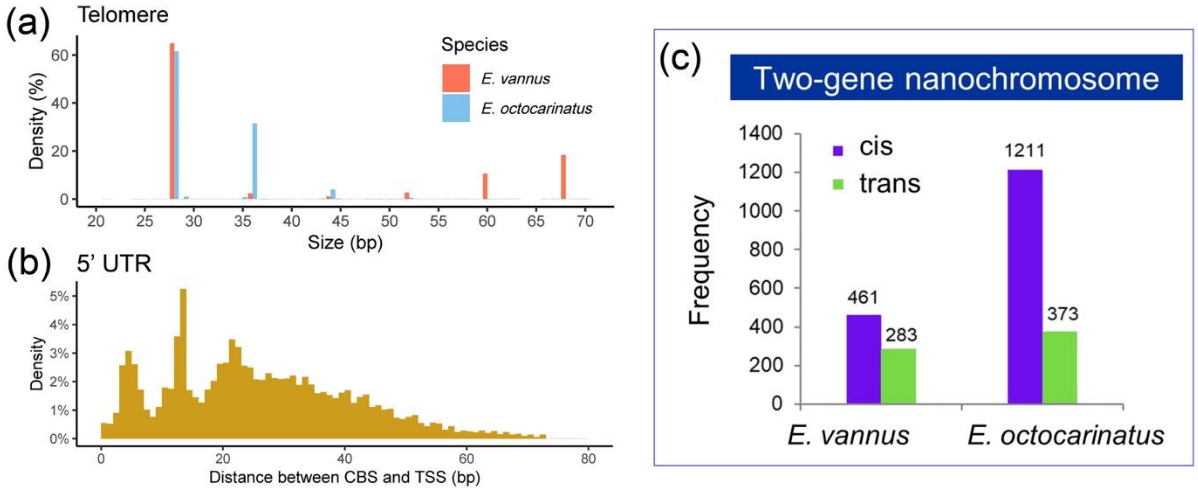

Figure S3. Size distribution of telomeres and 5' UTRs and the frequency of two-gene nanochromosomes in *Euplotes vannus* MAC. (a) Size distribution of telomeres in *Euplotes vannus* and *E. octocarinatus*. (b) Distribution of relative distance (5' UTR) between chromosome breakage sites (CBSs) and transcription start sites (TSSs) on 6588 *Euplotes vannus* nanochromosomes. (c) Two-gene nanochromosomes detected in MACs of *E. vannus* and *E. octocarinatus*. Violet bars denote the frequency of nanochromosomes containing genes transcribed in a same strand (*cis*) and green bars denote that of nanochromosomes containing genes transcribed in two different strands (*trans*).

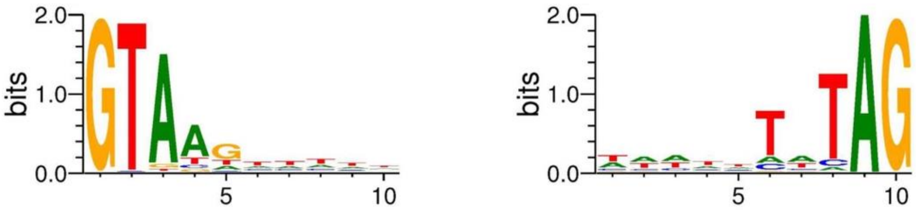

Figure S4. Schematic representation of the exon/intron boundaries with WebLogo in all 78661 introns in *Euplotes vannus* MAC genome. The GTR and YAG motifs are well conserved.

