

Genome-Wide Identification of Chitinase Genes in *Thalassiosira Pseudonana* and Analysis of Their Expression Under Abiotic Stresses

Haomiao Cheng (✉ chenghaomiao1997@163.com)

Institute of Oceanology Chinese Academy of Sciences <https://orcid.org/0000-0003-0906-7431>

Zhanru Shao

Institute of Oceanology Chinese Academy of Sciences

Chang Lu

Institute of Oceanology Chinese Academy of Sciences

Delin Duan

Institute of Oceanology Chinese Academy of Sciences

Research article

Keywords: Abiotic stress, Chitinase, Enzymatic activity, Gene expression, Gene family, *Thalassiosira pseudonana*

Posted Date: November 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-105434/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Plant Biology on February 10th, 2021. See the published version at <https://doi.org/10.1186/s12870-021-02849-2>.

Abstract

Background: The nitrogen-containing polysaccharide chitin is the second most abundant biopolymer on earth and is found in the cell walls of diatoms, where it serves as a scaffold for biosilica deposition. Diatom chitin is an important source of carbon and nitrogen in the marine environment, but surprisingly little is known about basic chitinase metabolism in diatoms.

Results: Here, we identify and fully characterize 24 chitinase genes from the model centric diatom *Thalassiosira pseudonana*. We demonstrate that their expression is broadly upregulated under abiotic stresses, despite the fact that chitinase activity itself remains unchanged, and we discuss several explanations for this result. We also examine the potential transcriptional complexity of the intron-rich *T. pseudonana* chitinase genes and provide evidence for two separate tandem duplication events during their evolution.

Conclusions: Given the many applications of chitin and chitin derivatives in suture production, wound healing, drug delivery, and other processes, new insight into diatom chitin metabolism has both theoretical and practical value.

1. Introduction

Chitin, β -1,4-linked *N*-acetyl-D-glucosamine, is the second most abundant natural polymer on earth after cellulose. Chitin is found in insect and crustacean exoskeletons, mollusc endoskeletons, and the cell walls of fungi and diatoms [1]. Chitin and its derivative, chitosan, are used in suture production, wound healing, drug delivery, vehicles, scaffolds, and other applications [2, 3]. Chitin has three main crystal structures: α -, β -, and γ -chitin [4]. Among these, β -chitin displays a parallel polysaccharide structure that confers greater reactivity, swelling, and solubility compared with the other two chitin types [5, 6]. β -chitin is only found in certain marine organisms, including molluscs [6, 7], tubeworms [8], and diatoms [9].

Diatoms are dominant primary producers that account for 40% of primary production in the oceans and perform 20% of annual global carbon fixation [10]. Chitin is a fundamental structure within the cell walls of diatoms and serves as a supporting scaffold for biosilica deposition [11]. Flexible chitin spines have been proposed to control cell buoyancy or sinking [12], and chitin can form a network around diatom cells that promotes the adhesion of bacteria such as *Vibrio parahaemolyticus* [13]. As a nitrogenous polysaccharide, chitin is an important source of carbon and nitrogen in the marine environment. It has been suggested that these two elements would quickly disappear from the ocean if chitin stopped circulating [14, 15]. To date, chitin metabolism has been investigated in two diatom genera, *Thalassiosira* and *Cyclotella*, due to the completion of their genome sequencing [16–18].

As known β -chitin producers, centric diatoms are presumed to contain special enzymes that catalyze chitin biosynthesis and degradation. Chitinases (EC 3.2.1.14) are one of the largest groups of hydrolases and decompose chitin into *N*-acetylglucosamines. Chitinase and its derivatives have been broadly applied in agriculture, pathogenicity, allergenicity, and health [19–21]. Chitinases have been found in diverse

organisms from different kingdoms of life, ranging from bacteria and higher plants to animals [22]. They are classified into glycosyl hydrolase families 18 and 19 (GH18 & GH19) based on the sequence homology of their catalytic domains [23]. GH18 chitinases are widely distributed among organisms such as viruses, bacteria, fungi, yeast, plants, and animals, whereas GH19 chitinases are found almost exclusively in plants [24]. The two chitinase families possess distinct sequence features and different 3D-structures, suggesting that they have descended from a different ancestor [25]. Liu et al. (2018) described the sequence homology, domain architectures, and gene expression patterns of chitinases in the oriental fruit fly and proposed their putative physiological functions [26]. Chen et al. (2018) characterized the chitinase gene family in *Brassica rapa* and examined its role in clubroot resistance [27]. Bartholomew et al. (2019) identified cucumber chitinase genes and characterized their evolution and expression responses to *Fusarium oxysporum* [28]. Previous work has focused on chitinase functions in nutrition, morphogenesis, and defense against pathogens in bacteria, fungi, insects, plants, animals, and other organisms [29], but few studies have examined chitinases in diatoms.

Here, we investigated the chitinase gene family in the model *Thalassiosira* species *Thalassiosira pseudonana*, making use of its completely sequenced genome [16]. We identified 24 chitinase genes and analyzed their sequence structural features, scaffold locations, phylogenetic relationships, stress-related *cis*-acting elements, subcellular locations, and responses to abiotic stresses at the transcriptional and enzymatic levels. This is the first study to investigate the characteristics of chitinase family members in diatoms. Our results provide insight into β -chitin degradation in diatoms and can support *in vitro* applications of chitinases derived from diatoms.

2. Materials And Methods

2.1 *T. pseudonana* cultivation

T. pseudonana (CCMP 1335) was obtained from the Microalgae Collection Center at Ningbo University (Ningbo, Zhejiang, China). Cells were grown in optimized f/2 liquid medium supplied by Shanghai Guangyu Biological Technology Co., Ltd. Cultures were grown at 19 °C with a 12 h:12 h light: dark diurnal cycle ($100 \mu\text{mol m}^{-2} \text{s}^{-1}$) and shaken at 100 rpm.

When cells grown in the starter culture had reached the exponential phase, 50 mL of cell culture were centrifuged under sterile conditions at 2,850 *g* for 15 min (Beckman Coulter, Allegra X-15R Centrifuge), washed once with medium, and inoculated into 50 mL of fresh medium. For temperature-limited treatment, cells were placed under a low temperature of 12 °C with $100 \mu\text{mol m}^{-2} \text{s}^{-1}$ of light, whereas for the light-limited treatment, cells were exposed to a low irradiance of $50 \mu\text{mol m}^{-2} \text{s}^{-1}$ at 19 °C. For the silica-limited treatment, cells were washed once with medium and inoculated into 50 mL of fresh, silica-free medium and were then cultured under normal light ($100 \mu\text{mol m}^{-2} \text{s}^{-1}$) and temperature (19 °C) conditions with the control group. The experiment was performed in flasks for 48 h, and there were three biological replicates per treatment.

2.2 Identification of chitinase family genes in *T. pseudonana*

The nucleotide sequences and corresponding protein sequences of *T. pseudonana* chitinase genes were retrieved from the Joint Genomics Institute PhycoCosm database (JGI PhycoCosm, <https://jgi.doe.gov/data-and-tools/phycocosm/>) using "chitinase" and other relevant keywords listed in Table S1. A total of 93 hits were annotated as chitinase-relevant genes. From these, we removed 41 genes that were redundantly annotated as full-length sequences but were in fact partial sequences of the other 52 genes (Table S2). Protein sequences of the remaining 52 genes were submitted to the SMART and Pfam websites to confirm the presence of chitinase domains Glyco_hydro_18 (PF00704) and Glyco_hydro_19 (PF00182) with a cut-off *E*-value of < 0.0001 [30–32]. Proteins with one or both of the Glyco_hydro_18 and Glyco_hydro_19 domains were regarded as putative chitinases.

2.3 Sequence analysis and structural characterization

All of the nucleotide sequences were analyzed using the ExPASy ProtParam tool to calculate their amino acid numbers, molecular weights, and theoretical isoelectric points (pIs) [33]. Genomic locations, intron numbers, and gene structure information were acquired from JGI PhycoCosm. MEME V5.1.1 online software was used to identify conserved motifs in the chitinase proteins with the following parameters: any number of repetitions, maximum of 20 motifs, and an optimum motif width of 6–200 amino acids [34]. Signal peptides and transmembrane helices (TMHs) were predicted using SignalP-5.0 (<http://www.cbs.dtu.dk/services/SignalP/>) and TMHMM Server v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>), respectively. Subcellular locations were predicted by combining the results of WoLF PSORT (<https://wolfpsort.hgc.jp/>) with those of HECTAR and ASAFind, the two software that are specific for heterokonts [35–37].

Secondary structures of *T. pseudonana* chitinase protein sequences were predicted using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) with default parameters. For each phylogenetic group of *T. pseudonana* chitinases that contained at least two members, protein sequences were aligned with ClustalW (<https://www.genome.jp/tools-bin/clustalw>), and representative secondary structures were then visualized with ESPript3.0 [38]. Three dimensional (3D) models of the *T. pseudonana* chitinases were constructed using the SWISS-MODEL server [39–43].

2.4 Scaffold localization and gene duplication identification

TBtools software was used to perform scaffold localization based on retrieved genomic location information [44]. Gene duplication events within the family were identified with BLASTp and MCScanX methods [45]. All *T. pseudonana* chitinase protein sequences were aligned with each other using BLASTp. The BLASTp result was used as input for MCScanX, and duplication events and gene collinearity were identified using default parameters.

