

RESEARCH

Open Access



# The genome and comparative transcriptome of the euryhaline model ciliate *Paramecium duboscqui* reveal adaptations to environmental salinity

Yu Fu<sup>1</sup>, Ping Ni<sup>1</sup>, Ying Zhang<sup>1</sup>, Fasheng Liang<sup>1</sup>, Naomi A. Stover<sup>2</sup> and Lifang Li<sup>1\*</sup>

## Abstract

**Background** As a potential model organism for studies of environmental and cell biology, *Paramecium duboscqui* is a special euryhaline species of *Paramecium* that can be found in fresh, brackish, or marine water in natural salinity ranges between 0‰ and 33‰. However, the genome information as well as molecular mechanisms that account for its remarkable halotolerant traits remain extremely unknown. To characterize its genome feature, we combined PacBio and Illumina sequencing to assemble the first high-quality and near-complete macronuclear genome of *P. duboscqui*. Meanwhile, comparative transcriptomic profiles under different salinities gave underlying insight into the molecular mechanism of its adaptations to environmental salinity.

**Results** The results showed that the MAC genome of *P. duboscqui* comprises 160 contigs, with 113 of them possessing telomere (~28.82 Mb haploid genome size). Through comparative genomic analyses with the other ciliate, we found that gene families encoding transmembrane transporter proteins have been expanded in *P. duboscqui*, showing enormous potential in salinity adaptation. Like other *Paramecium*, *P. duboscqui* utilizes TGA as its only termination codon and has reassigned TAA and TAG to encode glutamine. *P. duboscqui* showed different growth rates under different salinities, with an optimum growth rate in 5‰ salinity. A comparison of the transcriptomic profiles among *P. duboscqui* grown under different concentrations showed that genes involved in protein folding, oxygen respiration, and glutathione-dependent detoxification were upregulated in the high-salt group, whereas genes encoding DNA-binding proteins and transcription factors were upregulated in the low-salt group, suggesting distinct mechanisms for responding to low and high salinity. Weighted gene coexpression network analysis (WGCNA) linked the hub genes expressed at 30‰ salinity to cysteine-type peptidase activity, lipid transfer, sodium hydrogen exchange, and cell division, with the hub genes expressed at 0‰ salinity involved in transmembrane transport and protein localization.

**Conclusions** This study characterizes a new euryhaline model *Paramecium*, provides novel insights into *Paramecium* evolution, and describes the molecular mechanisms that have allowed *P. duboscqui* to adapt to different osmotic environments.

**Keywords** *Paramecium*, Macronuclear genome, Codon usage bias, Salinity adaptation

\*Correspondence:

Lifang Li

lifangli@sdu.edu.cn; qd\_liliy@sina.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Ciliated protozoa (ciliates) are one of the most diverse groups of single-celled microorganisms and have been reported as the dominant eukaryotic microbiota in aquatic ecosystems [1–3]. Among this group, the genus *Paramecium* (Ciliophora, Oligohymenophorea, Peniculida) is one of the best studied and most widespread [4]. As with all other ciliates, *Paramecium* harbors two types of nucleus, a diploid germline micronucleus (MIC) and a highly polyploid somatic macronucleus (MAC) [5]. Whereas the exact number of micronuclear chromosomes in *Paramecium* species has been difficult to determine, the macronuclear genome is known to comprise complete chromosomes that are amplified to about 800 copies [6]. During MAC development, imprecise removal of transposable elements and other repeated sequences leads to chromosome fragmentation, de novo telomere addition, and variable internal deletions, which along with its high polyploidy make it difficult to assemble complete somatic MAC genomes for *Paramecium* species [7]. To date, more than 40 *Paramecium* morphospecies have been found, but only 17 have been sequenced at the genomic level [8–10]. The rapid development of third generation sequencing technology has facilitated sequencing of complete MAC genomes from additional new species of *Paramecium*.

*Paramecium duboscqui* was established by Chatton and Brachon, who cataloged its truncated “*bursaria*” type body shape and overall length of 80–150  $\mu\text{m}$  [11]. It was initially considered a rare or even nonvalid species, and its distinction as a separate species was uncertain [10, 12]. Since 1997, this species has been revisited, and its morphology, mating types, and adaptation to temperature have been further described [13–15]. Unlike most *Paramecium* species that live exclusively in freshwater, *P. duboscqui* can be found in fresh, brackish, and marine water (around 0‰–33‰ salinity) and is recognized as a euryhaline and halotolerant species [15].

Adaptations to high salinity environments have been widely studied in prokaryotic unicellular organisms like Bacteria and Archaea, but microbial eukaryotes have received much less attention [16–18]. Early studies reported the existence of diverse populations of unicellular eukaryotes in hypersaline waters [19–22], with more recent studies showing that eukaryotic species may surpass prokaryotes in ~8–18% salt environments [23–28]. Halotolerant protozoa have received more attention during the past several years, with key findings made on community dynamics and the effects of environmental quality on protozoa [29–31]. However, the majority of these studies focused on ecology and biodiversity rather than the sequencing and gene expression information

needed to elucidate the molecular mechanisms of adaptation of halotolerant protozoa.

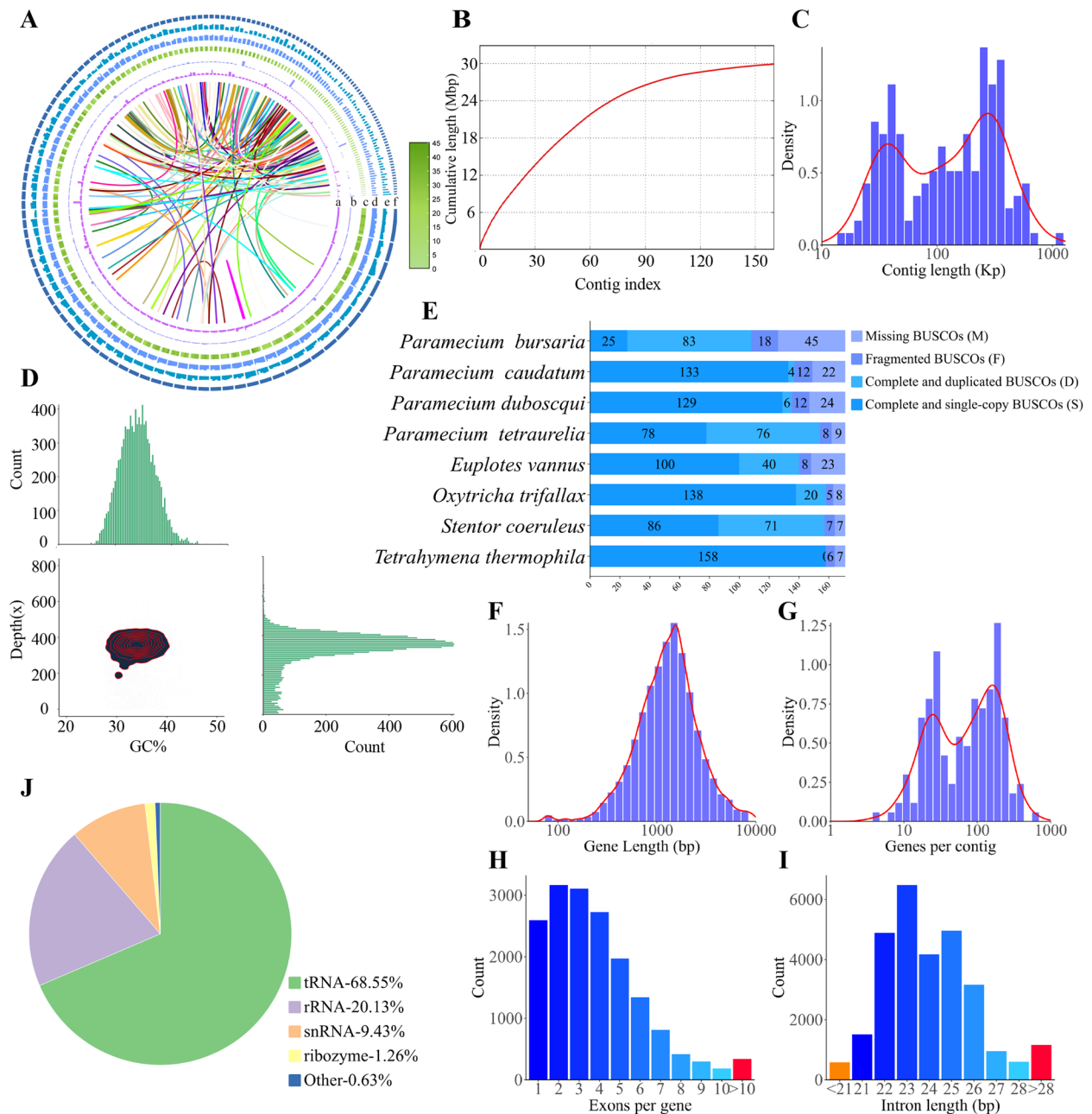
Unlike many other groups of eukaryotic microbes, ciliates like *P. duboscqui* are not enclosed by a cell wall and manage osmotic stress using contractile vacuoles [32]. The additional molecular mechanisms by which this particular species of *Paramecium* has adapted to high salinity are unknown. Furthermore, as one of the oldest branches of *Paramecium* species, *P. duboscqui* also serves as an important point of comparison with the other *Paramecium* species for which genome data are available and can help answer questions regarding *Paramecium* genome evolution, gene family contraction and expansion events, and the pathways available to respond to environmental stress. *P. duboscqui* is therefore a valuable model for understanding the evolution of halotolerance in *Paramecium*.

In this study, we employed both Illumina and PacBio sequencing technologies to assemble the somatic genome of *Paramecium duboscqui* and described the characteristics of its genome for the first time. In addition, through comparative genomic and transcriptomic analyses, we further systematically explored its genetic basis for adaptation to a range of salinity environments. These results provide unique insights into *Paramecium* genome evolution and the ability of *P. duboscqui* specifically to adapt to different salinity stress.

## Results

### Sequencing and assembly of the macronuclear genome of *P. duboscqui*

A total of 10.58 Gb of MAC Illumina clean reads were used to estimate the primary characteristics of the genome (Additional file 1: Table S1). The estimated genome size and heterozygosity were ~25.83 Mb and ~0.466%, respectively (Additional file 1: Fig. S3, Additional file 2: Table S2). To solve the assembly challenges arising from the highly polyploid and fragmented somatic MAC genomes of ciliates, we employed PacBio sequencing to generate high-fidelity (HiFi) reads for *P. duboscqui*. We then assembled a 28.82 Mb draft genome with 160 contigs and a contig N50 of 308 kb (Fig. 1A–C, Table 1, Additional file 2: Table S3). The GC content distribution of assembled contigs did not show any bacterial contaminants with discordant GC content (Fig. 1D, Additional file 1: Fig. S4). We found that 99.64% of the PacBio DNA-seq reads and 94.92% of Illumina RNA-seq reads could be properly aligned to the assembled genome (Table 1). We further evaluated the completeness of the assembled *P. duboscqui* genome by BUSCO [33] (Fig. 1E) and determined that it was comparable to that of other well-studied ciliate species (Additional file 2: Table S4).