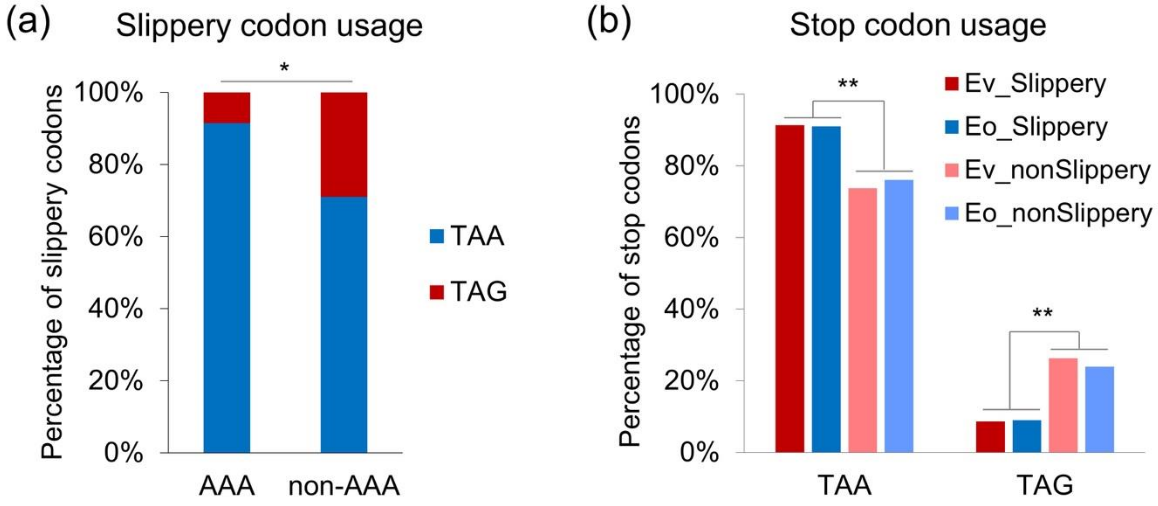

Figure S5. Slippery and stop codon usage in frameshifting transcripts. (a) Slippery codon usage in frameshifting sites with AAA and non-AAA upstream codons of *Euplotes vannus*. The asterisk denotes the statistically significant difference in Fisher's exact test (p = 0.0382). (b) Slippery and non-slippery stop codon usage in transcripts of *Euplotes vannus* (Ev) and *E. octocarinatus* (Eo). Asterisks denote the statistically significant difference in analysis of variance (ANOVA) (p = 0.005024).

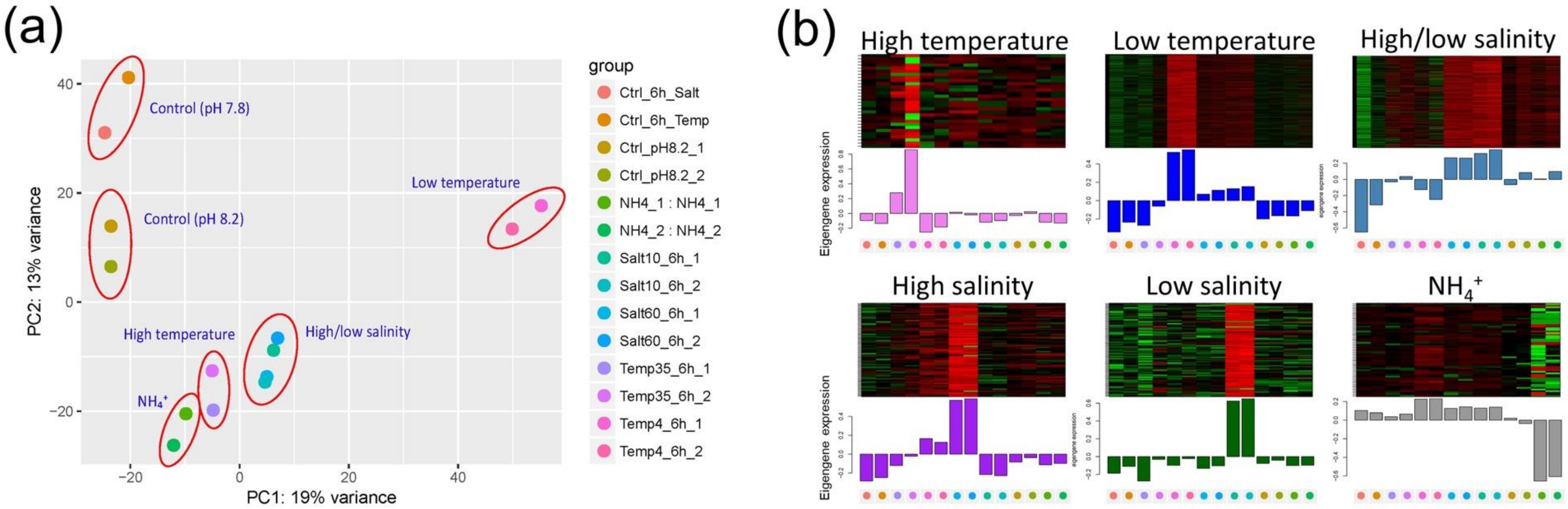

Figure S6. Differential gene expression and gene co-expression analyses under temperature, salinity and free ammonia stresses in *Euplotes vannus* (relative to Figure 5). (a) PCA analysis on gene expression under different stresses. (b) Different environmental stresses activated or deactivated different gene modules. The samples under different environmental stresses are denoted by dots on x-axis in the same color annotation as Figure S5a. The co-expressed genes in each module see Table S7.