2.5 Phylogenetic analysis and classification

To study the evolutionary relationships among chitinases from *T. pseudonana*, *T. Thalassiosira oceanica*, and *Thalassiosira weissflogii*, an unrooted neighbor-joining phylogenetic tree was constructed using MEGA X software (v10.1.8) with 1000 bootstrap replicates [46]. The *T. oceanica* protein sequences were identified and downloaded from the JGI PhycoCosm database using the keyword "chitinase". To acquire *T. weissflogii* protein sequences, the 24 *T. pseudonana* chitinase protein sequences were used as BLASTp queries against the *T. weissflogii* transcriptomes released by the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) [47]. The *T. oceanica* and *T. weissflogii* candidate sequences were submitted to SMART and Pfam to verify the presence of one or both of the Glyco_hydro_18 and Glyco_hydro_19 domains [30–32]. In total, 22 full-length *T. oceanica* chitinases (Table S3), 76 *T. weissflogii* chitinases (Table S4), and the 24 newly identified chitinases from *T. pseudonana* were used for phylogenetic analysis.

2.6 Prediction of *cis*-acting elements

To predict the *cis*-acting regulatory elements in the promoter regions of *T. pseudonana* chitinase genes, 1.5 kb of sequence upstream from each initial codon (ATG) was downloaded from the JGI PhycoCosm database. Stress response-related *cis*-acting elements in the promoter sequences of the chitinase genes were then investigated using the PlantCARE website (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) [48].

2.7 Enzyme assays and transcriptional profile analysis

After 48 h of stress treatment, a 1 mL sample was taken from each flask for measurement of chitinase activity using the Solarbio Chitinase Assay Kit (Cat#BC0820) following the manufacturer's instructions. The remaining 49 mL of *T. pseudonana* cells in each flask were collected by centrifugation at 2,850 *g*, washed once with fresh medium, and flash-frozen in liquid nitrogen. The *T. pseudonana* cell pellets were homogenized by vortexing in TRIzol reagent (Invitrogen, Waltham, MA, USA), then centrifuged at 9,700 *g*. The supernatants were used for RNA extraction with chloroform and isopropanol. RNA pellets were resuspended in diethyl pyrocarbonate-treated water, followed by elimination of genomic DNA and synthesis of complementary DNA (cDNA) with a PrimeScript RT reagent kit (Takara, Japan). The synthesized cDNA was then used for quantitative real-time PCR (qRT-PCR) experiments.

The transcriptional profiles of 14 randomly selected *T. pseudonana* chitinase genes were obtained by qRT-PCR using TB Green™ Premix Ex Taq™ II (Takara, Japan) and a Thermal Cycle Dice™ Real Time System (Takara, Japan). Gene-specific qRT-PCR primers were designed using the NCBI Primer-BLAST website [49]. Information on primers is provided in Table S5. The beta tubulin (TUB3) gene was used as the reference gene for normalizing the expression of the *T. pseudonana* chitinase genes [50].

3. Results

3.1 Genome-wide identification and analysis of chitinase genes in *T. pseudonana*

A total of 24 non-redundant chitinase genes were identified in the genome of *T. pseudonana* and designated *TpChi1–24* based on their chromosomal locations. Gene names, gene IDs, chromosomal locations, exon numbers, amino acid sequence lengths, molecular weights, and pIs are presented in Table 1. The sequence lengths of *TpChi* proteins ranged from 417 (*TpChi1*) to 3512 (*TpChi6*) amino acid (aa) residues, with an average length of 1,517 aa. The relative molecular mass varied from 46.4 kDa (*TpChi1*) to 373.1 kDa (*TpChi6*), and the pI values ranged from 4.09 (*TpChi6*) to 5.42 (*TpChi16*).

Table S6 lists predicted signal peptides, TMHs, and subcellular localizations for *T. pseudonana* chitinase proteins. HECTAR predictions indicated that up to half of the proteins (12 out of 24) possessed signal peptides. Nine (*TpChi9*, *TpChi13*, *TpChi14*, *TpChi17*, *TpChi18*, *TpChi20*, *TpChi21*, *TpChi22*, and *TpChi23*) were predicted to contain signal peptides by all three programs (SignalP-5.0, HECTAR, and ASAFIND), suggesting that these chitinases are involved in the secretory pathway. Twelve chitinases (*TpChi1*, *TpChi2*, *TpChi3*, *TpChi4*, *TpChi5*, *TpChi7*, *TpChi8*, *TpChi10*, *TpChi11*, *TpChi15*, *TpChi16*, and *TpChi19*) were predicted by HECTAR to have type Π signal anchors, indicating that they may be Type Π transmembrane proteins with an N-in orientation and maybe localized on chloroplast or mitochondrial membranes. Six proteins (*TpChi2*, *TpChi5*, *TpChi9*, *TpChi14*, *TpChi22*, *TpChi23*) were predicted to contain TMHs, suggesting a subcellular localization at the plasma membrane or the endomembrane system.

Table 1
Features of the identified *T. pseudonana* chitinase genes

Name	JGI gene ID	Scaffold	Genomic location	Exon	AA	MW(KDa)	pI
<i>TpChi1</i>	21369	2	chr_2:1,497,142-1,498,555 (+)	2	417	46.40	4.86
<i>TpChi2</i>	21501	2	chr_2:2,082,322-2,089,884 (-)	16	2070	232.39	5.13
<i>TpChi3</i>	21626	3	chr_3:306,112–310,877 (-)	4	1464	159.27	4.64
<i>TpChi4</i>	21917	3	chr_3:1,759,345-1,764,662 (-)	7	1498	165.71	5.21
<i>TpChi5</i>	22088	4	chr_4:199,894 - 204,931 (-)	17	1087	118.17	4.51
<i>TpChi6</i>	22238	4	chr_4:960,301–971,814 (+)	8	3512	373.14	4.09
<i>TpChi7</i>	22376	4	chr_4:1,635,653-1,638,316 (+)	5	684	74.97	4.66
<i>TpChi8</i>	5406	5	chr_5:1,117,857-1,121,711 (+)	14	901	96.61	4.75
<i>TpChi9</i>	22909	5	chr_5:2,040,965-2,049,241 (+)	9	2494	273.21	4.44
<i>TpChi10</i>	22915	5	chr_5:2,100,817-2,105,674 (-)	13	785	87.47	4.61
<i>TpChi11</i>	6636	7	chr_7:81,934 - 86,292 (+)	5	1349	146.67	4.47
<i>TpChi12</i>	23384	7	chr_7:448,791 - 455,002 (-)	8	1819	196.99	4.49
<i>TpChi13</i>	23470	7	chr_7:953,351–956,745 (+)	3	1031	109.60	4.28
<i>TpChi14</i>	23629	7	chr_7:1,845,016 - 1,855,061 (-)	15	2773	304.08	4.31
<i>TpChi15</i>	23700	8	chr_8:274,001-278,110 (-)	7	1116	119.55	4.44
<i>TpChi16</i>	8279	10	chr_10:177,498 - 181,104 (+)	10	911	101.63	5.42
<i>TpChi17</i>	24326	11a	chr_11a:339,905 - 344,411 (-)	7	1113	118.52	5.08
<i>TpChi18</i>	24421	12	chr_12:128,580 - 139,130 (-)	6	3325	359.94	4.25
<i>TpChi19</i>	11110	19a_19	chr_19a_19:163,538 - 171,325 (-)	8	2241	241.48	4.37
<i>TpChi20</i>	11125	19a_19	chr_19a_19:210,121–212,567 (-)	8	656	71.64	4.46

Name	JGI gene ID	Scaffold	Genomic location	Exon	AA	MW(KDa)	pI
<i>TpChi21</i>	11236	19a_19	chr_19a_19:528,078–531,770 (+)	4	1122	121.35	4.59
<i>TpChi22</i>	11237	19a_19	chr_19a_19:532,139–535,390 (+)	8	877	94.94	4.35
<i>TpChi23</i>	11963	22	chr_22:1,024,768-1,030,124 (-)	13	1438	156.88	4.31
<i>TpChi24</i>	25849	23	chr_23:244567–250756 (+)	9	1735	187.00	4.66

3.2 Gene structures, motifs, and conserved domains of the *TpChis*

All chitinase genes contained at least two exons and one intron, and *TpChi5* contained 17 exons, the most of all the chitinase genes. One quarter of the *TpChis* had more than ten introns (6, 25%), over one half had seven or more introns (14, 58%), and only *TpChi1* had a single intron (Fig. 1).