**Fig. 1** Sequencing and assembly of the macronuclear genome of *Paramecium duboscqui*. **A** Characteristics of the 160 assembled contigs of *P. duboscqui*. Tracks from the inside out represent gene collinearity, the distribution of genome coverage of HiFi reads, genome coverage of Illumina RNA-seq reads, gene density, GC skew, GC density, and contig length, respectively, with values calculated in 3-kb sliding windows. **B** The cumulative distribution of contig length. **C** The distribution of contig length. **D** The GC content and depth of the genome. **E** Result of BUSCO for the *P. duboscqui* genome and other closely related species. **F–I** The distribution of gene length, gene number in each contig, exon number in each gene, and intron length. **J** Percentage of non-coding RNA genes found in the genome

The Eugene pipeline [34] was used to annotate the MAC genome, producing a final annotation that contained 16,774 protein-coding genes. Among the predicted protein-coding genes, almost all of them (93.2%) were supported by RNA-seq data, and 81.3% had

significant hits in the InterPro database. The gene number in each contig had a statistically significant correlation with contig length ( $r=0.984$ ,  $p$ -value<0.01). The median gene length was 1713 bp, and each contig had an average of ~212 genes (Fig. 1F, G). On average, each gene

**Table 1** Statistics of the *P. duboscqui* genome assembly

Genomic characteristics	
Genome size (Mb)	28.82
N50 (kb)	308
N75 (bp)	202,900
Longest contigs (bp)	1,164,585
Contigs number	160
Contigs number with telomere	113
GC content (%)	34
Mismatches	0
2 telomeres	20
1 telomere	93
Gene number	16,774
Gene length (bp)	1713.63
Mapping ratio of Illumina RNA seq	94.92%
Mapping ratio of PacBio seq	99.64%

contained ~ 3.92 exons and ~ 2.92 introns (Fig. 1H). Most introns are between 22 and 26 bp, with an average length of ~ 25 bp (Fig. 1I), which is similar to other *Paramecium* species. We also found that the *P. duboscqui* genome contained a typical intron motif with a highly conserved branch point adenosine (Additional file 1: Fig. S5). The genome of *P. duboscqui* is compact and streamlined, with an average intergenic region length of only ~ 67 bp. Notably, 113 of 160 assembled contigs (24.34 Mb) were capped with telomere, which is far more than is seen for other *Paramecium* species (Table 1, Additional file 2: Tables S5). The completeness of the predicted *P. duboscqui* proteins was 97.07% by BUSCO (Additional file 2: Table S6). We also predicted tRNAs using tRNAscan-SE [35] and searched for other non-coding RNA genes against the Rfam database. From most to least numerous, we detected 109 tRNAs, 32 rRNAs, 15 snRNAs, and 2 ribozymes (Fig. 1J).

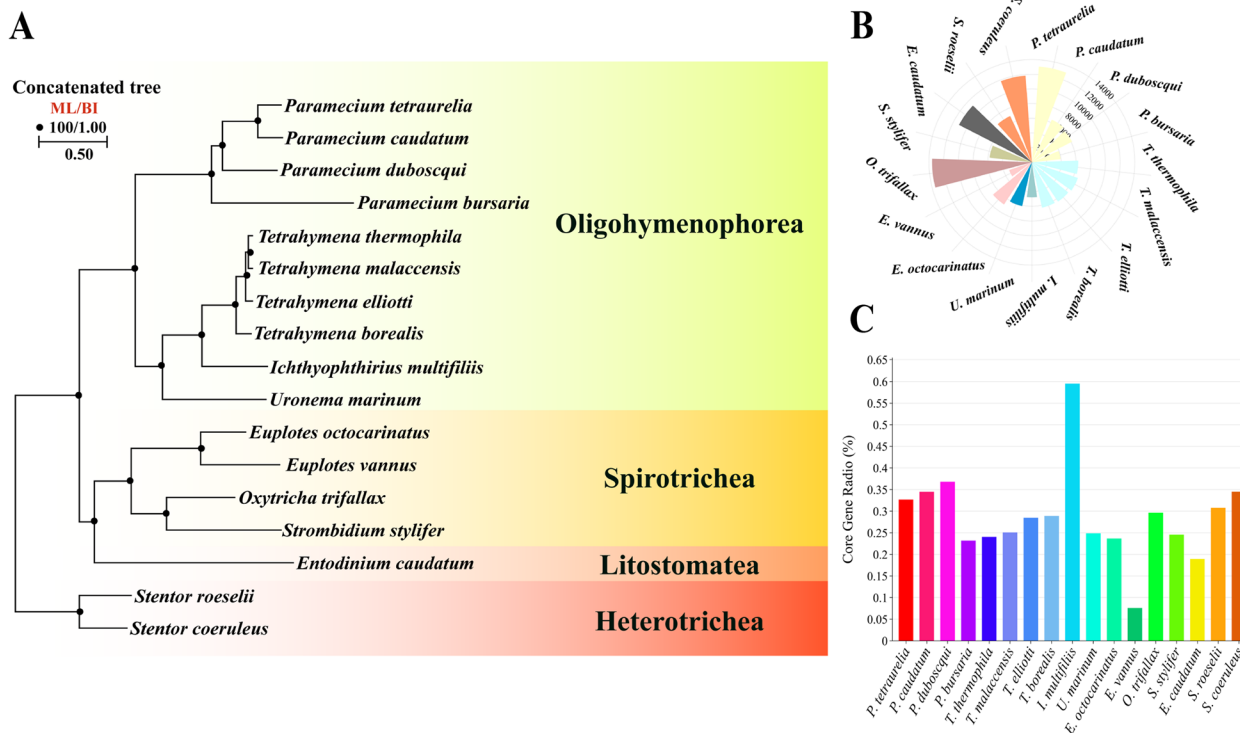
### Evolutionary history of *P. duboscqui*

To anchor the evolutionary position of *P. duboscqui*, we selected 17 typical ciliate species, with two heterotrichs as outgroups, to construct phylogenomic trees (Fig. 2A, Additional file 2: Table S7). Given that *P. duboscqui* belongs to the order Peniculida within the class Oligohymenophorea undoubtedly, in which several well-studied genomes have been reported, we carried out additional comparative genomic analyses to determine the evolutionary position of *P. duboscqui* (Additional file 1: Fig. S6, Additional file 2: Table S8).

For comparative genomics analysis, a total of 417,878 genes were assigned to 60,220 orthogroups. A total of 2556 orthogroups were found in more than 13 ciliates

and thus were considered as ciliate core orthogroups (CCOs) [28]. A total of 6146 of the 16,774 coding genes in *P. duboscqui* could be classified into CCOs. The ratio of CCOs in *P. duboscqui* (36.8%) was markedly higher than that of its relative *P. bursaria* (23.2%) and only lower than that of *I. multifiliis* (59.5%) (Fig. 2B, C). We performed GO functional and KEGG enrichment analyses on the extracted genes and found that among 948 extracted orthogroups, most could be assigned with GO term annotations. GO terms with the most genes were assigned to GO:0022857 (transmembrane transporter activity), GO:0015297 (antiporter activity), GO:0004888 (transmembrane signaling receptor activity), and GO:0055085 (transmembrane transport). The most abundant KEGG pathway assignments included transporter, cysteine and methionine metabolism, peroxisome proliferator-activated receptor (PPAR) signaling, lipid biosynthesis, fatty acid biosynthesis, and fatty acid degradation (Fig. 3A, Additional file 1: Fig. S7). When compared to other *Paramecium* species, the extracted genes were significantly enriched with genes annotated with GO:0022857, GO:0055085, and GO:0015385 (sodium:proton antiporter activity), which may contribute to the ability of *P. duboscqui* to maintain intracellular ion concentration homeostasis in different salinity environments (Fig. 3B).

Previous studies have shown that the ancestor of *Paramecium* and *Tetrahymena* diverged from other ciliates after an ancient whole genome duplication (WGD), and a recent series of WGD events led to the formation of *Paramecium aurelia* [6]. We identified WGD events by a method that calculates the density distribution of synonymous substitution rates per gene (Ks) between collinear paralogous genes. Accordingly, we aligned the protein sequences encoded by the *P. duboscqui* genome against the *P. tetraurelia* and *P. duboscqui* genomes and then calculated Ks values for paralogous gene pairs by filtering using WGDI [36]. The resulting Ks values between orthologous gene pairs in *P. duboscqui* showed a peak distribution ( $K=3$ ), which was also found in *P. tetraurelia*, suggesting that these *Paramecium* species genomes shared a common WGD event during their early evolution. Furthermore, the resulting Ks values between orthologous gene pairs between *P. duboscqui* and *P. tetraurelia* showed that *P. tetraurelia* underwent one additional round of WGD following its divergence from *P. duboscqui* ( $K=1$ ) (Fig. 3C) [6]. In our analysis, we did not find traces of the oldest WGD event known to have occurred in *Paramecium*, which may indicate that synonymy substitution has reached saturation. We identified two peaks from the distribution of Ks values in *P. duboscqui*, the first of which may be caused by the problem of high heterozygosity in the *P. duboscqui* genome assembly



**Fig. 2** Comparative genomic analysis among *P. duboscqui* and its relatives. **A** Phylogenomic tree estimated from a concatenated dataset of 60,220 orthogroups by maximum likelihood (ML) and Bayesian inference (BI) methods. The numbers at the nodes are the bootstrap values of ML out of 1000 replicates and the posterior probability of Bayesian analysis, respectively. The black dots represent the full support values in both ML and BI trees. The scale bar corresponds to 50 substitutions per 100 nucleotide positions. **B** The Rose diagram shows the gene number in CCOs for each species. **C** The proportion of CCO genes in each ciliate