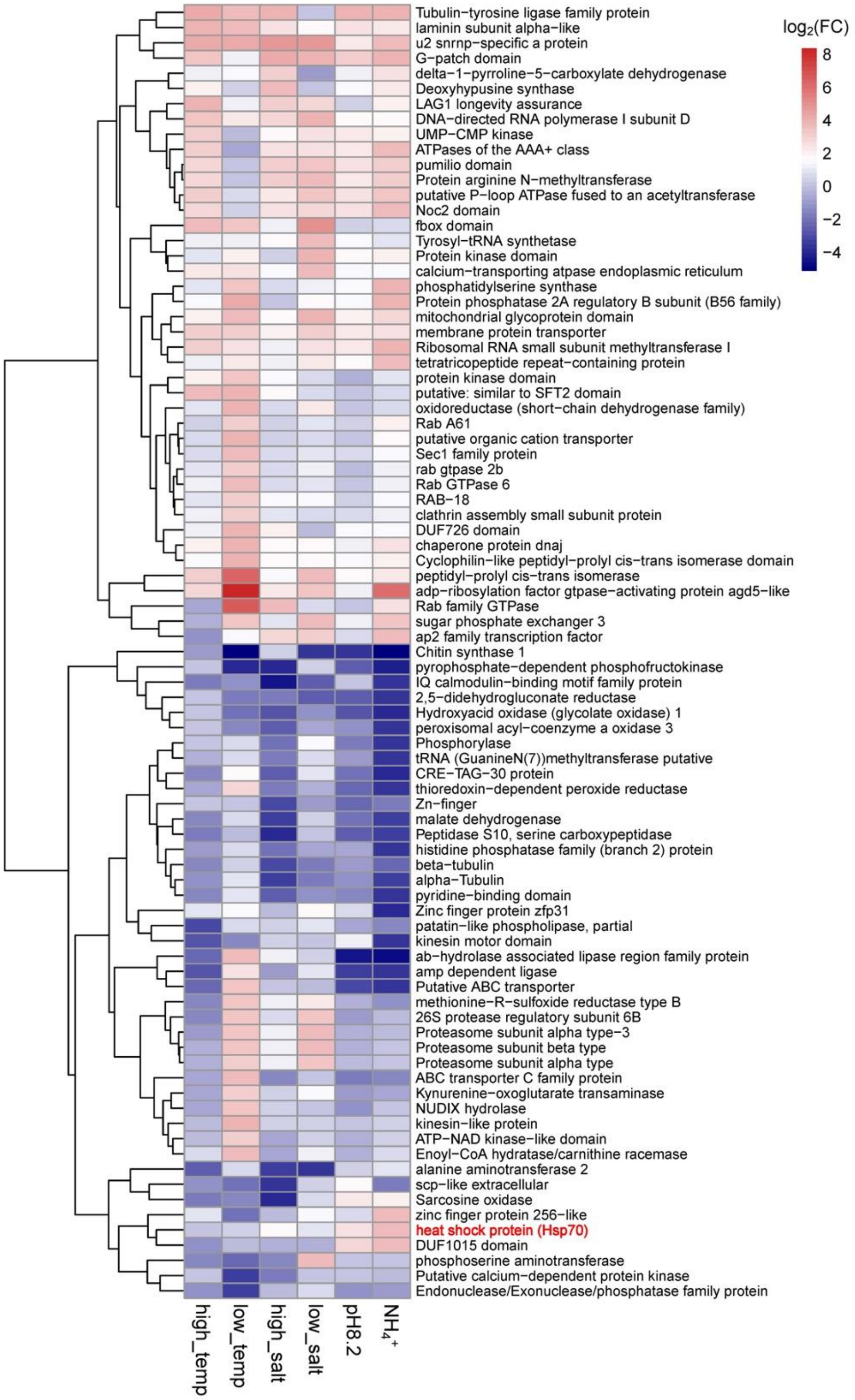

Figure S7. Top 100 differentially expressed genes under temperature, salinity and free ammonia stresses in *Euplotes vannus* (relative to Figure 5 and Figure 6). The numbers in the scale bar denote the log2 values of gene expression fold change between cells under stresses and cells as control. Analysis based on two biological replicates.