The sequence differences within the chitinase family were further analyzed by examining their conserved domains and motifs using the Pfam database and the MEME program, respectively. Twenty conserved motifs were identified in the *T. pseudonana* chitinase proteins. Their lengths varied from 15 to 65 aa, and their details are presented in Table S7. The Glyco_hydro_18 domain was identified in nineteen chitinases and the Glyco_hydro_19 domain in six. Notably, *TpChi12* was found to possess both Glyco_hydro_18 and Glyco_hydro_19 domains. Three other auxiliary domains, Chitin_bind_1 (PF00187), CBM_14 (Peritrophin A, PF01607), and LPMO_10 (PF03067), were detected in 10 sequences (41.7%). In addition, three motif combinations corresponding to the two Glyco_hydro domains were identified. The combination of motifs 19, 11, 17, 13, and 10 was detected in the Glyco_hydro_19 domain, which was identified in *TpChi1*, *TpChi12*, *TpChi13*, *TpChi15*, *TpChi19*, and *TpChi20*. Two other highly conserved motif combinations were identified in the Glyco_hydro_18 domain. *TpChi3*, *TpChi6*, *TpChi14*, *TpChi18*, *TpChi21*, *TpChi22*, and *TpChi24* contained the combination of motifs 18, 16, 1, 7, 3, 12, and 5, whereas *TpChi2*, *TpChi4*, *TpChi5*, *TpChi7*, *TpChi8*, *TpChi9*, *TpChi10*, *TpChi11*, *TpChi12*, *TpChi16*, and *TpChi23* contained the combination of motifs 2, 8, 1, 7, 15, 3, 4, 6, 3, and 9. *TpChi17* was an exception, containing only motifs 1 and 14. In general, chitinases in the same group had similar motif compositions (Fig. 1).

3.3 Phylogenetic analysis of representative *Thalassiosira* chitinases

An unrooted phylogenetic tree was constructed to reveal the evolutionary relationships among the chitinases of *T. pseudonana*, *T. oceanica* (22, Table S3) and *T. weissflogii* (76, Table S4). A total of 122 protein sequences from the three *Thalassiosira* species were used to build the phylogenetic tree (Fig. 2). The *T. pseudonana* chitinases were classified into eight groups (I–VIII) based on the topology of the

phylogenetic tree and the domain architectures of the protein sequences. Group I was the largest and contained 6 members from *T. pseudonana*, 5 from *T. oceanica*, and 12 from *T. weissflogii*. A total of 15 domains were detected in the 122 chitinases from all three *Thalassiosira* species (Fig. 2). Groups II–IV all contained chitinases with the Glyco_hydro_18 catalytic domain, and the chitinases in Group I (except for TwChi50) contained only the Glyco_hydro_19 domain. Some chitinases in Group I (*TwChi9*, *TwChi24*, *TwChi58*, and *TwChi65*) contained two Glyco_hydro_18 domains. Interestingly, seven chitinases (*TpChi12*, *TpChi11*, *TwChi29*, *TwChi31*, *TwChi52*, *TwChi59*, and *TwChi74*) in Group I and *TwChi50* in Group II contained both the Glyco_hydro_18 and the Glyco_hydro_19 catalytic domains. Overall, chitinases in the same group had similar domain architectures.

3.4 Scaffold locations and gene duplication of the *TpChis*

The 24 *TpChis* were assigned to 12 scaffolds (Fig. 3 and Table 1), and the distribution of chitinase genes on each scaffold was uneven. Chr7 and Chr19a_19 contained the largest number of chitinase genes (4), whereas only a single chitinase gene was located on Chr8, Chr10, Chr11a, Chr12, Chr22, and Chr23. No segmental duplications were observed among the chitinase gene family members. However, there were two tandemly duplicated gene pairs: *TpChi11* and *TpChi12*, and *TpChi21* and *TpChi22*. These were located on Chr7 and Chr19a_19, respectively, and accounted for 16.7% of the chitinase genes. No collinearity among chitinase gene family members was observed with MCScanX.

3.5 Stress-related *cis*-acting elements in the *T. pseudonana* chitinase gene promoters

To investigate possible regulation mechanisms of the *TpChis* under abiotic or biotic stress, 1.5-kb upstream sequences (promoter regions) were analyzed using the PlantCARE database to identify *cis*-regulatory elements. Twenty-six types of 259 stress-related *cis*-acting elements were identified, including light response-related (G-box and 21 other types), low-temperature responsiveness-related (LTR), drought-inducibility-involved (MBS), and two other *cis*-acting elements involved in defence and stress responsiveness (i.e., the CCAAT-box and TC-rich repeats) (Fig. 4). All the chitinase genes contained 2–9 stress-related *cis*-acting elements in their promoter regions. The G-box was found in all *TpChi* promoters and accounted for almost half of the identified elements (122). In total, 22 types of 199 light-responsive *cis*-acting elements accounted for 76.8% of the stress-related *cis*-acting elements.

3.6 Secondary and 3D structures of the *T. pseudonana* chitinases

Four secondary structures (α -helix, extended strand, β -turn, and random coil) were found in the *T. pseudonana* chitinases (Table S8). The α -helix and random coil were the dominant components of chitinase secondary structure and accounted for 24.79% and 47.46% on average. Extended strand constituted 19.63% of the secondary structures and β -turn constituted 8.12%. Figure 5 shows representative secondary structures for each group. 3D molecular modelling provides dynamic

information that is usually difficult to acquire from experiments, and 3D protein models were constructed for the *T. pseudonana* chitinases to better understand their structural properties (Table S9). Most top-hit models (75%) were chitin-degrading enzymes from a wide range of phylogenetic origins, including viruses (*TpChi5*), bacteria (*TpChi2*, *TpChi7*, *TpChi8*, *TpChi10*, *TpChi11*, *TpChi16*, *TpChi17*), yeast (*TpChi20*), fungi (*TpChi12*, *TpChi23*), rice (*TpChi1*, *TpChi19*), insects (*TpChi14*, *TpChi18*), and humans (*TpChi3*, *TpChi21*, *TpChi24*). *T. pseudonana* chitinases in the same phylogenetic groups (Fig. 2) were modelled with similar 3D structures (Table S9).

3.7 Expression analysis of *TpChis* under abiotic stresses

To investigate individual gene responses to low irradiance ($50 \mu\text{mol m}^{-2}\text{s}^{-1}$), low temperature (12 °C), and silica deficiency stress at the transcriptional level, 14 genes were randomly selected for qRT-PCR analysis. Generally, *TpChis* were upregulated after 48 h of exposure to the stresses (Fig. 6). Silica depletion produced the greatest increases in transcript abundance for 12 of the 14 genes (85.7%) upregulated under silica limitation. The expression levels of 9 (64.3%) and 11 (78.6%) genes were enhanced under light-limited and temperature-limited stresses, respectively. Although most genes were upregulated under stress, the most remarkably upregulated gene was *TpChi4*, whose relative expression level increased 14-, 5-, and 5-fold under light-limited, temperature-limited, and silica-limited conditions compared with the control. By contrast, only one chitinase gene (*TpChi17*) was down regulated overall, decreasing to one-tenth of the control expression level under temperature-limited condition.

To further investigate the involvement of the chitinase gene family in abiotic stress responses at the enzymatic level, we measured chitinase activity of *T. pseudonana* under conditions of low irradiance, low temperature, and silica depletion. The activity of the *T. pseudonana* chitinases remained steady under low irradiance, low temperature, and silica deficiency ($p > 0.05$) (Fig. 6).

4. Discussion

Chitin is the predominant polysaccharide in the diatom cell wall, and its various functions permit diatoms to adjust to the changeable ocean environment. Chitinase, the enzyme that degrades chitin into more valuable derivatives with shorter sugar chains and lower molecular weights, therefore deserves more attention in diatoms. Although the chitinase gene family has been widely investigated in insects, higher plants, and fungi [26, 27, 51, 52], there is little information about this gene family in Stramenopiles, and a systematic investigation has yet to be undertaken. The completeness of the *T. pseudonana* genome sequence enables the identification of putative diatom chitinase genes [16]. In the present study, *T. pseudonana* chitinase genes were identified from genome sequences, and their phylogenetic relationships within *Thalassiosira*, gene structures, conserved motifs and domains, scaffold locations, duplication events, stress-related *cis*-acting elements, and transcriptional expression profiles and enzymatic activities under abiotic stress were investigated in an integrated manner.

Diversity in structure and intron abundance may generate chitinases with various functions in *T. pseudonana*

In total, 24 chitinase genes were retrieved from the *T. pseudonana* genome and verified based on the identification of conserved chitinase domains. Similar numbers of chitinase genes have been reported previously in *T. pseudonana* (more than 20 chitinase-encoding genes) [53] and *Cyclotella cryptica* (22 chitinases) [17]. Members of the *T. pseudonana* chitinase family displayed different secondary and 3D structures (Tables S8, S9), suggesting that they may function in multiple biological processes, such as stress-related responses, growth, and developmental operations. Intron gain and loss amplifies the complexity of gene organization [54], and we found diverse exon-intron organization in the *T. pseudonana* chitinase genes (Fig. 1) that may have contributed to their structural diversity during evolution. Importantly, the *TpChis* generally possessed a large number of introns. Twenty-five percent had more than 10 introns, over half had 7 or more introns (14, 58%), and members of GroupV had the most abundant introns (Fig. 1). When abundant introns are embedded in gene sequences, alternative splicing (AS) is likely to occur during gene transcription. In eukaryotes, AS spatially and temporally regulates gene expression by selectively preserving or removing some exons, thereby producing multiple mRNAs [55]. The many possible AS variants of the *TpChis* may therefore enlarge the diversity of the *T. pseudonana* chitinase transcriptome and proteome.