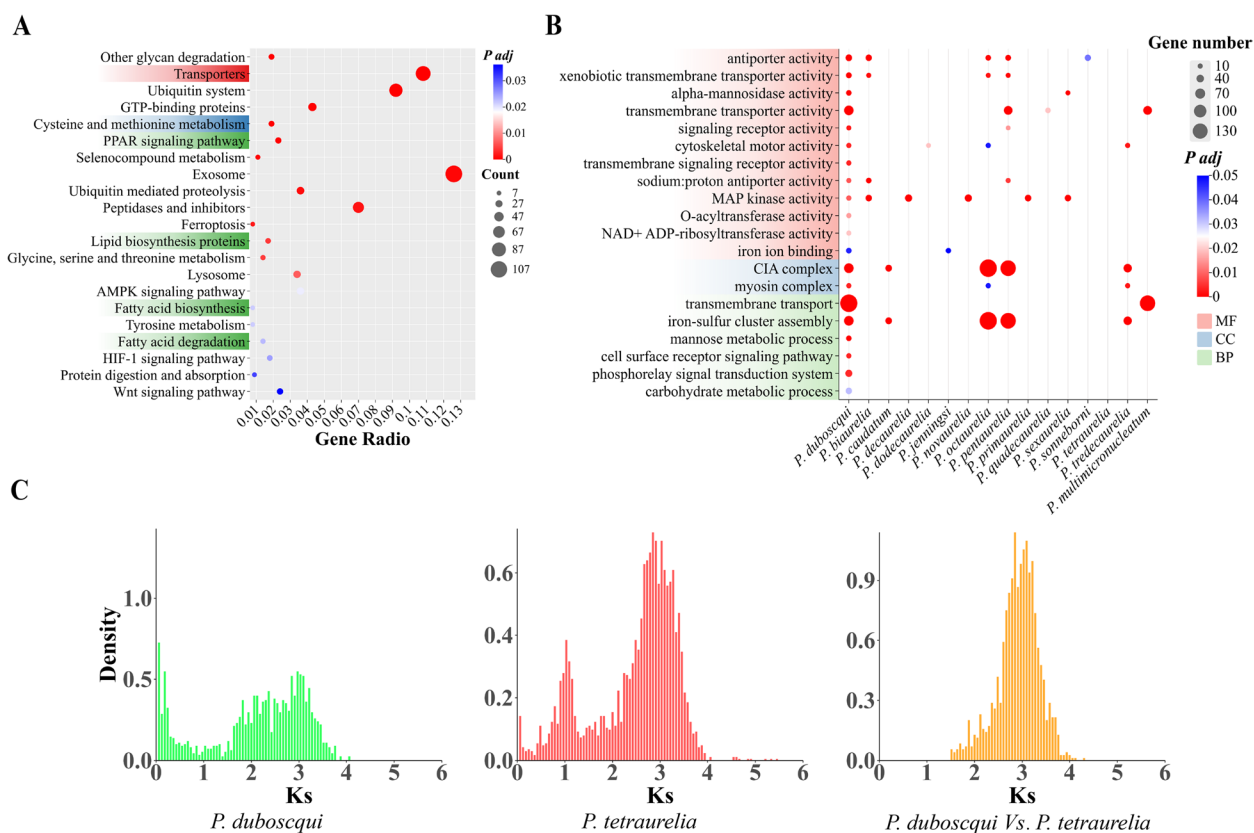
arising from its extreme polyploidy. Alternatively, this peak may suggest that *P. duboscqui* has very recently undergone a round of WGD.

#### Codon usage bias for *P. duboscqui*

To explore codon usage patterns of *P. duboscqui*, we scanned its protein-coding sequences for codons biases, as well as for stop codon reassignments and alternative genetic codes, which are common among ciliates. The result indicated that like other *Paramecium* species, *P. duboscqui* employed only TGA as a termination codon, with TAA and TAG reassigned to encode glutamine (Additional file 2: Table S9). Nucleotide composition bias, specifically the GC content, is the prime factor of codon usage bias [37]. Genomic GC content and GC content at different codon positions of *P. duboscqui* are higher than the other *Paramecium* species, especially at the third of codon position (Additional file 2: Table S10). Relative synonymous codon usage (RSCU) values suggested that *P. duboscqui* prefers to use codons ending with A/T, which is similar to other *Paramecium* (Fig. 4A). However, unlike the other *Paramecium*, the putative optimal codon of lysine is AAG rather than AAA, which may be

the reason for the difference in the GC content of protein coding sequences at the third codon position between *P. duboscqui* and the other *Paramecium* (Additional file 1: Table S11). The effective number of codons (ENC) is a typical parameter to measure the magnitude of synonymous codon usage bias for any gene [38]. ENC values range from 20 (extreme bias for using one codon) to 61 (no bias for using synonymous codons) [39, 40]. The ENC value of *P. duboscqui* is higher than the other *Paramecium* at around 41, which indicated that *P. duboscqui* has a weaker codon usage bias (Additional file 2: Table S12). A PR2-Plot showed that the base bias at the 3rd codon position is not obvious between A, T and G, C (Fig. 4B).

Neutrality and ENC plots were produced to explore factors influencing the codon usage pattern in *P. duboscqui*. Previously, it has been shown that a regression line slope of the neutrality plot close to or equal to 1 indicates that mutation pressure is the sole determinant of codon usage bias; a slope closer to 0 indicates that natural selection is the major factor [41]. As shown in Fig. 4C, the slope of the regression line value was 0.136 and the correlation between GC12 and GC3 was low ( $R=0.27$ ,  $p$ -value  $< 2.2e-16$ ), which suggests that natural selection



**Fig. 3** Functional enrichment of extracted gene family and WGD of *P. duboscqui* and its relatives. **A** KEGG enrichment analysis on the extracted genes of *P. duboscqui* with other 16 ciliates (adjusted *p*-value by Benjamini–Hochberg procedure < 0.05). **B** GO term analysis on the extracted genes of *P. duboscqui* and other *Paramecium* (adjusted *p*-value by Benjamini–Hochberg procedure < 0.05). **C** Whole-genome duplication events during the evolution of *P. duboscqui* and *P. tetraurelia*, as evidenced by the distribution of the peak Ks values

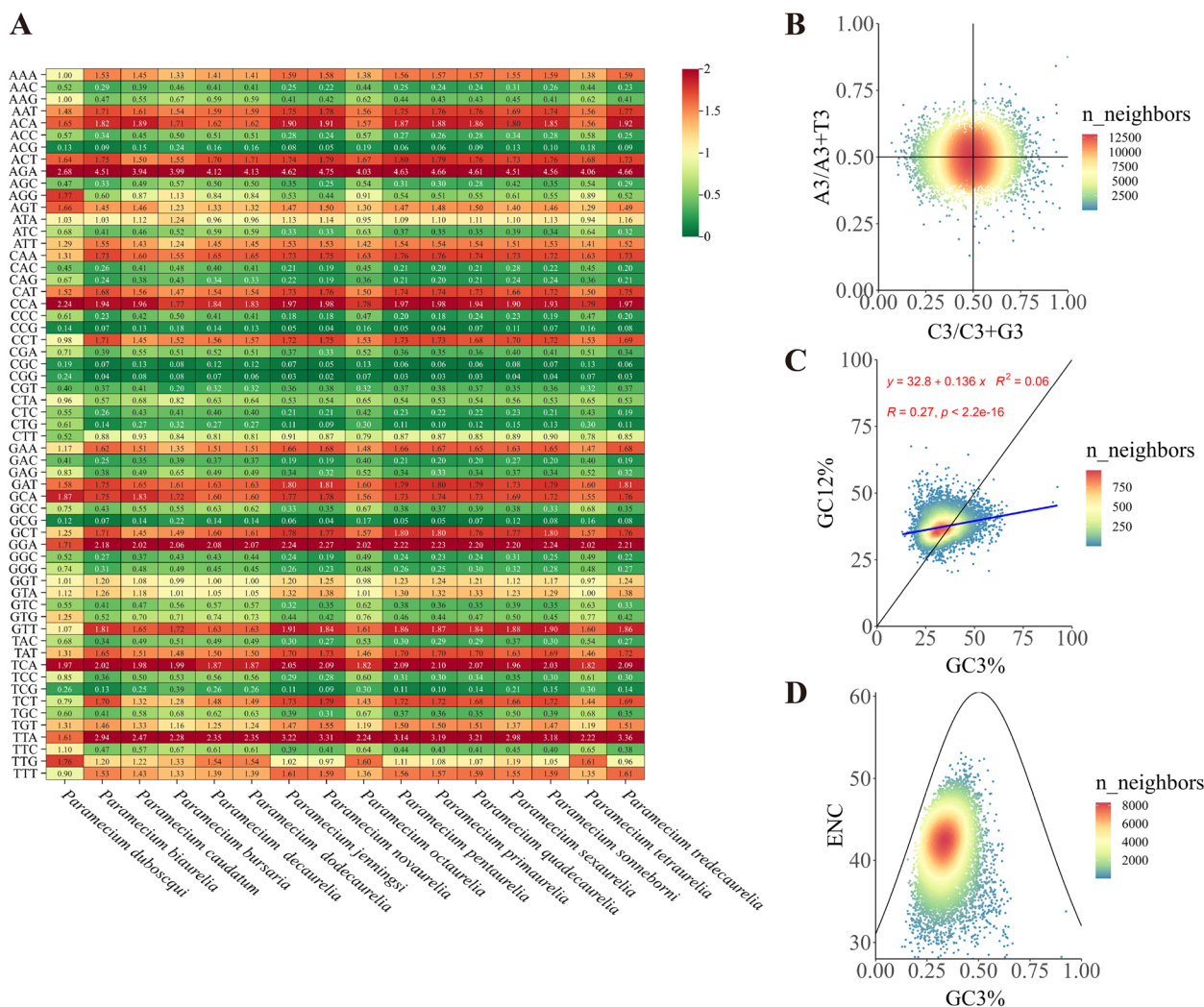
was the main influence on codon usage bias in *P. duboscqui* (Additional file 2: Table S13). It is notable that the slope of this regression line for *P. duboscqui* is greater than those of other *Paramecium*, which indicates that the mutation pressure is stronger in this species and that natural selection is weaker than in other *Paramecium*, resulting in a weaker codon usage bias. If codon usage were limited only by mutation bias, the predicted ENC value would be on or near the standard curve [39]. As shown in Fig. 4D, most of the genes were located above or below the standard curve, suggesting that natural selection plays the main function in codon usage bias.

**Genetic basis for salinity tolerance**

To better understand the salinity tolerance of *P. duboscqui*, we calculated 30-day growth curves under a range of salinities, defined here as Salinity‰ = 0.030 + 1.8050Cl<sup>-</sup>‰, between 0‰ and 30‰ (Fig. 5A). All groups under 30‰ reached logarithmic growth and a stable population by the 8th day. On

the 22nd day, the 30‰ group reached the logarithmic growth phase. From these results, we determined that 5‰ salinity provides the optimum growth rate for *P. duboscqui*. Meanwhile, the length–width ratio of the cells at 30‰ salinity was significantly lower than 0‰ and 5‰ groups on the 30th day (Fig. 5B).