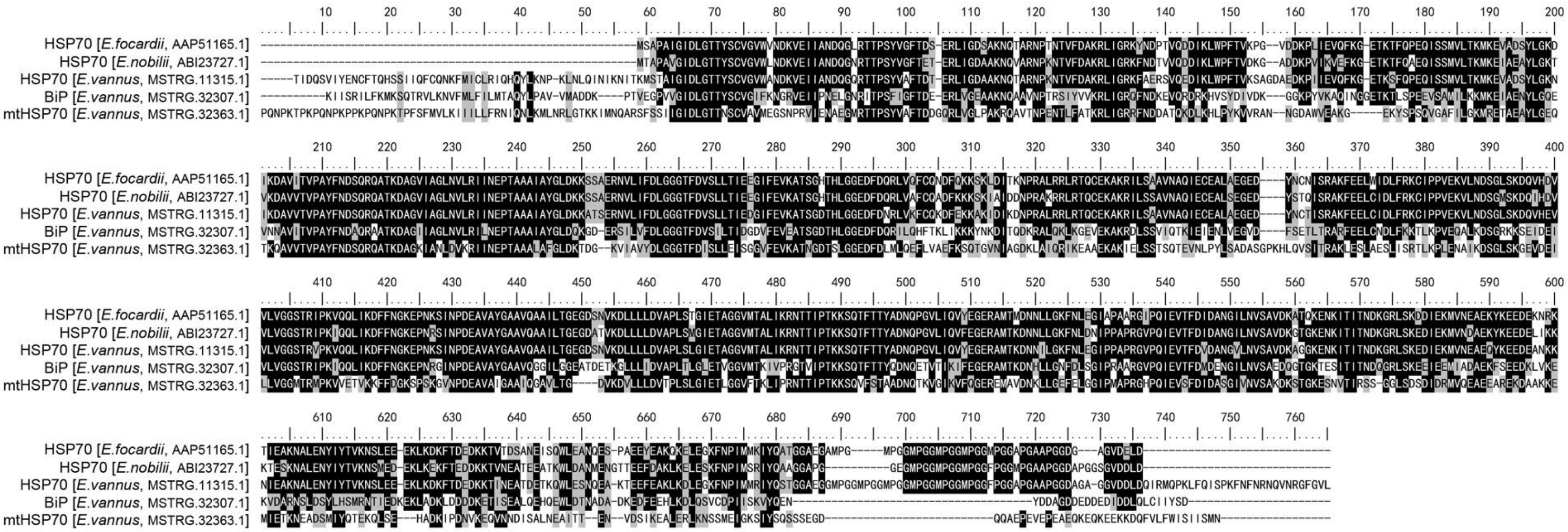

Figure S8. Sequence alignment of Hsp70 homologs in *Euplotes vannus*, *E. focardii* and *E. nobilii*, as well as other two relatively distant homologs of Hsp70, BiP (Binding immunoglobulin protein) and mtHsp70 (mitochondrial Hsp70).