Tandem duplication and genetic diversification may co-affect the evolution of the *TpChi* gene family

T. pseudonana chitinase genes were classified into eight groups (I–VIII) based on a phylogenetic tree that also included proteins from two additional *Thalassiosira* species. Chitinases in each group displayed similar domain architectures, motif compositions (Fig. 2), and conserved secondary (Fig. 5) and 3D structure (Table. S9), suggesting that they may have a closer evolutionary history and similar cellular functions.

Over the course of evolution, segmental and tandem duplications have propelled the expansion of gene families [56]. Here, we found two tandemly duplicated pairs of *T. pseudonana* chitinase genes (*TpChi11* and *TpChi12*, *TpChi21* and *TpChi22*) and no segmental duplication events, showing that tandem duplication has played a more important role in the expansion of the *T. pseudonana* chitinase gene family. *TpChi21* and *TpChi22* were located close together on Chr19a_19; they were classified into GroupVIII (Figs. 2 and Fig. 3), and exhibited conserved domains and motif compositions (Fig. 1), as well as similar secondary and 3D structures (Fig. 5 and Table. S9). The other pair, *TpChi11* and *TpChi12*, were distantly located on Chr_7 (362 kb) and were classified into different phylogenetic groups. *TpChi11* and *TpChi12* contained a common Glyco_hydro_18 domain. However, *TpChi12* also possessed a Glyco_hydro_19 domain and a longer sequence length. Genetic and functional divergence of duplicated genes are known to occur during evolution [57]. *TpChi12* was identified as a tandem duplicate of *TpChi11*, although it had an additional chitin-degrading domain. These observations suggest either the acquisition of a Glyco_hydro_19 domain by *TpChi12* or its loss in *TpChi11*. Furthermore, the exon-intron structures and the gene expression profiles under silica limitation differed between *TpChi11* and *TpChi12*, indicating that

this gene pair has undergone functional divergence. Based on these results, we speculate that the chitinase gene family may have undergone duplication first, then experienced genetic diversification that drove further functional diversification.

Several subcellular locations are predicted for TpChi proteins

Chitin fibrils are located between the silica frustule and the cell membrane of chitin-producing diatoms [58], but the locations of chitin degradation sites are unknown. We therefore investigated the subcellular locations of the *T. pseudonana* chitinases. HECTAR predictions indicated that half of the chitinase family members are involved in the secretory pathway, consistent with the finding that a majority of chitinase genes in *B. rapa* and *Arabidopsis thaliana* are secretory [24, 27]. This suggests that chitinases have extracellular activity and important *in vitro* applications. The remaining chitinases, however, were predicted to reside on the chloroplast or mitochondrial membranes. Given the secondary endosymbiosis exhibited by diatoms [59], *T. pseudonana* chitinases anchored to the chloroplast or mitochondrial membranes may be encoded by genes originally acquired from red algae or non-photosynthetic exosymbionts.

Different chitinase expression patterns may support chitin metabolism and other physiological needs of *T. pseudonana* in changeable environments

Twenty-six types of stress-related *cis*-regulatory elements were found in the promoter regions of the 24 *TpChi* genes (Fig. 4), indicating that they may undergo complex transcriptional regulation. A total of 199 light-responsive *cis*-acting elements were unevenly distributed in the *TpChi* gene promoters, suggesting that the expression of most chitinase genes was indispensably induced by light. Consistent with this observation, LTR, TC-rich repeat, and MBS elements are also important for the induction of chitinase genes in cucumber, *B. rapa*, *B. juncea*, and *Camelina sativa* [27, 28, 51].

Multiple studies have demonstrated variations in the transcription of chitinase family genes during biotic disease resistance in plants [27, 28, 51]. Cao & Tan (2019) investigated the expression of tomato chitinase genes in response to the pathogen *Sclerotinia sclerotiorum* and under low temperature, high temperature, drought, and salt stresses [60]. They observed that the expression of chitinase genes was broadly elevated under these stress conditions. The present study demonstrated a similar pattern of overall upregulation of *TpChi* genes in response to low irradiance, low temperature, and silica depletion. The different transcriptional profiles of *T. pseudonana* chitinase genes under different abiotic stress conditions suggest that they have various response and regulation mechanisms and may be involved in a variety of physiological processes.

No significant change in *in vivo* chitinase activity was detected in response to abiotic stresses, despite the increase in chitinase gene transcripts. This can probably be ascribed to the restriction in enzymatic activity or a decrease in the translation of chitinase proteins in adverse environments. The inconsistency between gene transcription and enzyme activity may support the proposal that *T. pseudonana* chitinase genes are involved in other cellular processes apart from chitin metabolism, or it may represent a

compensatory adjustment of chitin metabolism in shifting environments. The inconsistent results also suggest the possibility of post-transcriptional regulation during chitin degradation in *T. pseudonana*. By contrast, the expression of seabuckthorn chitinase genes was elevated at both the transcriptional [61] and protein activity levels [62] under cold stress. Moreover, chitinase activity of cucumber and tobacco also increased after treatment with biotic or abiotic stress inducers [63]. The results presented here suggest that there may be functional differences in the regulation and activity of chitinases in diatoms compared to those of higher plants.

Durkin et al. (2009) discovered that silica deposition and chitin exposure on the *T. pseudonana* cell wall were inversely correlated [53]. They suggested that chitin biosynthesis is enhanced in stressed cells that are unable to precipitate silica. In addition, co-regulation of transcript quantity by silicic acid and iron depletion is reported to be associated with cell wall processes [64]. In the present study, chitinase transcripts accumulated while enzymatic activity remained steady under silica depletion. Therefore, stable chitinase activity and higher chitin accumulation may allow chitin to serve as a substitute cell wall material when silica cannot be deposited. Moreover, as silica is required for diatom cell growth, chitin metabolism may also be affected by the abnormal growth of cells, further affecting cycles of carbon and other nutrients in the ocean [65]. However, further investigation of other genes involved in chitin metabolism (e.g., chitin synthases and chitin deacetylases) is required.

Auxiliary domains may assist in chitin binding by the *T. pseudonana* chitinases

Most chitinases contain a catalytic domain and additional auxiliary domains such as carbohydrate-binding modules (CBMs) [66], which are classified into 87 families at the CAZy database (<http://www.cazy.org/>). The CBMs anchor the enzyme to the substrate and disrupt its crystalline structure, creating free chain ends and thereby enhancing the enzyme's catalytic activity towards insoluble substrates [66, 67]. Chitin binding has been demonstrated for a protein containing only two CBM_14 domains [68]. In the present study, the CBM_14, Chitin_bind_1, and LMPO_10 domains, which are typically found in chitin-binding proteins, were identified in the protein sequences of *TpChi3*, *TpChi5*, *TpChi8*, *TpChi11*, *TpChi16*, *TpChi17*, *TpChi21*, *TpChi22*, *TpChi23*, and *TpChi24* (41.7%), five of which possessed two or more CBM_14 domains (*TpChi3*, *TpChi21*, *TpChi22*, *TpChi23*, and *TpChi24*) (Fig. 1), suggesting chitin-binding ability of *T. pseudonana* chitinases. Moreover, seven chitinases (*TpChi3*, *TpChi11*, *TpChi16*, *TpChi21*, *TpChi22*, *TpChi23*, and *TpChi24*) that contained the CBM_14 domain were predicted to be extracellular (containing a signal peptide) or located in the chloroplast or mitochondrial membrane (Table S6). This finding suggests more subcellular location of diatom chitinases containing the CBM_14 domain than the space between the plasma membrane and the silica frustule of diatoms as Traller et al. (2016) discussed [17].

5. Conclusions

We identified 24 *TpChi* genes from the *T. pseudonana* genome and divided them into eight subgroups. The gene family was analyzed from the perspective of gene structure, phylogeny, scaffold location, gene

duplication, secondary and 3D structures, and expression at the transcriptional and enzymatic levels under abiotic stresses. In general, chitinase transcript levels were upregulated but enzymatic activity remained unchanged under abiotic stresses. These findings suggest that the *T. pseudonana* chitinase gene family is crucial for responses to changeable environments and plays an indispensable as yet uncharacterized role in cellular processes. Our report contributes to a better understanding of chitin metabolism in diatoms and paves the way for future functional characterization of the *TpChi* genes and the *in vitro* application of *T. pseudonana* chitinases.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The reference genome of *T. pseudonana* was released in PhycoCosm at the Joint Genomics Institute (JGI) PhycoCosm with the accession number of *Thalassiosira pseudonana* CCMP 1335 (<https://jgi.doe.gov/data-and-tools/phycocosm/>). Gene sequences and protein sequences of *T. pseudonana* chitinase were downloaded from JGI PhycoCosm (<https://jgi.doe.gov/data-and-tools/phycocosm/>). Chitinase protein sequences of *T. oceanica* were downloaded from JGI PhycoCosm (<https://jgi.doe.gov/data-and-tools/phycocosm/>). Chitinase protein sequences of *T. weissflogii* were downloaded from iMicrobe website (<https://www.imicrobe.us/>). All of the datasets supporting the results of this article are included within the article and its Additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work received financial support from the National Natural Science Foundation of China (41806175) and Youth project of Marine Biology and Biotechnology Laboratory in Pilot National Laboratory for Marine Science and Technology Project (YQ2018N006). These funders financially supported the instruments and reagents purchase for the research. Each of the funding bodies took part in designing the study, collecting, analysing and interpreting the data, and writing the manuscript.