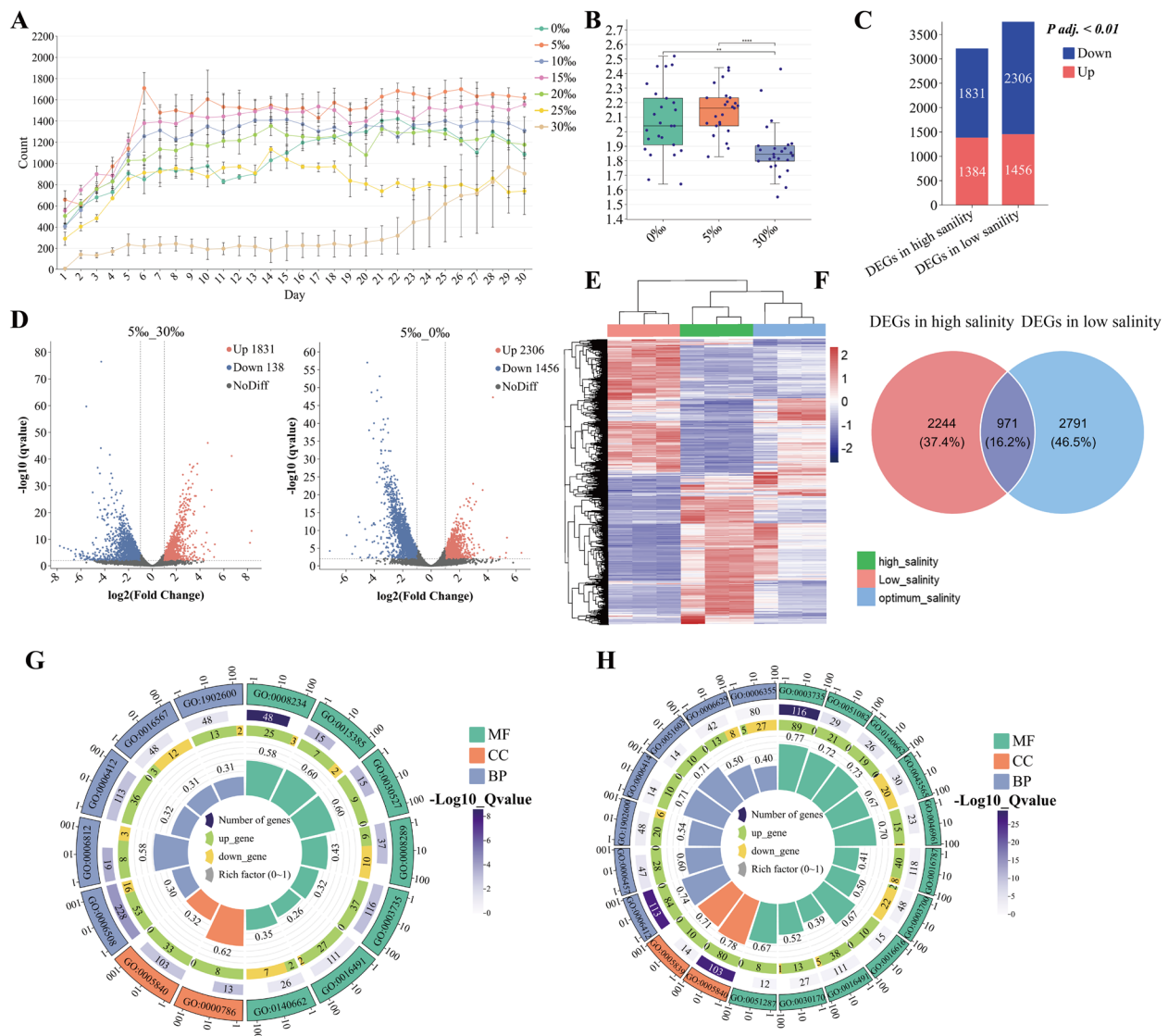
To investigate the genetic adaptations that allowed *P. duboscqui* to survive in a range of saline concentrations, we calculated its gene expression profiles at three different salinity levels: 0‰ (low), 5‰ (optimum), and 30‰ (high). A total of 3215 differentially expressed genes (DEGs) were found between the high and optimum salinities, and 3762 DEGs were found between the optimum and low salinities (Fig. 5C, D). Interestingly, only 19.2% of the DEGs were shared in both groups, and heatmap clustering analysis of all DEGs showed very little overlap between the low and high salinity groups (Fig. 5E, F), suggesting that *P. duboscqui* may have distinct mechanisms to adapt to low and high salinity environments.



**Fig. 4** Codon usage pattern of *P. duboscqui*. **A** The RSCU values of the codons of *P. duboscqui*. A gradient from dark blue to dark red indicates the RSCU value increases from low to high. **B** PR2-bias plot of A3/(A3 + T3) against G3/(G3 + C3) of *P. duboscqui*. **C** Neutrality plots, showing the relationship between GC3 and GC12 of *P. duboscqui*. **D** Relationship between the ENC and GC3 content at the third codon position of *P. duboscqui*

Functional enrichment analysis of DEGs in the high salinity group showed that upregulation of genes were annotated with the GO terms GO:0003735 (structural constituent of ribosome), GO:0000786 (nucleosome), GO:0005840 (ribosome), and GO:0006412 (translation), indicating that *P. duboscqui* initiated protein translation processes at a high rate in response to a high salinity environment. These findings were further supported by KEGG functional enrichment analysis (Fig. 5G, Additional file 1: Fig. S8, Additional file 1: Fig. S9, Additional file 2: Table S14). Genes upregulated were enriched in GO terms including GO:0015385 (sodium:proton antiporter activity), GO:0006812 (cation transport), and GO:1902600 (proton transmembrane transport), showing

that *P. duboscqui* responded to the high osmotic pressure environment by controlling the flow of positive ( $\text{Na}^+/\text{H}^+$ ) ions. Previous studies have shown that reactive oxygen species (ROS) induced from salt stress led to the oxidative modification of cysteine, which in turn promotes tolerance to salt stress [42, 43]. In this study, we found that genes with GO terms including GO:0008234 (cysteine-type peptidase activity) and GO:0016491 (oxidoreductase activity) were upregulated, suggesting that oxidative modification of cysteine may promote osmotic defense of *P. duboscqui* in high salt environments. KEGG functional enrichment analysis of DEGs showed that most DEGs under high salt encode chaperones and transporters or proteins involved in amino acid and glycerolipid



**Fig. 5** Genetic adaptation for salinity tolerance of *P. duboscqui*. **A** Population density of *P. duboscqui* grown in different salinities. **B** Length–width ratio at different salinities on the 30th day (0‰, 5‰, 30‰). **C** The number of differentially expressed genes in the high (30‰) and low (5‰) salinity groups. Both groups were compared against the optimum (5‰) salinity group. **D** Volcano map of differentially expressed genes. **E** Heatmap for differentially expressed genes. **F** Venn diagram of differentially expressed genes in the two groups. **G** GO term enrichment analysis on high salinity DEGs of *P. duboscqui* (adjusted *p*-value by Benjamini–Hochberg procedure < 0.05). **H** GO term enrichment analysis on low salinity DEGs of *P. duboscqui* (adjusted *p*-value by Benjamini–Hochberg procedure < 0.05)

metabolism. These findings mirror the results of previous studies, which showed that activation of protein folding pathways is needed to help organisms deal with misfolded proteins resulting from the accumulation of ROS under salt stress [44]. Transporters play a key role in the salt adaptation response as part of the SOS (salt overly sensitive) signaling pathway [45]. Glycerophospholipid and dissociative amino acids, such as proline and alanine, are also important osmotic pressure modulators [46, 47].

Functional enrichment analysis of DEGs in the low salinity group showed downregulated GO terms related to translation (GO:0003735, GO:0005840, and GO:0006412), indicating that *P. duboscqui* slowed this process in low salinity environments, which was different from the high salt group (Fig. 5H, Additional file 1: Fig. S10, Additional file 1: Fig. S11, Additional file 2: Table S14). Instead, transcription initiation genes with GO terms including GO:0006355 (regulation of transcription, DNA-templated), GO:0043565



(sequence-specific DNA binding), and GO:0003700 (DNA-binding transcription factor activity) were upregulated under low salt, which showed that *P. duboscqui* may repress expression under these conditions by binding transcription factors to DNA. Genes annotated to GO:0046961 (proton-transporting ATPase activity, rotational mechanism) and GO:1902600 were also downregulated in low salinity, showing the rate of ion exchange was inhibited. Conversely, we found that chaperones and metabolic enzymes were downregulated in low salinity indicating that *P. duboscqui* growth did not depend on misfolded protein repair and oxidation–reduction processes under these conditions as they did in high salinity.

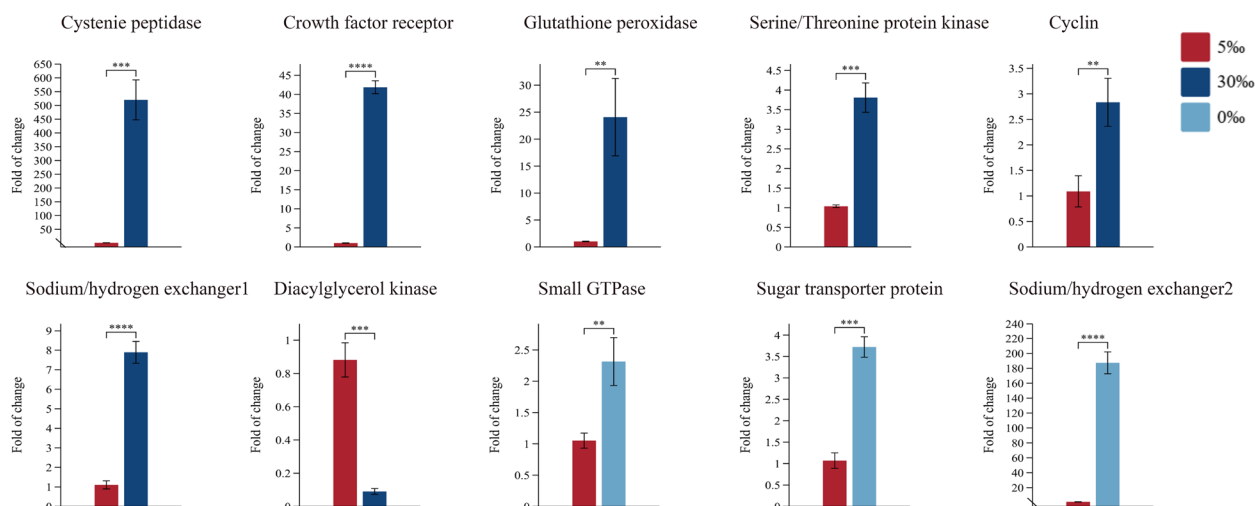
### RT-qPCR validation of RNA sequencing

In order to validate the results obtained from transcriptome analyses, ten genes that showed different expression patterns from RNA-seq were selected for RT-qPCR. Genes encoding homologs of a cysteine peptidase (CP), growth factor receptor (CFR), glutathione peroxidase (GSHPx), serine/threonine protein kinase (STPK), cyclin, Na<sup>+</sup>/H<sup>+</sup> exchanger (NHE), diacylglycerol kinase (DAGK), small GTPase (Ras), and sugar transporter protein (STP) were differentially expressed under varying salinity conditions. Of these, six differentially expressed genes (CP, CFR, GSHPx, STPK, cyclin, NHE1) were upregulated, and one differentially expressed gene (DAGK) were downregulated in the high-salt group, while the other three differentially expressed genes (Ras, NHE2, and STP) were upregulated in the low-salt group. The change in expression observed for these genes under their respective conditions using RT-qPCR and RNA-seq was largely consistent (Fig. 6, Additional file 2: Table S15, Additional file 2: Table S16).