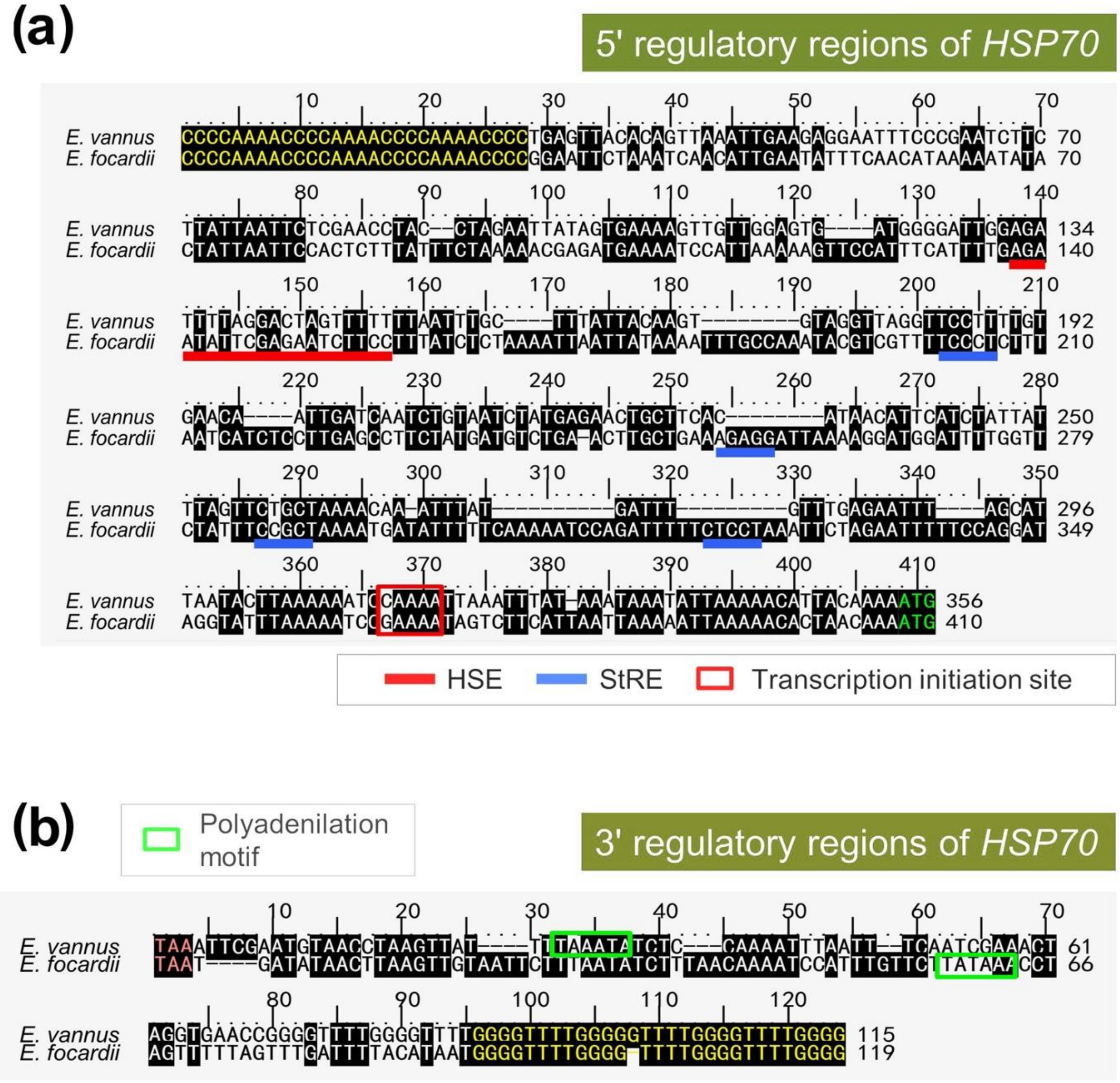

Figure S9. Sequence alignment of 5' and 3' regulatory regions of Hsp70 genes in *Euplotes vannus* and *E. focardii* (relative to Figure 6c). (a) Nucleotide sequence alignment of the 5’ regulatory regions of HSP70 genes between *E. vannus* and *E. focardii*. The identities are shaded; the telomeric C_4A_4 repeats and translation initiation codons (ATG) are in yellow and green, respectively; putative sites for the transcription initiation are boxed; sequence motifs bearing agreement with HSE and StRE elements are indicated by red and blue bars, respectively. (b) Nucleotide sequence alignment of the 3’ regulatory regions of the *E. vannus* and *E. focardii* HSP70 genes; neither of them carries ARE elements in the 3’ regulatory region. The identities are shaded; the telomeric G_4T_4 repeats and translation stop codons (TAA) are in red; putative polyadenylation motifs are boxed.

Table S1. Information of accession of genome/transcriptome assemblies of 32 ciliates.