Authors' contributions

HC, ZS and DD conceived and designed research. HC conducted experiments, analysed the data and wrote the manuscript. ZS and DD gave critical revision of the manuscript. All authors have read and approved the manuscript.

Acknowledgements

The authors would like to thank TopEdit (www.topeditsci.com) for its linguistic assistance during the preparation of this manuscript. We would also like to thank Guanjie Zhang from Xi'an Jiaotong University and BeiningXue from the Institute of Oceanology, CAS, for their statistical analysis assistance and encouragements.

References

1. Pareek N, Vivekanand V, Singh R. Chitin deacetylase: characteristic molecular features and functional aspects. In: Shukla P, Pletschke BI, editors. *Advances in enzyme biotechnology*. New Delhi, India: Springer; 2013. p. 125-36.
2. Shigemasa Y, Minami S. Applications of chitin and chitosan for biomaterials. *Biotechnol Genet Eng Rev*. 1996;13(1):383-420.
3. Yang TL. Chitin-based materials in tissue engineering: applications in soft tissue and epithelial organ. *Int J Mol Sci*. 2011;12(3):1936-63.
4. Rudall KM. The Chitin/Protein Complexes of Insect Cuticles. *Adv Insect Physiol*. 1963;1:257-313.
5. Lamarque GC, M; Viton, C; Domard, A. New route of deacetylation of alpha- and beta-chitins by means of freeze-pump out-thaw cycles. *Biomacromolecules*. 2005;6(3):1380-8.
6. Cuong HN, Minh NC, Van Hoa N, Trung TS. Preparation and characterization of high purity beta-chitin from squid pens (*Loligo chinensis*). *Int J Biol Macromol*. 2016;93(Pt A):442-7.
7. Dweltz NE. The Structure of β -chitin. *Biochim Biophys Acta*. 1961;51:283-74.
8. Gaill F, Shillito B, Lechaire J, Chanzy H, Goffinet G. The Chitin Secreting System from Deep-sea Hydrothermal Vent Worms. *Biol Cell*. 1992;76(2):201-4.
9. J M, Mcinnes A, Falk M. Studies of the chitin (chitin-poly-N-acetylglucosamine) fibres of diatom *Thalassiosira fluviatilis* hustedt: I. production and isolation of chitin fibres. *Can J Bot*. 1965;43(6):707-13.
10. Field C, Behrenfeld M, Randerson J, Falkowski P. Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science*. 1998;281(5374):237-40.
11. Brunner E, Richthammer P, Ehrlich H, Paasch S, Simon P, Ueberlein S, et al. Chitin-based organic networks: an integral part of cell wall biosilica in the diatom *Thalassiosira pseudonana*. *Angew Chem Int Ed Engl*. 2009;48(51):9724-7.
12. Walsby AE, Xypolyta A. The form resistance of chitan fibres attached to the cells of *Thalassiosira fluviatilis* Hustedt. *Eur J Phycol*. 1977;12(3):215-23.

13. Frischkorn KR, Stojanovski A, Paranjpye R. *Vibrio parahaemolyticus* type IV pili mediate interactions with diatom-derived chitin and point to an unexplored mechanism of environmental persistence. *Environ Microbiol.* 2013;15(5):1416-27.
14. Johnstone J. *Conditions of life in the sea.* Cambridge, United Kingdom: University Press; 1908. 332 p.
15. Souza CP, Almeida BC, Colwell RR, Rivera IN. The importance of chitin in the marine environment. *Mar Biotechnol (NY).* 2011;13(5):823-30.
16. Armbrust E, Berges J, Bowler C, Green B, Martinez D, Putnam N, et al. The Genome of the Diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism. *Science.* 2004;306(5693):79-86.
17. Traller JC, Cokus SJ, Lopez DA, Gaidarenko O, Smith SR, McCrow JP, et al. Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. *Biotechnol Biofuels.* 2016;9:258.
18. Shao Z, Thomas Y, Hembach L, Xing X, Duan D, Moerschbacher BM, et al. Comparative characterization of putative chitin deacetylases from *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* highlights the potential for distinct chitin-based metabolic processes in diatoms. *New Phytol.* 2019;221(4):1890-905.
19. Ortiz-Rodriguez T, de la Fuente-Salcido N, Bideshi DK, Salcedo-Hernandez R, Barboza-Corona JE. Generation of chitin-derived oligosaccharides toxic to pathogenic bacteria using ChiA74, an endochitinase native to *Bacillus thuringiensis*. *Lett Appl Microbiol.* 2010;51(2):184-90.
20. Nagpure A, Choudhary B, Gupta RK. Chitinases: in agriculture and human healthcare. *Crit Rev Biotechnol.* 2014;34(3):215-32.
21. Patel S, Goyal A. Chitin and chitinase: Role in pathogenicity, allergenicity and health. *Int J Biol Macromol.* 2017;97:331-8.
22. Adrangi S, Faramarzi MA. From bacteria to human: a journey into the world of chitinases. *Biotechnol Adv.* 2013;31(8):1786-95.
23. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J.* 1991;280(2):309–16.
24. Passarinho PA, de Vries SC. Arabidopsis Chitinases: a Genomic Survey. *The Arabidopsis Book.* 2002;2002(1):e0023.
25. Hamel F, Boivin R, Tremblay C, Bellemare G. Structural and Evolutionary Relationships Among Chitinases of Flowering Plants. *J Mol Evol.* 1997;44(6):614-24.
26. Liu SH, Li HF, Yang Y, Yang RL, Yang WJ, Jiang HB, et al. Genome-wide identification of chitinase and chitin deacetylase gene families in the oriental fruit fly, *Bactrocera dorsalis* (Hendel). *Comp Biochem Physiol Part D Genomics Proteomics.* 2018;27:13-22.
27. Chen J, Piao Y, Liu Y, Li X, Piao Z. Genome-wide identification and expression analysis of chitinase gene family in *Brassica rapa* reveals its role in clubroot resistance. *Plant Sci.* 2018;270:257-67.
28. Bartholomew ES, Black K, Feng Z, Liu W, Shan N, Zhang X, et al. Comprehensive Analysis of the Chitinase Gene Family in Cucumber (*Cucumis sativus* L.): From Gene Identification and Evolution to

- Expression in Response to *Fusarium oxysporum*. *Int J Mol Sci*. 2019;20(21):5309.
29. Adrangi S, Faramarzi MA, Shahverdi AR, Sepehrizadeh Z. Purification and characterization of two extracellular endochitinases from *Massilia timonae*. *Carbohydr Res*. 2010;345(3):402-7.
 30. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res*. 2015;43(Database issue):D257-60.
 31. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res*. 2018;46(D1):D493-D6.
 32. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47(D1):D427-D32.
 33. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press; 2005. p. 571-607.
 34. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37(Web Server issue):W202-8.
 35. Gschloessl B, Guermeur Y, Cock JM. HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinf*. 2008;9:393.
 36. Gruber A, Rocap G, Kroth PG, Armbrust EV, Mock T. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *Plant J*. 2015;81(3):519-28.
 37. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*. 2004;340(4):783-95.
 38. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014;42(Web Server issue):W320-4.
 39. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*. 2009;30 Suppl 1:S162-73.
 40. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Sci Rep*. 2017;7(1):10480.
 41. Bienert S, Waterhouse A, de Beer TA, Tauriello G, Studer G, Bordoli L, et al. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res*. 2017;45(D1):D313-D9.
 42. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296-W303.
 43. Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics*. 2020;36(6):1765-71.
 44. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant*. 2020;13(8):1194-202.
 45. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40(7):e49.

46. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol.* 2018;35(6):1547-9.
47. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* 2014;12(6):e1001889.
48. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002;30(1):325-7.
49. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinf.* 2012;13:134.
50. Kettles N, Kopriva S, Malin G. Insights into the Regulation of DMSP Synthesis in the Diatom *Thalassiosira pseudonana* through APR Activity, Proteomics and Gene Expression Analyses on Cells Acclimating to Changes in Salinity, Light and Nitrogen. *PLOS ONE.* 2014;9(4):e94795.
51. Mir ZA, Ali S, Shivaraj SM, Bhat JA, Singh A, Yadav P, et al. Genome-wide identification and characterization of Chitinase gene family in *Brassica juncea* and *Camelina sativa* in response to *Alternaria brassicae*. *Genomics.* 2020;112(1):749-63.
52. Watanabe TK, R; Kawase, T; Tanabe, T; Mitsutomi, M; Sakuda, S; Miyashita, K. Family 19 chitinases of *Streptomyces* species: characterization and distribution. *Microbiology.* 1999;145:3353-63.
53. Durkin CA, Mock T, Armbrust EV. Chitin in diatoms and its association with the cell wall. *Eukaryotic Cell.* 2009;8(7):1038-50.
54. Cao J, Li X. Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum*). *Planta.* 2015;241(3):757-72.
55. Xie X, Wu N. Introns in higher plant genes. *Chin Sci Bull.* 2002;47(17):1409-15.
56. Cannon SB, Mitra A, Baumgarten A, Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 2004;4:10.
57. Moore RC, Purugganan MD. The evolutionary dynamics of plant duplicate genes. *Curr Opin Plant Biol.* 2005;8(2):122-8.
58. Smucker R. Chitin Primary Production. *Biochem Syst Ecol.* 1991;19(5):357-69.
59. Prihoda J, Tanaka A, de Paula WB, Allen JF, Tirichine L, Bowler C. Chloroplast-mitochondria cross-talk in diatoms. *J Exp Bot.* 2012;63(4):1543-57.
60. Cao J, Tan X. Comprehensive Analysis of the Chitinase Family Genes in Tomato (*Solanum lycopersicum*). *Plants.* 2019;8(3):52.
61. Kashyap P, Deswal R. A novel class I Chitinase from *Hippophae rhamnoides*: Indications for participating in ICE-CBF cold stress signaling pathway. *Plant Sci.* 2017;259:62-70.
62. Gupta R, Deswal R. Low temperature stress modulated secretome analysis and purification of antifreeze protein from *Hippophae rhamnoides*, a Himalayan wonder plant. *J Proteome Res.*

2012;11(5):2684-96.