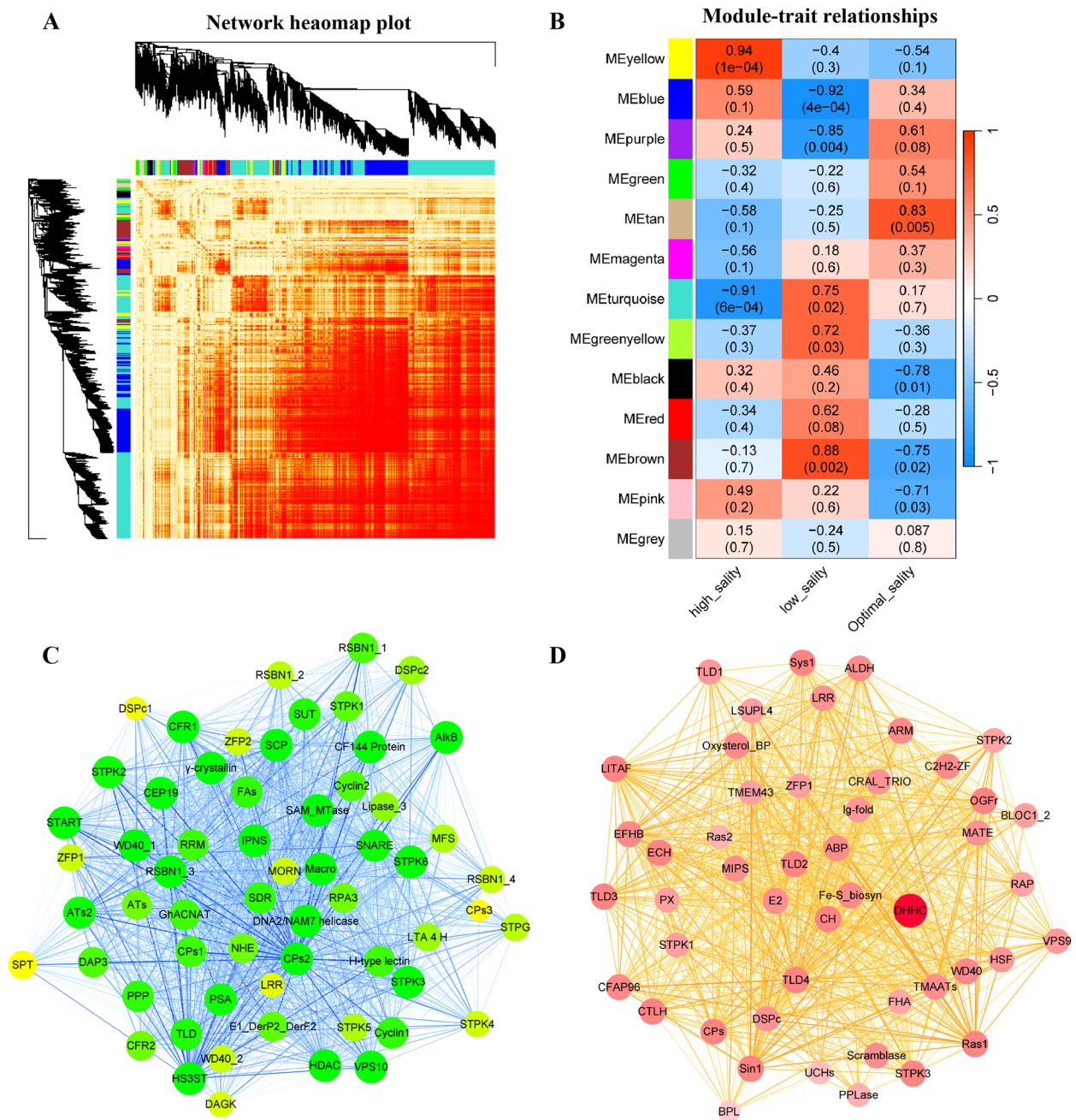
### Weighted gene co-expression network (WGCNA) construction

*P. duboscqui* showed different phenotypic characteristics and growth rates in different salinity environments, and transcriptome differential expression analysis showed that *P. duboscqui* adapted to high salt and low salt environments by expressing different sets of genes. To identify hub genes likely to play roles in these osmotic stress responses, we constructed correlations between gene expression modules and different salinity values (0‰, 5‰, and 30‰). After filtering out 3810 genes (the last 25% of median absolute deviation, MAD), 11,428 gene expression values were utilized to create co-expression modules (Additional file 1: Fig. S12). The automatic network construction function blockwiseModules clustered these genes into 13 modules based on expression pattern (Fig. 7A, Additional file 1: Fig. S13). We used Pearson's method to further statistically compute the correlation coefficients for salinity-to-module eigenvalue correlations and utilized a “*t* test” to obtain *p* values.

Two modules that were significantly correlated with changes in salinity were identified by this method (yellow module-30‰\_salinity,  $r=0.94$ ,  $p < 1e-4$ ; brown module-0‰\_salinity,  $r=0.88$ ,  $p < 0.002$ ) (Fig. 7B). We obtained 59 and 48 core genes for the yellow and brown modules, respectively, and visualized their protein interaction relationships using Cytoscape (Fig. 7C, D). Hub genes in the response to 0‰ salinity encoded protein localization and transmembrane transport proteins, whereas hub genes in the response to 30‰ salinity were involved in cysteine-type peptidase activity, lipid transfer, sodium/hydrogen exchange, and cell division. These results were consistent with those seen during the analysis of differentially expressed genes.



**Fig. 6** Relative transcription levels of RT-qPCR. Error bars indicate  $\pm$  S.D. of biological triplicates



**Fig. 7** WGCNA analysis of different salinities stress experiments using RNA high-throughput sequencing. **A** Gene co-expression clustering tree and co-expression topology heat map. Note, stronger colors represent higher expression similarity between genes. **B** The constructed relationship matrix of modules-traits, where positive values represent module eigengenes positively correlated with the trait, and negative values represent their negative correlation with the trait. The significance threshold was set to 0.15. **C** Core genes arrangement for the yellow module presented in **B**. **D** Core genes arrangement for the brown module presented in **B**

### Discussion

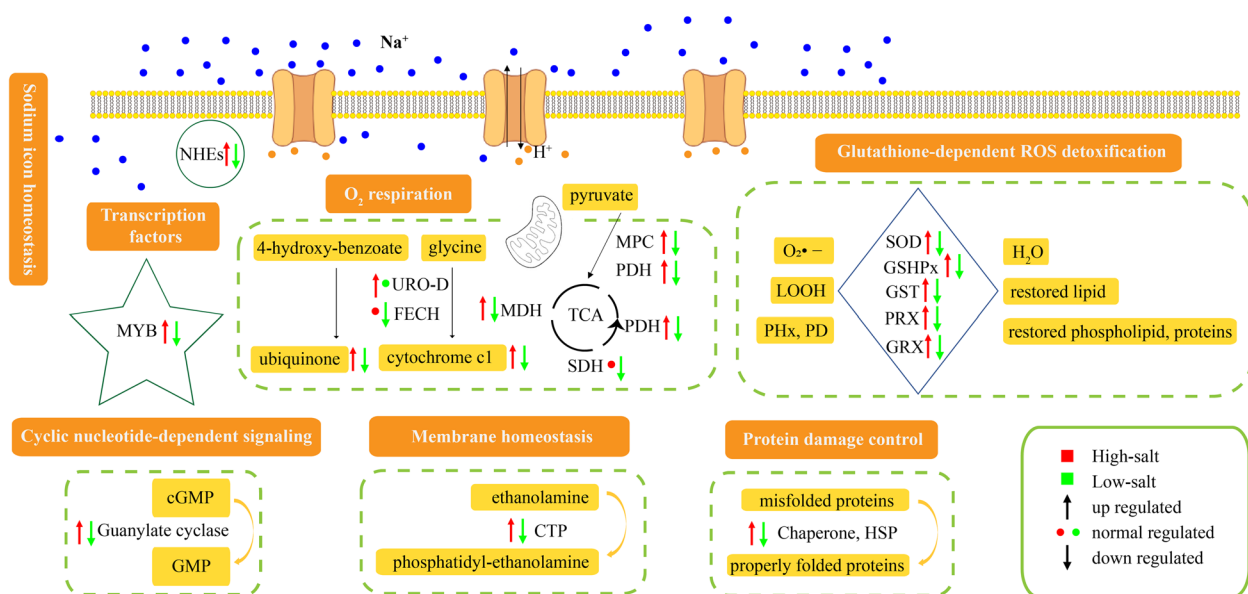
*Paramecium duboscqui* widely distributes in fresh, brackish, marine water habitats, which is a unique and representative euryhaline *Paramecium* species [15]. The

analyses of the genome of *P. duboscqui* are vital for the subsequent study of its salinity adaptation mechanism and the further study of *Paramecium* genome evolution model.