No.	Class	Subclass	Order	Species	Data source	Accession Number	Reference
1	Oligohymenophorea	Hymenostomatia	Tetrahymenida	*Tetrahymena thermophila*	NCBI	GCA_000189635.1	Eisen, et al (2006)
2	Oligohymenophorea	Hymenostomatia	Ophryoglenida	*Ichthyophthirius multifiliis*	NCBI	GCA_000220395.1	Coyne, et al. (2011)
3	Oligohymenophorea	Scuticociliatia	Philasterida	*Anophryoides haemophila*	MMETSP	MMETSP1018	Gentekaki, et al. (2017)
4	Oligohymenophorea	Scuticociliatia	Philasterida	*Uronema* sp.	MMETSP	MMETSP0018	Gentekaki, et al. (2017)
5	Oligohymenophorea	Peniculia	Peniculida	*Paramecium tetraurelia*	NCBI	GCA_000165425.1	Aury, et al. (2006)
6	Colpodea	-	Cyrtolophosidida	*Platyophrya macrostoma*	MMETSP	MMETSP0127	Gentekaki, et al. (2017)
7	Colpodea	-	Cyrtolophosidida	*Aristerostoma* sp.	MMETSP	MMETSP0125	Gentekaki, et al. (2017)
8	Spirotrichea	Choreotrichia	Tintinnida	*Favella taraikaensis*	MMETSP	MMETSP0434	-
9	Spirotrichea	Choreotrichia	Tintinnida	*Schmidingerella arcuata* [*Favella ehrenbergii*]	MMETSP	MMETSP0123	Gentekaki, et al. (2017)
10	Spirotrichea	Choreotrichia	Choreotrichida	*Strombidinopsis acuminatum*	MMETSP	MMETSP0126	Gentekaki, et al. (2017)
11	Spirotrichea	Choreotrichia	Choreotrichida	*Strombidium sulcatum*	NCBI	PRJNA335783	Chen, et al. (2015)
12	Spirotrichea	Choreotrichia	Choreotrichida	*Strombidium inclinatum*	MMETSP	MMETSP0208	Gentekaki, et al. (2017)
13	Spirotrichea	Choreotrichia	Choreotrichida	*Strombidium rassoulzadegani*	MMETSP	MMETSP0449	-
14	Spirotrichea	Stichotrichia	Sporadotrichida	*Oxytricha trifallax*	NCBI	GCA_000295675.1	Swart, et al. (2013)
15	Spirotrichea	Stichotrichia	Sporadotrichida	*Stylonychia lemnae*	NCBI	GCA_000751175.1	Aeschlimann et al. (2014)
16	Spirotrichea	Stichotrichia	Urostylida	*Pseudokeronopsis* sp. 1	MMETSP	MMETSP1396	-
17	Spirotrichea	Stichotrichia	Urostylida	*Pseudokeronopsis* sp. 2	MMETSP	MMETSP0211	-
18	Spirotrichea	Hypotrichia	Euplotida	*Euplotes focardii*	MMETSP	MMETSP0205	Gentekaki, et al. (2017)
19	Spirotrichea	Hypotrichia	Euplotida	*Euplotes harpa*	MMETSP	MMETSP0213	Gentekaki, et al. (2017)
20	Spirotrichea	Hypotrichia	Euplotida	*Euplotes octocarinatus*	-	-	Wang, et al. (2018)
21	Spirotrichea	Hypotrichia	Euplotida	*Euplotes crassus*	MMETSP	MMETSP1380	-
22	**Spirotrichea**	**Hypotrichia**	**Euplotida**	***Euplotes vannus***	**-**	**-**	**The current work**
23	Litostomatea	Haptoria	Pleurostomatida	*Litonotus* sp.	MMETSP	MMETSP0209	Gentekaki, et al. (2017)
24	Protocruziea	Protocruziidia	Protocruziida	*Protocruzia adherens*	MMETSP	MMETSP0216	Gentekaki, et al. (2017)
25	Mesodiniea	-	-	*Myrionecta rubra*	MMETSP	MMETSP0798	-

26	Mesodiniea	-	-	*Mesodinium pulex*	MMETSP	MMETSP0467	-
27	Prostomatea	-	Prorodontida	*Tiarina fusus*	MMETSP	MMETSP0472	-
28	Heterotrichea	-	Heterotrichida	*Climacostomum virens*	MMETSP	MMETSP1397	-
29	Heterotrichea	-	Heterotrichida	*Blepharisma japonicum*	MMETSP	MMETSP1395	-
30	Heterotrichea	-	Heterotrichida	*Fabrea salina*	MMETSP	MMETSP1345	-
31	Heterotrichea	-	Heterotrichida	*Stentor coeruleus*	NCBI	GCA_001970955.1	Slabodnick, et al. (2017)
32	Heterotrichea	-	Heterotrichida	*Condylostoma magnum*	MMETSP	MMETSP0210	Gentekaki, et al. (2017)

Table S2. Genomic and transcriptomic reads mapping information.