63. Schneider S, Ullrich WR. Differential induction of resistance and enhanced enzyme activities in cucumber and tobacco caused by treatment with various abiotic and biotic inducers. *Physiol Mol Plant Pathol*. 1994;45:291-304.
64. Mock T, Samanta MP, Iverson V, Berthiaume C, Robison M, Holtermann K, et al. Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proc Natl Acad Sci U S A*. 2008;105(5):1579-84.
65. Dugdale R, Wilkerson F, Minas H. The role of a silicate pump in driving new production. *Deep Sea Res, Part I*. 1995;42(5):697-719.
66. Guillen D, Sanchez S, Rodriguez-Sanoja R. Carbohydrate-binding domains: multiplicity of biological roles. *Appl Microbiol Biotechnol*. 2010;85(5):1241-9.
67. Vaaje-Kolstad G, Horn SJ, van Aalten DM, Synstad B, Eijsink VG. The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *J Biol Chem*. 2005;280(31):28492-7.
68. Shen Z, Jacobs-Lorena M. A type I peritrophic matrix protein from the malaria vector *Anopheles gambiae* binds to chitin - Cloning, expression, and characterization. *J Biol Chem*. 1998;273(28):17665-70.

Additional Files

Additional file 1: Table S1. Table S1. Keywords used for searching *T. pseudonana* chitinase genes on JGI.xlsx

Additional file 2: Table S2. All transcripts annotated as chitinase in *T. pseudonana* genome from JGI database.xlsx

Additional file 3: Table S3. The information of chitinase genes from *T. oceanica*.xlsx

Additional file 4: Table S4. The information of chitinase genes from *T. weissflogii*.xlsx

Additional file 5: Table S5. Primers used in qRT-PCR studies.xlsx

Additional file 6: Table S6. Analysis and prediction of signal peptides, Transmembrane helices (TMHs) and Subcellular localizations of *TpChis*.xlsx

Additional file 7: Table S7. List of the putative motifs of the *TpChis*.xlsx

Additional file 8: Table S8. The proportion of secondary structure of *TpChis*.xlsx

Additional file 9: Table S9. Top 3D models predicted of the *TpChis* proteins.xlsx

Figures

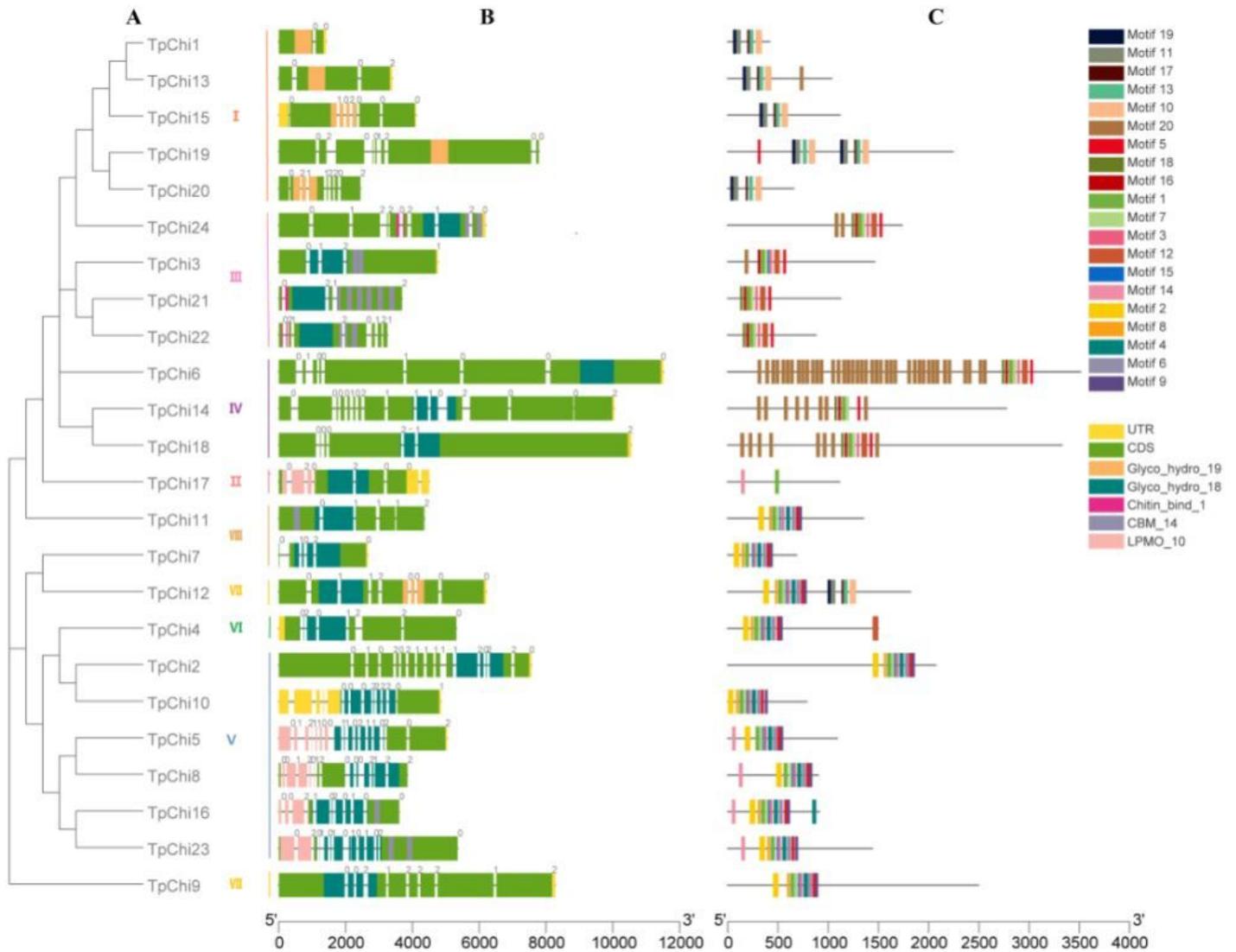


Figure 1

Phylogenetic relationships, gene structures, and analysis of domains and conserved motifs of the *T. pseudonana* chitinases. A) Phylogenetic tree of 24 *T. pseudonana* chitinase proteins. B) Gene exon/intron organization and domain distribution of the *T. pseudonana* chitinases. Grey lines represent introns, and the numbers 0, 1, and 2 represent their splicing phases. Yellow boxes represent untranslated regions (UTRs), and green boxes represent coding sequences (CDSs). Five chitinase-related domains, Glyco_hydro_19, Glyco_hydro_18, Chitin_bind_1, CBM_14, and LPMO_10, are indicated with orange, blue-green, magenta, grey, and salmon boxes, respectively, in the CDS translated regions. The lengths of exons and introns can be inferred by the scale at the bottom. C) Distribution of conserved motifs in the *T. pseudonana* chitinases. Twenty putative motifs are indicated with different coloured boxes. Table S7 presents details of the motifs.