In our study, PacBio HiFi sequencing was employed to de novo assembly the genome of *P. duboscqui*. The MAC genome of *P. duboscqui* comprises 160 contigs with 113 of them possessing telomere, the ratio of which is far more than the other *Paramecium* genomes by next-generation sequencing. This demonstrates the superiority of the third generation sequencing in the assembly of *Paramecium* genomes. As one of the oldest branches of *Paramecium* species, *P. duboscqui* is highly similar to the other *Paramecium* species in the intron length, intron conserved motif, and intergenic region length [4]. However, the GC content of *P. duboscqui* is higher than the other *Paramecium* species, which was reflected in the difference of codon usage pattern. The Ks value distribution indicated that *P. duboscqui* and *P. tetraurelia* shared a common WGD event during their early evolution [6].

Previous studies have shown that changes in salinity can have wide-ranging effects on cells, particularly in the ER where high salt conditions have been found to disrupt protein homeostasis, redox homeostasis, and calcium homeostasis [48–51]. All of these conditions lead to misfolding or denaturation of proteins in the ER. When the protein folding burden of the ER exceeds its folding capacity, cells typically express chaperones to increase the folding capacity of the ER, along with other proteins to stall most protein translation processes and accelerate protein degradation, collectively referred to as the unfolded protein response (UPR) [52]. Maintaining ER homeostasis must therefore play a central role in *P. duboscqui* adaptation to different salinity environments.

The *P. duboscqui* differential gene expression and WGCNA analyses produced a concise list of genes involved in the response to different salinities, along with putative functions for many of these genes in a variety of metabolic pathways. Here, we present a model for how these genes work together in response to high salinity environments (Fig. 8). Upon encountering a high salt environment, expression of genes coding for transcription factors, such as the MYB family of proteins, are increased to promote the expression of additional genes. These positively regulated target genes likely encode some or all of the following proteins and protein functions: (1) NHEs, which are responsible for the extrusion of sodium ions to maintain ion homeostasis at high salt concentrations; (2) CTP (phosphoethanolamine cytidyltransferase), which is involved in salt-dependent adjustment of the membrane lipid composition in order to maintain cell membrane fluidity; (3) proteins involved in oxygen respiration, to promote the redox reactions needed to provide energy for maintaining cell homeostasis; (4) ER chaperones and cytosolic HSR (heat shock response) proteins, to repair cellular proteins that become denatured in the high-salt environment; (5) glutathione-dependent detoxification proteins needed to manage the oxidative stress induced by high salt conditions; (6) genes involved in cGMP-dependent signal transduction, which can stimulate additional gene expression. Many of these genes also show a corresponding downregulation under low salt conditions, further highlighting their roles in the adaptation of *P. duboscqui* to varying salt conditions.



**Fig. 8** Schematic representation of biological pathways in *P. duboscqui* affected by different salinities

## Conclusions

In the current study, we used both Illumina and PacBio sequencing technologies to assemble a complete and high-quality MAC genome of the representative euryhaline *Paramecium duboscqui*. The genome of this species is extraordinarily rich in coding sequences and has undergone at least one round of WGD, making it an ideal model to study ciliate genome expansion and contraction events. When facing changes in environmental salinity, *P. duboscqui* increases transcription of genes needed for ion exchange, cell membrane fluidity, respiration, and glutathione-dependent detoxification, making it one of the rare halotolerant eukaryotic microorganisms. Our study provides a unique and valuable resource on *Paramecium* biology and establishes the genomic tools needed for future discoveries in this unique model organism.

## Methods

### Sample collection and cultivation

*P. duboscqui* cells were collected from marine water habitat in Weihai, Shandong, China (122° 1' 30' N; 37° 30' 43' E; 26 °C), isolated under the stereomicroscope, and then washed several times with sterilized seawater. Species identification was conducted using morphological characteristics and SSU rRNA gene sequencing. Clonal cultures were established using *Klebsiella pneumoniae* cells as the food source at 20 °C. After 4–5 days of culture, the population density reached 3000 cells/ml of medium.

### Nucleic acid isolation and sequencing

*P. duboscqui* cells were isolated through nylon filter membrane and washed three times by Tris–HCl (pH 7.5). DNA extraction was carried out using the MasterPure™ Complete DNA & RNA Purification Kit following the manufacturer's instructions (Qiagen, Germany). Approximately 300,000 cells were harvested and generated 30 µg DNA for PacBio HiFi sequencing. PacBio single-molecule real-time (SMRT) bell library preparation was performed using SMRTbell® Express Template Prep Kit 2.0 (Pacific Biosciences, PN 101–853-100) in accordance with the manufacturer's instructions. The library was prepared for sequencing on the Sequel II/IIe system to create a 30 h Movie file (Pacific Biosciences, CA, USA).

In addition, the Illumina paired-end sequencing libraries were generated using the NEB Next® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations. The DNA libraries were sequenced on an Illumina NovaSeq 6000 platform, and 150 bp paired-end reads were generated.

For the transcriptome, *P. duboscqui* total RNA and RNA from cells grown at different salinities (0‰, 5‰,

30‰) over 30 days were extracted after 2 h of starvation with TRIzol Reagent (Molecular Research Center, Inc., USA), and sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations. The library preparations were sequenced on an Illumina NovaSeq 6000 platform, and 150 bp paired-end reads were generated.

### Genome assembly

We first utilized the Fastp (v0.23.1) [53] software to filter out low-quality and adapter sequence from raw Illumina DNA reads. The Jellyfish (v2.01) [54] and GenomeScope (v2.0) [55] software with parameter (K-mer=17) were used to estimate genome size and heterozygosity. MAC genome was assembled using Canu (v2.2) [56] with parameters (-pacbio-hifi, GenomeSize=25 M). Mitochondrial and bacterial genome sequences were downloaded from GenBank as databases to remove contamination caused by mitochondria or bacteria. The contigs assembled by Canu with tag “suggestBubble=yes” or “suggestRepeat=yes” were regarded as redundant sequences and were eliminated. The clean PacBio and Illumina DNA reads were aligned to the *P. duboscqui* genome using the minimap2 (v2.26) [57] software and Bowtie2 (v2.5.2) [58] with default parameters, respectively. Due to the peculiarity of the short introns in *Paramecium* species, we compiled the source code of Hisat2 (v2.1.1) [59] to guarantee more accurate intron prediction. The clean RNA reads were aligned to the *P. duboscqui* genome using Hisat2 with parameter (-minintronlen 9). Telomeres were detected using a custom Python script that recognized the telomere repeat 6-mer 5'-(C4A2)n-3' or 5'-(T2G4)n-3' at the ends of contigs. BUSCO (v5.4) was used to assess the genome completeness with default parameters.

### Gene prediction and annotation

Somatic MAC genome assembly of *P. duboscqui* was annotated through the following pipeline. First, the repetitive sequence families were predicted using Repeatmolder [60]. Second, transcriptome reads were aligned to the MAC genome by Hisat2 and were assembled to transcripts using Trinity (v2.15.1) [61]. Third, the exact locations of introns were predicted using Truc (v1.1). Finally, the protein sequences of *P. tetraurelia* were selected as homologous proteins, and gene models were predicted by Eugene (v1.5), combined with repetitive sequence families, transcripts, homologous proteins, and intron positions as input. All protein-coding genes were annotated using six databases: NR, SwissProt, eggNOG, and Interpro, GO, and KEGG. The tRNA and other ncRNA

genes were detected in the MAC genome by tRNAscan-SE (v1.3.1) and Rfam (v11.0) [62], respectively.

### Phylogenetic analysis

OrthoFinder (v2.5.4) [63] was used to infer orthogroups based on 17 ciliate protein sequences with parameters (-M msa -S diamond) (Additional file 2: Table S17). Sequences in each orthogroup were aligned using MAFFT (-auto), trimmed with TrimAl (v1.2) (-automated1) and concatenated [64]. A supermatrix containing 267,094 amino acid sites was then generated. Phylogenies were inferred by both maximum likelihood (ML) and Bayesian inference (BI) methods implemented in the RAXML (v8.2.12) [65] (model=LG+Γ+F, rate categories=4, bootstrap=1000) and PHYLOBAYES MPI 1.5a (model=CAT-GTR+Γ, chains=4, generations=10,000, Maxdiff<0.3, burn-in=10%) [66], respectively. Phylogenomic trees were visualized using MEGA (v7.0.20) [67]. Divergence times were estimated using r8s (v1.81) [68]. Calibration times were obtained on the TimeTree website (<https://www.timetree.org/>) showing that the divergence time between *Tetrahymena thermophila* and *P. tetraurelia* was ~918 Ma.

### Orthology analysis

OrthoFinder was used to infer orthogroups based on 17 ciliate species, and a total of 417,878 (95.81%) of 436,163 genes were assigned to 60,220 orthogroups. Among them, 2921 orthogroups were considered CCOs, as they were detected in 13 species.

CAFE (v4.0) [69] was used to determine gene family evolution among *P. duboscqui* and the related ciliates based on the results of OrthoFinder and the ultrametric tree extracted by r8s. Among the 4501 orthogroups in *P. duboscqui*, we detected 3553 contraction events (involving 1507 genes) and 948 expansion events (involving 4273 genes).

WGDI (v0.54) was used with default parameters to detect gene duplication events in *P. duboscqui* and *P. tetraurelia*.

### Codon usage bias

The index of codon usage bias was calculated using BCAWT (Bio Codon Analysis Workflow Tool) [36]. Codetta (version 2.0) was used to predict the genetic code from nucleotide sequence data [70].

### Growth rate at different salinity levels

We set up 7 consecutive salinities (Salinity‰ = 0.030 + 1.8050Cl<sup>-</sup>‰) gradients between 0‰ and 30‰, with three parallel experiments for each group during one month. The volume of each culture flask was controlled at

200 ml, and the initiation number was about 200 cells per ml. *Escherichia coli* was provided as food. We cultivated for 30 days to acquire growth curves of *P. duboscqui*. As a result, the optimum growth rate was determined to be 5‰ salinity. We grew cells at 0‰ (the minimum salinity), 5‰ (the optimum salinity), and 30‰ (the maximal salinity) with three replicates for the following RNA-seq.

### Transcriptome sequencing and functional enrichment analysis

We performed the following standard processing protocol on gene expression data for *P. duboscqui* grown in different salinities (0‰, 5‰, 30‰). The clean data reads were aligned to the *P. duboscqui* genome using the Hisat2 (v2.1.1) software with parameter (-min-intronlen 9). FeatureCounts (v1.3.1) [71] was then employed to determine the count of reads aligned to each gene in the reference genome. Subsequently, the functions variance-stabilizing transformation (VST) and principal component analysis (PCA) in DESeq2 were used to normalize count data and evaluate the similarity among samples. Differential expression analysis was conducted using the DESeq2 [72]. According to the expression criterion ( $|\log_2\text{fold change}| \geq 1$ ) and a *p*-adj value < 0.05, we detected differentially expressed genes in the comparisons of 5‰ vs. 0‰ and 5‰ vs. 30‰, respectively.