Sample	**Source**	**Input reads**	**Mapped reads**	**Note**
Genome	DNA	559039802	392763937 (70.3%)	Coverage = 640X
Starvation_1	RNA	20175266	15411886 (76.4%)	Starved cells, replicate 1
Starvation_2	RNA	25939381	23501079 (90.6%)	Starved cells, replicate 2
Vegetative_1	RNA	22130139	15634943 (70.7%)	Vegetative cells, replicate 1
Vegetative_2	RNA	22656934	18753144 (82.8%)	Vegetative cells, replicate 2
Ctrl_1	RNA	27410007	27410007 (82.8%)	Control, pH 7.8, 20 ^oC and 35 psu, replicate 1
Ctrl_2	RNA	28426577	28426577 (83.9%)	Control, pH 7.8, 20 ^oC and 35 psu, replicate 2
Temp4_1	RNA	24378854	24378854 (74.3%)	Cells under low temperature 4^oC, replicate 1
Temp4_2	RNA	24728188	24728188 (72.4%)	Cells under low temperature 4^oC, replicate 2
Temp35_1	RNA	27392914	27392914 (80.1%)	Cells under high temperature 35^oC, replicate 1
Temp35_2	RNA	27802672	27802672 (83.1%)	Cells under high temperature 35^oC, replicate 2
Salt10_1	RNA	24108428	24108428 (84.0%)	Cells under low salinity 10 psu, replicate 1
Salt10_2	RNA	27871505	27871505 (80.1%)	Cells under low salinity 10 psu, replicate 2
Salt60_1	RNA	25259917	25259917 (82.8%)	Cells under high salinity 60 psu, replicate 1
Salt60_2	RNA	27558276	27558276 (72.1%)	Cells under high salinity 60 psu, replicate 2
pH8.2_1	RNA	26705063	26705063 (84.7%)	Cells under pH 8.2, 20 ^oC and 35 psu, replicate 1
pH8.2_2	RNA	24499855	24499855 (79.6%)	Cells under pH 8.2, 20 ^oC and 35 psu, replicate 2
NH4_1	RNA	28061657	28061657 (84.2%)	Cells under presence of free ammonia, replicate 1
NH4_2	RNA	24413430	24413430 (81.1%)	Cells under presence of free ammonia, replicate 2

Table S3. List of identified ncRNAs in *Euplotes vannus* MAC genome.

Rfam acc.	Rfam id	*E. vannus* MAC
RF00001	5S_rRNA	4
RF00002	5_8S_rRNA	5
RF00003	U1	2
RF00004	U2	6
RF00005	tRNA	109
RF00009	RNaseP_nuc	1
RF00010	RNaseP_bact_a	1
RF00012	U3	1
RF00015	U4	2
RF00017	SRP_euk_arch	2
RF00020	U5	2
RF00023	tmRNA	1
RF00025	Telomerase-cil	2
RF00026	U6	
RF00029	Intron_gpII	3
RF00059	TPP	1
RF00162	SAM	
RF00177	SSU_rRNA_5	4
RF00513	Trp_leader	
RF00876	mir-684	1
RF01059	mir-598	1
RF01118	PK-G12rRNA	1
RF01726	SAM-II_long_loops	
RF01848	ACEA_U3	
RF01852	tRNA-Sec	
RF01960	SSU_rRNA_eukarya	

Table S4. Annotation information of repeats in the MAC genome assembly of *Euplotes vannus* (separated supplementary data file).

Table S5. Expression and annotation information of *Euplotes vannus* genes (separated supplementary data file).

Table S6. Gene rearrangement events in *Euplotes vannus* (separated supplementary data file).

Table S7. Co-expressed genes in *Euplotes vannus* under dramatic environmental changes, including low/high temperature, low/high salinity and the presence of free ammonia (separated supplementary data file).