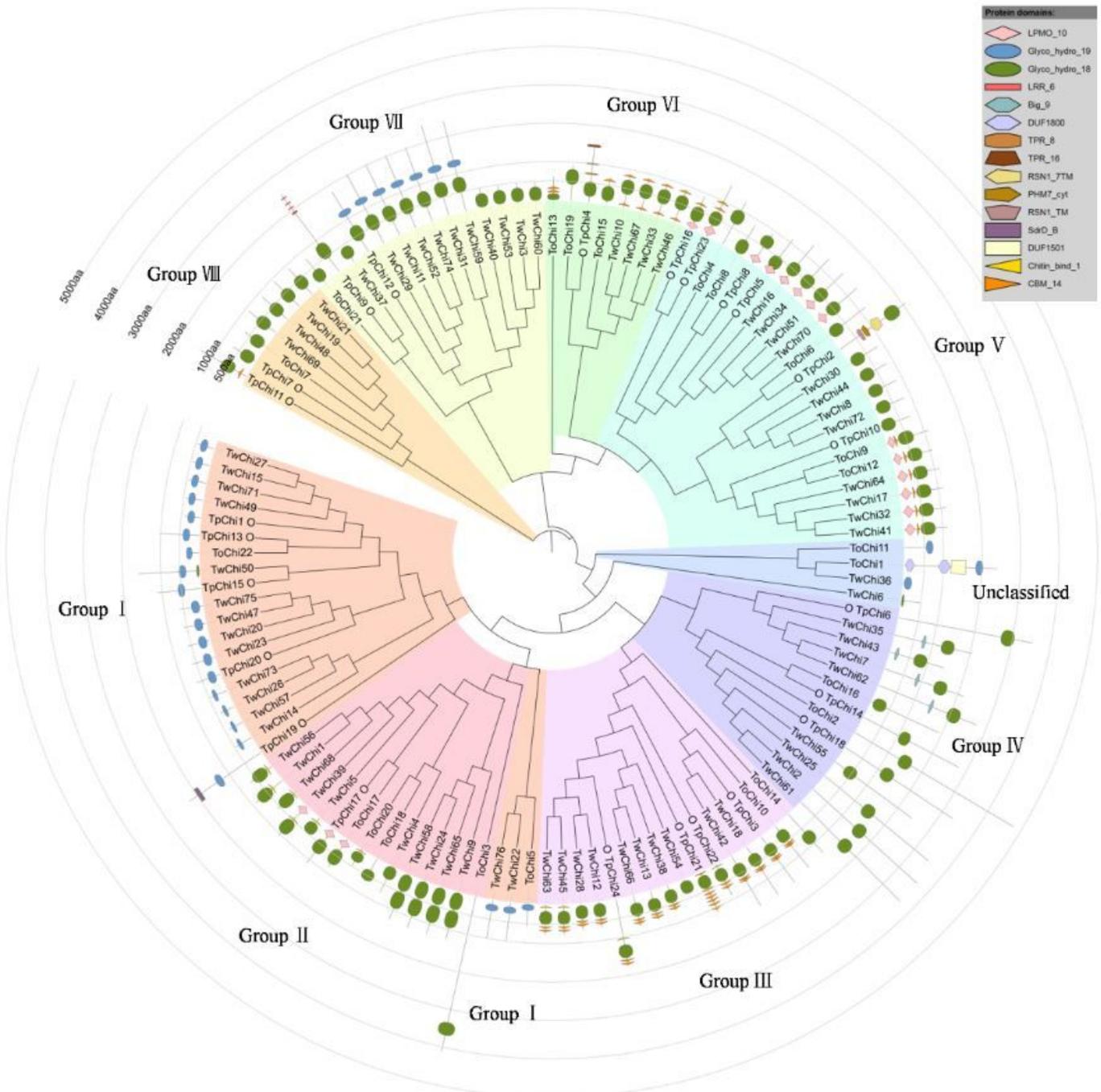


Figure 2

Phylogenetic relationships and domain architectures of the 122 chitinases (Chis) from *Thalassiosira oceanica* (To), *Thalassiosira pseudonana* (Tp), and *Thalassiosira weissflogii* (Tw). The *T. pseudonana* chitinases are indicated using the symbol "O". Inner layer: An unrooted phylogenetic tree constructed with MEGA using the NJ (neighbor-joining) method and 1000 bootstrap replicates. The eight groups (I–VIII) are distinguished by colour, and the four unclassified chitinases are shown in a light blue shade. Outer layer: Distribution of the chitinase protein domains, denoted by boxes of different colours and shapes. The

glyco_hydro_19 domains are represented by blue oval boxes, and the glyco_hydro_18 domains are represented by green oval boxes; their positions can be inferred by the circular scales around the tree.

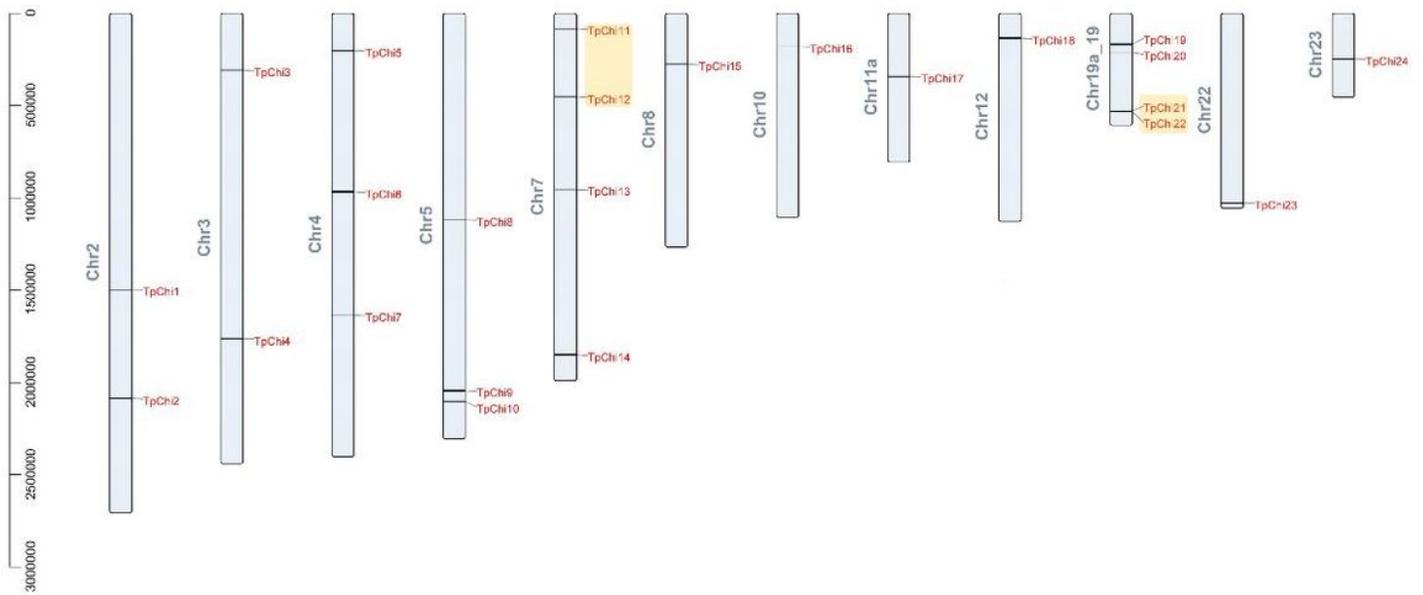


Figure 3

Scaffold locations and gene duplications of the TpChis. The tandem duplicated gene pairs are marked by yellow rounded rectangles.

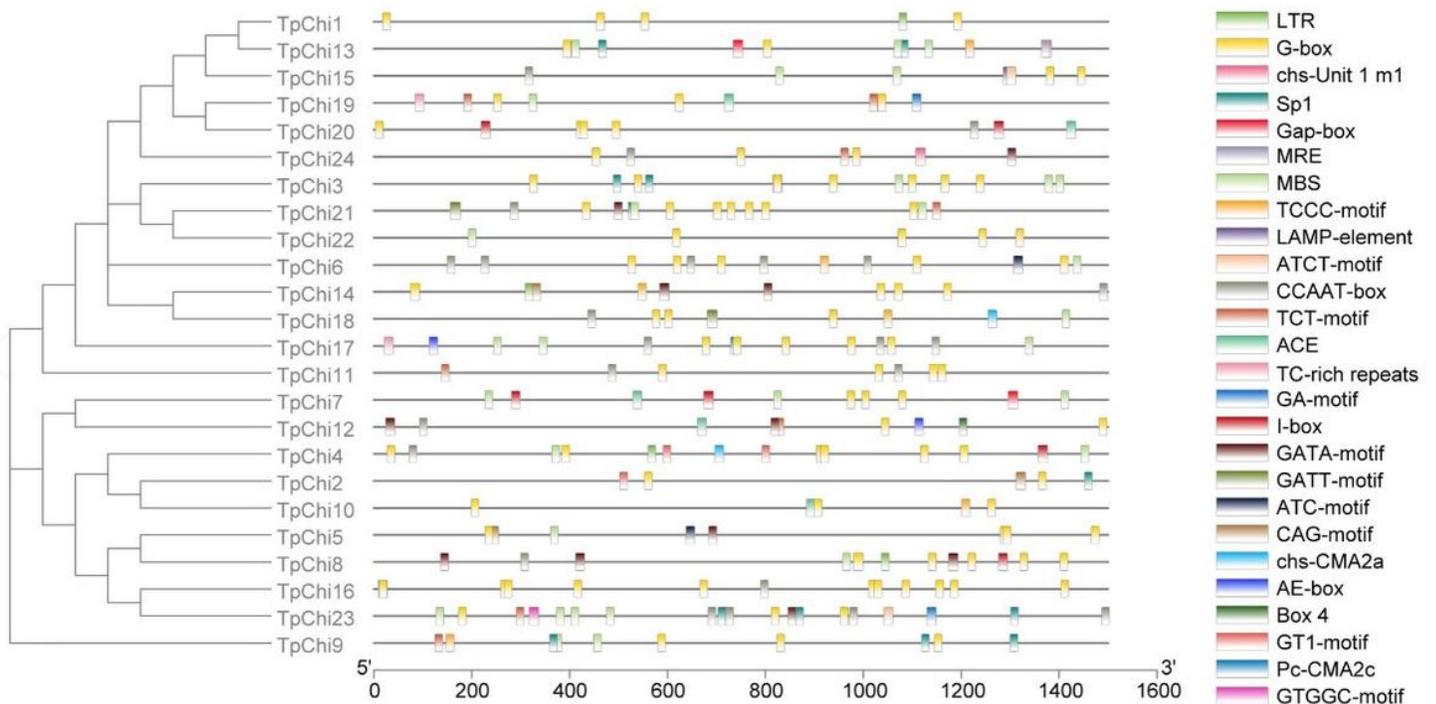


Figure 4

Predicted stress-related cis-acting elements in the promoter regions of the TpChis.