The GO and KEGG function enricher in the R package “clusterProfiler” (v3.12.0) (<https://github.com/YuLab-SMU/clusterProfiler>) were used to perform enrichment analysis. Cytoscape (<https://cytoscape.org/>) was used to visualize the network.

### Abbreviations

MIC	Micronucleus
MAC	Macronucleus
HiFi	High fidelity
CCO	Ciliate core orthogroup
WGD	Whole genome duplication
Ks	Synonymous substitution rates
RSCU	Relative synonymous codon usage
ENC	Effective number of codons
DEGs	Differentially expressed genes
ROS	Reactive oxygen species
SOS	Salt overly sensitive
CP	Cysteine peptidase
CFR	Growth factor receptor
GSHPx	Glutathione peroxidase
STPK	Serine/threonine protein kinase
NHE	Na <sup>+</sup> /H <sup>+</sup> exchanger
DAGK	Diacylglycerol kinase
Ras	GTPase
STP	Sugar transporter protein
WGCNA	Weighted gene coexpression network analysis
MAD	Median absolute deviation
UTR	Unfolded protein response
SMRT	Single-molecule real-time
BCAWT	Bio Codon Analysis Workflow Tool
VST	Variance-stabilizing transformation
PCA	Principal component analysis

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-02026-5>.

Additional file 1: Figs. S1–S13: Fig. S1. Photomicrographs of *P. duboscqui* from cells in vivo (A, C) and after staining (B, D, E). Fig. S2. This diagram illustrates the workflow employed for the genome assembly and annotation of *P. duboscqui* in this study. Fig. S3. The estimation of *P. duboscqui* genome size. Fig. S4. The GC content of *P. duboscqui* genome. Fig. S5. Weblogo plot of 22~26 bp intron of *P. duboscqui*. Fig. S6. Phylogenomic tree estimated from a concatenated dataset of 36,900 orthogroups by maximum likelihood (ML) and Bayesian inference (BI) methods. Fig. S7. GO term analysis on extracted gene of *P. duboscqui*. Fig. S8. GO term enrichment analysis on high salinity DEGs gene of *P. duboscqui*. Fig. S9. KEGG enrichment analysis on high salinity DEGs gene of *P. duboscqui*. Fig. S10. GO term enrichment analysis on low salinity DEGs gene of *P. duboscqui*. Fig. S11. GO term enrichment analysis on low salinity DEGs gene of *P. duboscqui*. Fig. S12. Scale-free topological analysis to determine optimal beta and co-expression matrix and module determination for WGCNA. Fig. S13. Construction of gene dendrograms and module colors for identifying co-expression modules in WGCNA.

Additional file 2: Tables S1–S17: Table S1. Statistics for the sequencing data of the *P. duboscqui* genome. Table S2. Estimation of genome size and heterozygosity. Table S3. Features from different assembly strategies used for the *P. duboscqui* genome. Table S4. Result of BUSCO for *P. duboscqui* genome and the other closely related species. Table S5. The number of telomeres in the MAC genome of *Paramecium*. Table S6. Result of BUSCO for *P. duboscqui* proteins. Table S7. 17 typical ciliate species for comparative genomic analyses. Table S8. 17 *Paramecium* and 4 *Tetrahymena* species for comparative genomic analyses. Table S9. The codon table of *P. duboscqui*. Table S10. Nucleotide composition of *Paramecium*. Table S11. The putative optimal codons of *Paramecium*. Table S12. The average ENC value of *Paramecium*. Table S13. The slope of Neutrality plot of *Paramecium*. Table S14. GO term functional annotation. Table S15. Annotation of genes involved in RT-qPCR. Table S16. Primer sequence of genes involved in RT-qPCR. Table S17. The accession number of ciliate species.

## Acknowledgements

We acknowledge the computing resources provided on IEMB-1, a high-performance computing cluster operated by the Institute of Evolution and Marine Biodiversity (Ocean University of China). Our thanks are due to Prof. Weibo Song (Ocean University of China) for his kind help and advice on preparing the manuscript. We also would like to thank Dr. Mingzhen Ma (Shandong University, China), Mr. Zhiwei Wen (Shandong University, China) and Ms. Congjun Li (Shandong University, China) for their help in experiment assistance.

## Authors' contributions

Y.F., F.L. and Y.Z. performed the analysis. Y.F., N.P. and Y.Z. performed experiments. Y.F. wrote the original draft. N.A.S. and L.L. revised the manuscript. L.L. did the conceptualization, review & editing the manuscript, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

## Funding

This work was supported by the National Natural Science Foundation of China (project number: 31772431).

## Availability of data and materials

The datasets presented in this study can be found in National Center for Biotechnology Information Database with accession number shown in Additional file: Table S17. All *Paramecium* information can be found in ParameciumDB [73].

## Data availability

No datasets were generated or analysed during the current study.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Laboratory of Marine Protozoan Biodiversity and Evolution, Marine College, Shandong University, Weihai 264209, China. <sup>2</sup>Department of Biology, Bradley University, Peoria 61625, USA.

Received: 10 July 2024 Accepted: 30 September 2024

Published online: 15 October 2024

## References

- Caron DA, Countway PD, Jones AC, Kim DY, Schnetzer A. Marine protistan diversity. *Annu Rev Mar Science*. 2012;4:467–93. <https://doi.org/10.1146/annurev-marine-120709-142802>.
- Fernandes NM, Campello-Nunes PH, Paiva TS, Soares CAG, Silva-Neto ID. Ciliate diversity from aquatic environments in the Brazilian Atlantic Forest as revealed by high-throughput DNA sequencing. *Microb Ecol*. 2021;81:630–43. <https://doi.org/10.1007/s00248-020-01612-8>.
- Weisse T. Functional diversity of aquatic ciliates. *Eur J Protistol*. 2017;61:331–58. <https://doi.org/10.1016/j.ejop.2017.04.001>.
- Long H, Johri P, Gout JF, Ni J, Hao Y, Licknack T, Wang Y, Pan J, Jiménez-Marín B, Lynch M. *Paramecium* genetics, genomics, and evolution. *Annu Rev Genet*. 2023;57:391–410. <https://doi.org/10.1146/annurev-genet-071819-104035>.
- Riley JL, Katz LA. Widespread distribution of extensive chromosomal fragmentation in ciliates. *Mol Biol Evol*. 2001;18:1372–7. <https://doi.org/10.1093/oxfordjournals.molbev.a003921>.
- Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N, Arnaiz O, Billaut A, Beisson J, Blanc I, Bouhouche K, Câmara F, Duhaucourt S, Guigo R, Gogendeau D, Katinka M, Keller AM, Kissmehl R, Klotz C, Koll F, Le Mouél A, Lepère G, Malinsky S, Nowacki M, Nowak JK, Plattner H, Poulain J, Ruiz F, Serrano V, Zagulski M, Dessen P, Bétermier M, Weissenbach J, Scarpelli C, Schächter V, Sperling L, Meyer E, Cohen J, Wincker P. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*. 2006;444:171–8. <https://doi.org/10.1038/nature05230>.
- Le Mouél A, Butler A, Caron F, Meyer E. Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. *Eukaryot Cell*. 2003;2:1076–90. <https://doi.org/10.1128/EC.2.5.1076-1090.2003>.
- Fokin SI. *Paramecium* genus: biodiversity, some morphological features and the key to the main morphospecies discrimination. *Protistology*. 2010/2011;6:227–35.
- Fokin SI, Przybos' E, Chivilev SM, Beier CL, Horn M, Skotarczak B, Wodecka B, Fujishima M. Morphological and molecular investigations of *Paramecium schewiakoffi* sp. nov. (Ciliophora, Oligohymenophorea) and current status of distribution and taxonomy of *Paramecium* spp. *Eur J Protistol*. 2004;40:225–43. <https://doi.org/10.1016/j.ejop.2004.02.001>.
- Wichterman R. *The Biology of Paramecium*. 2nd ed. New York: Springer; 1989.
- Chatton E, Brachon S. Sur une paramécie à deux races: *Paramecium duboscqui*, n. sp. *CR Biol*. 1933;14:988–91.
- Vivier E. Morphology, taxonomy, and general biology of the genus *Paramecium*. In: Van Wagtenonk WS, editor. *Paramecium: A Current Survey*. New York: Elsevier; 1974. p. 1–89.
- Shi X, Jin M, Liu G. Rediscovery of *Paramecium duboscqui* Chatton and Brachon, 1933, and a description of its characteristics. *J Eukaryotic Microbiol*. 1997;44:134–41. <https://doi.org/10.1111/j.1550-7408.1997.tb05950.x>.