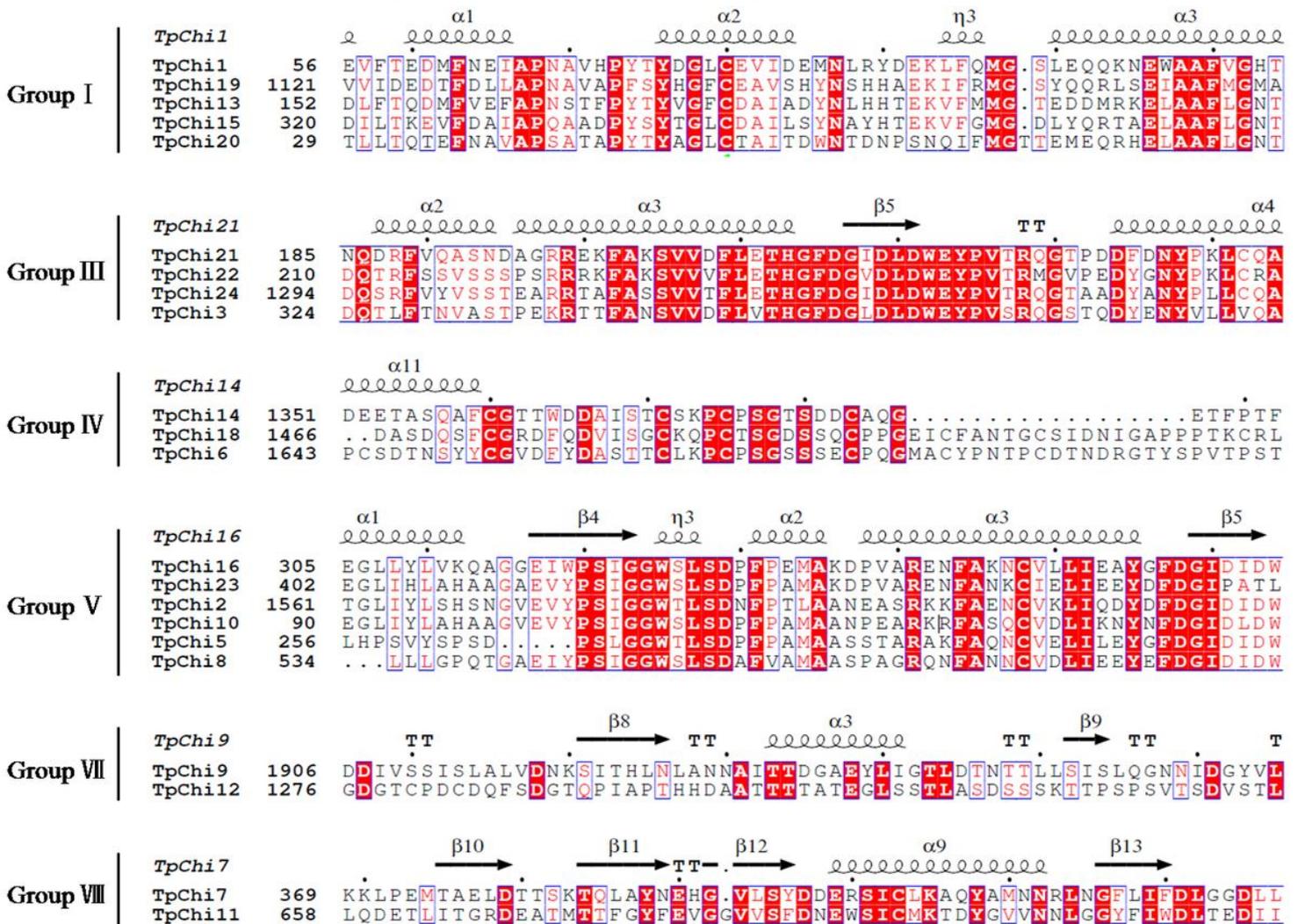


Figure 5

Representative secondary structures and sequence alignments of the *T. pseudonana* chitinases in each group. Amino acids in white on a red background are conserved sites, and similar amino acids are shown in red with blue rectangles. The secondary structures of the *T. pseudonana* chitinases are shown above the alignment. Alpha-helices are represented with helices symbols, β-strands with arrows, and turns with TT letters.

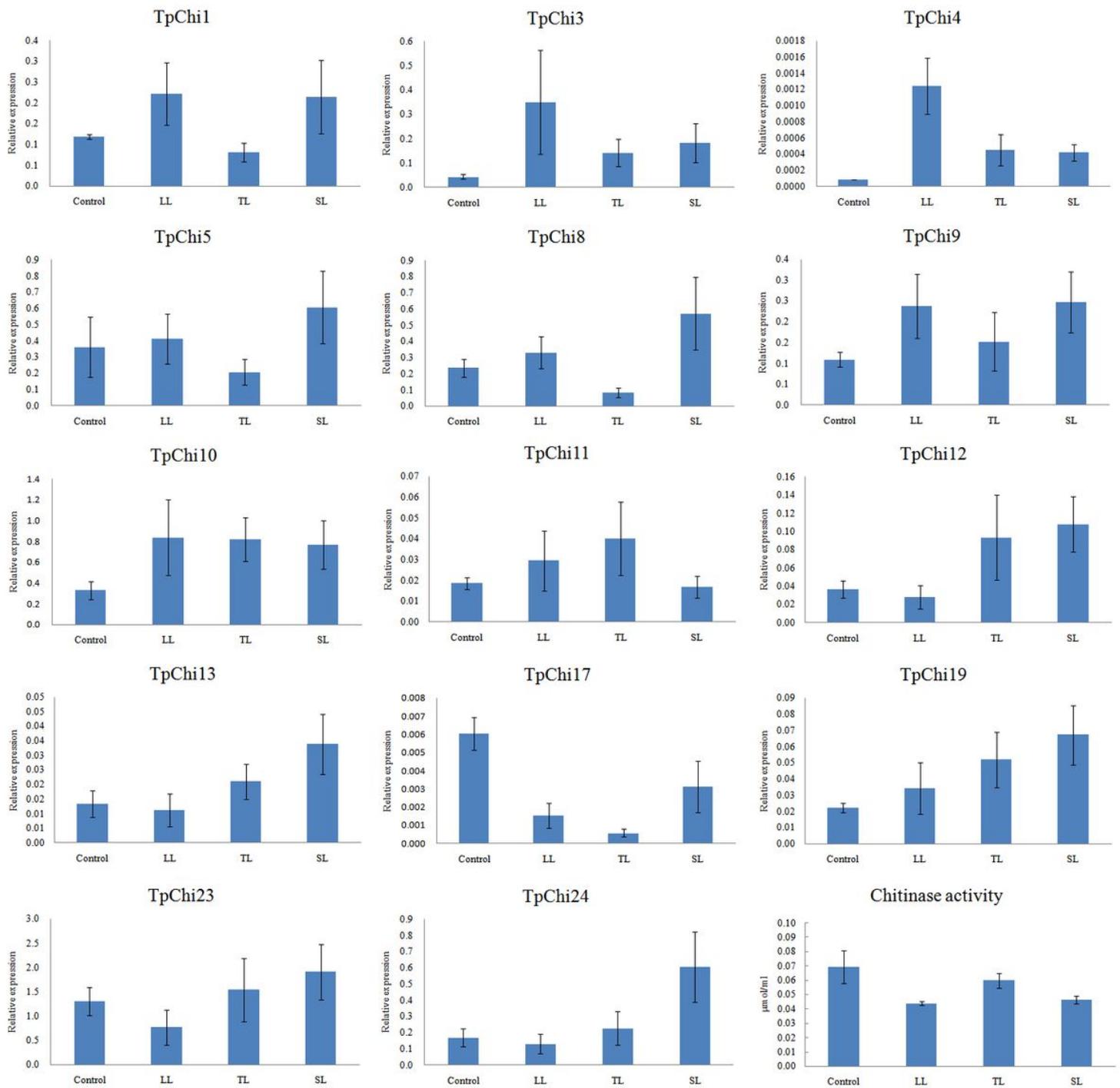


Figure 6

Expression profiles of 14 randomly selected chitinase genes under light-limited (LL), temperature-limited (TL), and silica-limited (SL) conditions. Data are presented as the mean \pm standard deviation of three replicates. The relative transcript levels of selected genes were calculated using the $2^{-\Delta\Delta\text{Ct}}$ method with beta-tubulin as the internal reference gene. Lower right: Chitinase activity of *T. pseudonana* cultures under LL, TL, and SL conditions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)
- [TableS8.xlsx](#)
- [TableS9.xlsx](#)