14. Boscaro V, Fokin SI, Verni F, Petroni G. Survey of *Paramecium duboscqui* using three markers and assessment of the molecular variability in the genus *Paramecium*. *Mol Phylogenet Evol.* 2012;65:1004–13. <https://doi.org/10.1016/j.ympev.2012.09.001>.
15. Fokin SI, Stoeck T, Schmidt H. *Paramecium duboscqui* Chatton, Brachon, 1933. Distribution, ecology and taxonomy. *Eur J Protistol.* 1999;35:161–7. [https://doi.org/10.1016/S0932-4739\(99\)80033-9](https://doi.org/10.1016/S0932-4739(99)80033-9).
16. Künzel A, Aharon Oren: Halophilic microorganisms and their environments. *Int Microbiol.* 2003;6:151–2. <https://doi.org/10.1007/s10123-003-0125-0>.
17. Jeilu O, Gessesse A, Simachew A, Johansson E, Alexandersson E. Prokaryotic and eukaryotic microbial diversity from three soda lakes in the East African Rift Valley determined by amplicon sequencing. *Front Microbiol.* 2022;13:999876. <https://doi.org/10.3389/fmicb.2022.999876>.
18. Leoni C, Volpicella M, Fosso B, Manzari C, Piancone E, Dileo MCG, Arcadi E, Yakimov M, Pesole G, Ceci LR. A differential metabarcoding approach to describe taxonomy profiles of bacteria and archaea in the saltern of Margherita di Savoia (Italy). *Microorganisms.* 2020;8:936. <https://doi.org/10.3390/microorganisms8060936>.
19. Entz GS. Die Fauna der kontinentalen Kochsalszwasser. *Math Naturwiss Ber Ungarn.* 1904;19:89–124.
20. Kirby H. Two protozoa from brine. *Trans Amer Microsc Soc.* 1932;51:8–15.
21. Nymysłowski B. Über unbekannt halophile Mikroorganismen aus dem Inneren des Salzbergwerkes Wieliczka. *Bull Int Acad Sci.* 1913;3:88–104.
22. Volcani BE. The microorganisms of the Dead Sea. In: Rehovoth, editor. *Papers Collected to Commemorate the 70th Anniversary of Dr. Chaim Weizmann.* Rehovot: Daniel Sieff Research Institute; 1944. p. 71–85.
23. Casamayor EO, Massana R, Benlloch S, Øvreås L, Díez B, Goddard VJ, Gasol JM, Joint I, Rodríguez-Valera F, Pedrós-Alió C. Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environ Microbiol.* 2002;4:338–48. <https://doi.org/10.1046/j.1462-2920.2002.00297.x>.
24. Clavero E, Hernandez-Marin eM, Grimalt J.O, Garcia-Pichel F. Salinity tolerance of diatoms from thalassic hypersaline environments. *J Phycol.* 2000;36:1021–34. <https://doi.org/10.1046/j.1529-8817.2000.99177.x>.
25. Elloumi J, Carrias JF, Ayadi H, Sime-Ngando T, Bouaïn A. Communities structure of the planktonic halophiles in the solar saltern of Sfax, Tunisia. *Estuar Coast Shelf S.* 2009;81:19–26. <https://doi.org/10.1016/j.ecss.2008.09.019>.
26. Estrada M, Henriksen P, Gasol JM, Casamayor EO, Pedrós-Alió C. Diversity of planktonic photoautotrophic microorganisms along a salinity gradient as depicted by microscopy, flow cytometry, pigment analysis and DNA-based methods. *FEMS Microbiol Ecol.* 2004;49:281–93. <https://doi.org/10.1016/j.femsec.2004.04.002>.
27. Lei Y, Xu K, Choi JK, Hong HP, Wickham SA. Community structure and seasonal dynamics of planktonic ciliates along salinity gradients. *Eur J Protistol.* 2009;45:305–19. <https://doi.org/10.1016/j.ejop.2009.05.002>.
28. Zhang B, Hou L, Qi H, Hou L, Zhang T, Zhao F, Miao M. An extremely streamlined macronuclear genome in the free-living protozoan *Fabrea salina*. *Mol Biol Evol.* 2022;39:msac062. <https://doi.org/10.1093/molbev/msac062>.
29. Kazmi SSUH, Tayyab M, Uroosa, Pastorino P, Barcelò D, Khan S, Yaseen ZM. Vertical variations and environmental heterogeneity drove the symphony of periphytic protozoan fauna in marine ecosystems. *Sci Total Environ.* 2024;932:173115. <https://doi.org/10.1016/j.scitotenv.2024.173115>.
30. Xu H, Zhang W, Jiang Y, Yang EJ. Use of biofilm-dwelling ciliate communities to determine environmental quality status of coastal waters. *Sci Total Environ.* 2014;470-471:511-8. <https://doi.org/10.1016/j.scitotenv.2013.10.025>.
31. Wang Q, Sun Z, Song S, Ali A, Xu H. Can salinity variability drive the colonization dynamics of periphytic protozoan fauna in marine environments? *Mar Pollut Bull.* 2024;198:115882. <https://doi.org/10.1016/j.marpolbul.2023.115882>.
32. Li C, Fu Y, Tian Y, Zang Z, Gentekaki E, Wang Z, Warren A, Li L. Comparative transcriptome and antioxidant biomarker response reveal molecular mechanisms to cope with zinc ion exposure in the unicellular eukaryote *Paramecium*. *J Hazard Mater.* 2023;453:131364. <https://doi.org/10.1016/j.jhazmat.2023.131364>.
33. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31:3210–12. <https://doi.org/10.1093/bioinformatics/btv351>.
34. Sallet E, Gouzy J, Schiex T. EuGene: An automated integrative gene finder for eukaryotes and prokaryotes. *Methods Mol Biol.* 2019;1962:97–120. [https://doi.org/10.1007/978-1-4939-9173-0\\_6](https://doi.org/10.1007/978-1-4939-9173-0_6).
35. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 2016;44:W54–7. <https://doi.org/10.1093/nar/gkw413>.
36. Sun P, Jiao B, Yang Y, Shan L, Li T, Li X, Xi Z, Wang X, Liu J. WGDl: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol Plant.* 2022;15:1841–51. <https://doi.org/10.1016/j.molp.2022.10.018>.
37. Plotkin JB, Kudla G. Synonymous but not the same: The causes and consequences of codon bias. *Nat Rev Genet.* 2011;12:32–42. <https://doi.org/10.1038/nrg2899>.
38. Anwar AM. BAWT: Automated tool for codon usage bias analysis for molecular evolution. *J Open Source Softw.* 2019;4:1500. <https://doi.org/10.21105/joss.01500>.
39. Wright F. The effective number of codons used in a gene. *Gene.* 1990;87:23–9.
40. Comeron JM, Aguadé M. An evaluation of measures of synonymous codon usage bias. *J Mol Evol.* 1998;47:268–74. <https://doi.org/10.1007/pl00006384>.
41. Sueoka N. Directional mutation pressure, mutator mutations, and dynamics of molecular evolution. *J Mol Evol.* 1993;37:137–53. <https://doi.org/10.1007/BF02407349>.
42. Sies H, Jones DP. Reactive oxygen species (ROS) as pleiotropic physiological signalling agents. *Nat Rev Mol Cell Biol.* 2020;21:363–83. <https://doi.org/10.1038/s41580-020-0230-3>.
43. Zhang W, Zhi W, Qiao H, Huang J, Li S, Lu Q, Wang N, Li Q, Zhou Q, Sun J, Bai Y, Zheng X, Bai M, Van Breusegem F, Xiang F. H<sub>2</sub>O<sub>2</sub>-dependent oxidation of the transcription factor GmN1L1 promotes salt tolerance in soybean. *Plant Cell.* 2023;36:112–35. <https://doi.org/10.1093/plcell/koad250>.
44. Zhang J, Liu D, Zhu D, Liu N, Yan Y. Endoplasmic reticulum subproteome analysis reveals underlying defense mechanisms of wheat seedling leaves under salt stress. *Int J Mol Sci.* 2021;22:4840. <https://doi.org/10.3390/ijms22094840>.
45. Saddhe AA, Mishra AK, Kumar K. Molecular insights into the role of plant transporters in salt stress response. *Physiol Plant.* 2021;173:1481–94. <https://doi.org/10.1002/jez.1402510303>.
46. Cronkite DL, Pierce SK. Free amino acids and cell volume regulation in the euryhaline ciliate *Paramecium calkinsi*. *J Exp Zool.* 1989;251:275–84.
47. Qin Y, Zhang B, Wang Y, Su R. Characterization of SEC14 family in wheat and the function of TaSEC14-7B in salt stress tolerance. *Plant Physiol Biochem.* 2023;202:107926. <https://doi.org/10.1016/j.plaphy.2023.107926>.
48. Walter P, Ron D. The unfolded protein response: from stress pathway to homeostatic regulation. *Science.* 2011;334:1081–6. <https://doi.org/10.1126/science.1209038>.
49. Li B, Niu F, Zeng Y, Tse MK, Deng C, Hong L, Gao S, Lo SW, Cao W, Huang S, Dagdas Y, Jiang L. Ufmylation reconciles salt stress-induced unfolded protein responses via ER-phagy in Arabidopsis. *PNAS.* 2023;120(5):e2208351120. <https://doi.org/10.1073/pnas.2208351120>.
50. S Narasimhan KK, Devarajan A, Karan G, Sundaram S, Wang Q, van Groen T, Monte FD, Rajasekaran NS. Reductive stress promotes protein aggregation and impairs neurogenesis. *Redox Biol.* 2020;37:101739. <https://doi.org/10.1016/j.redox.2020.101739>.
51. Cai C, Tu J, Najarro J, Zhang R, Fan H, Zhang FQ, Li J, Xie Z, Su R, Dong L, Arellano N, Ciboddo M, Elf SE, Gao X, Chen J, Wu R. NRAS mutant dictates AHCYL1-governed ER calcium homeostasis for melanoma tumor growth. *Mol Cancer Res.* 2024;22:386–401. <https://doi.org/10.1158/1541-7786.MCR-23-0445>.
52. Yu J, Li T, Liu Y, Wang X, Zhang J, Wang X, Shi G, Lou J, Wang L, Wang CC, Wang L. Phosphorylation switches protein disulfide isomerase activity to maintain proteostasis and attenuate ER stress. *EMBO J.* 2020;39:e103841. <https://doi.org/10.15252/emboj.2019103841>.
53. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics.* 2018;34:i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.

54. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011;27:764–70. <https://doi.org/10.1093/bioinformatics/btr011>.
55. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1432. <https://doi.org/10.1038/s41467-020-14998-3>.
56. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;27:722–36. <https://doi.org/10.1101/gr.215087.116>.
57. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–100. <https://doi.org/10.1093/bioinformatics/bty191>.
58. Langdon WB. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. *BioData Min*. 2015;8:1. <https://doi.org/10.1186/s13040-014-0034-0>.
59. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37:907–15. <https://doi.org/10.1038/s41587-019-0201-4>.
60. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *PNAS*. 2020;117:9451–7. <https://doi.org/10.1073/pnas.1921046117>.
61. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Muceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52. <https://doi.org/10.1038/nbt.1883>.
62. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res*. 2013;41:D226–32. <https://doi.org/10.1093/nar/gks1005>.
63. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:238. <https://doi.org/10.1186/s13059-019-1832-y>.
64. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3. <https://doi.org/10.1093/bioinformatics/btp348>.
65. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
66. Lartillot N, Rodrigue N, Stubbs D, Richer J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol*. 2013;62:611–5. <https://doi.org/10.1093/sysbio/syt022>.
67. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4. <https://doi.org/10.1093/molbev/msw054>.
68. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*. 2003;19:301–2. <https://doi.org/10.1093/bioinformatics/19.2.301>.
69. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*. 2006;22:1269–71. <https://doi.org/10.1093/bioinformatics/btl097>.
70. Shulgina Y, Eddy SR. Codetta: predicting the genetic code from nucleotide sequence. *Bioinformatics*. 2023;39:btac802. <https://doi.org/10.1093/bioinformatics/btac802>.
71. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30. <https://doi.org/10.1093/bioinformatics/btt656>.
72. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
73. Arnaiz O, Meyer E, Sperling L. ParameciumDB 2019: integrating genomic data across the genus for functional and evolutionary biology. *Nucleic Acids Res*. 2020;48:D599–605. <https://doi.org/10.1093/nar/gkz948>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.