BMC Genomic Data

RESEARCH Open Access

Genome-wide identification and characterization of stress-responsive genes in *Chlorella vulgaris*

Yasmeen Khizar[1†], Umer Farooq[2†], Kotb A Attia[3*], Obaid Ur Rehman[2], Asmaa M. Abushady[4,5], Sajid Fiaz[6*], Umar Zeb[7], Rashid Iqbal[8,9] and Muhammad Uzair[2*]

Abstract

Background *Chlorella vulgaris* is a significant green alga that has a role in the bioremediation of environmental pollutants, especially heavy metals. Therefore, to meet the emerging needs of sustainable bioremediation, it is the need of the hour to improve the bioremediation potential of *Chlorella vulgaris*. Stress-related genes play significant roles in homeostasis and stress management in algal species, including *C. vulgaris*. It deals with varying pH and temperature, toxic heavy metals, oxidative stress, and many others. While certain stress-responsive proteins such as Heat Shock Proteins (HSPs) and Antioxidant Enzymes have been previously reported in *C. vulgaris*, this study aims to expand the scope by identifying and characterizing a diverse range of genes from various gene families, many of which have not been studied before in *C. vulgaris*.

Method A comprehensive analysis of the stress-related genes was conducted in which comparative phylogenetic analysis; conserved motif detection, determination of gene structure, and their subcellular localization were performed.

Results As a result of this study, 15 stress-related genes in *C. vulgaris* were annotated and characterized. The phylogenetic analysis represented that these genes evolved independently in *C. vulgaris*. Twenty highly conserved motifs amino acid structures have been exhibited. These motifs have a potential role in stress management. The proteins are localized at different locations in the cells. In parallel to genome-wide analysis, an experiment was conducted in a wet lab to evaluate the growth curve of *C. vulgaris* under Cd and pH stress.

Conclusions The results revealed a probability that *C. vulgaris* has some mechanisms and genes that act as key players for survival. Moreover, this study not only provides identification and characterization of stress-related genes

†Yasmeen Khizar and Umer Farooq contributed equally to this work.

*Correspondence:
Kotb A Attia
kattia1.c@ksu.edu.sa
Sajid Fiaz
sajidfiaz50@yahoo.com
Muhammad Uzair
uzairbreeder@gmail.com

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

but also lays the foundation for further identification, annotation, and confirmation by expression profiling under different stress conditions such as toxic heavy metals and pH.

Keywords *Chlorella vulgaris*, Stress-related genes, In Silico, Heavy metals, Algae

Introduction

Algae are found in aquatic and terrestrial environments and can respond to various environmental and physiological stressors [1]. One of the most researched microalgae is the single-celled green alga *Chlorella vulgaris*, a Chlorophyceae family member. It is widely explored due to its simple culture requirements, rapid growth rate, and genetic tractability [2]. Additionally, it is well known for its potential for environmental adaptation and biotechnological uses in producing biofuels, food and feed, and wastewater treatment [3]. It has higher photosynthetic potential and can robustly grow under mixotrophic and heterotrophic conditions. *C. vulgaris* is a source of protein and a potential candidate as an ideal biofuel feedstock. This green alga can fix carbon dioxide effectively and remove nutrients such as nitrogen and phosphorous, making it an excellent candidate for greenhouse gas bio mitigation and bioremediation of wastewater [3]. In addition, it also exhibits potential as an alternative expression host for the production of recombinant protein, although challenges need to be addressed [4].

Despite its potential benefits, heavy metal contamination by human activity threatens aquatic species' health, including *C. vulgaris* [5]. For instance, Cd can accumulate in biological tissues and disrupt metabolic processes, thus highly hazardous to aquatic species [6]. Additionally, heavy metals can interact with other environmental parameters, including pH, to impact *C. vulgaris*' physiology and development [7]. *C. vulgaris* exhibits diverse responses to various stressors, including temperature, nutrient deficiencies, and oxidative stress [8]. For instance, in an experimental study, *C. vulgaris* was more tolerant of Cd at higher pH levels and more sensitive to Cd at lower pH levels [9]. Additionally, pH substantially impacts the ultrastructure of *C. vulgaris* cells, with low pH values causing cells to have fewer and thicker cell walls and high pH values causing cells to have more and thinner cell walls [10]. Stress-responsive genes are present in various organisms, including plants, algae, and bacteria, which are vital for survival. They enable adaptation to various environmental stressors, both biotic and abiotic [11]. For instance, recent studies on stress-responsive genes in *Chlamydomonas reinhardtii* provide in-depth knowledge of the genetic mechanisms underlying stress tolerance in microalgae. A study identified the genetic response to singlet oxygen in a *C. reinhardtii* glutathione peroxidase knockout (gpx5) mutant, highlighting the role of antioxidant enzymes in oxidative stress management [12]. Another study reported increased mRNA levels of HSP70s and chaperone-related genes in *C. reinhardtii* under low temperatures, further signifying the importance of genes in stress tolerance [13]. These findings highlight the need for a comprehensive study to identify and characterize diverse gene families contributing to stress-response mechanisms. Gene expression regulation at transcriptional and post-transcriptional levels might influence and control important biological activities and processes such as perception of compounds, signal transduction, cellular morphogenesis, and environmental stresses [14]. Insights into stress-responsive genes can significantly contribute to enhanced stress tolerance in organisms. It can also enhance agricultural practices, environmental management, and biotechnological applications [15]. Studying the role of these genes in stress resistance has been an area of interest for a long time and is advancing with new technologies, e.g. Genome-wide analysis [16].

Understanding stress responses, genes, and respective proteins in *C. vulgaris* can provide insight into its adaptation, mechanisms, and potential biotechnological applications. Despite the significance of stress-responsive genes and proteins in *C. vulgaris*, there is a lack of comprehensive genome-wide study. In this study, we aimed to conduct a genome-wide identification and characterization of the stress-responsive genes in *C. vulgaris* for significant insights into the genetic basis of stress response. The identified stress-responsive genes and proteins may serve as targets for genetic engineering to improve stress tolerance and biomass productivity in *C. vulgaris.*

Material and methods

Gene selection process

A total of 65 stress-responsive genes were initially identified through a comprehensive review of the literature by focusing on genes known for their role in stress-responsive mechanisms such as oxidative stress, heavy metals detoxification, and pH adaption in model species *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, and *Coccomyxa subellipsoidea*. The genes were analyzed using a local BLASTp algorithm to access homology against *Chlorella vulgaris* (taxid: 3077, strain 122/11P) [17]. From this evaluation, 15 genes were selected based on high sequence homology, query coverage, low E-value, % identity, and sufficient alignment length, ensuring their relevance to stress responses in *C. vulgaris*. The prioritization of these genes ensures their relevance to the study objectives while serving as a foundation for future research.

Identification and annotation of stress-related genes in *Chlorella vulgaris*

Different genes were identified and selected by a thorough review of the literature in model species of algae to conduct a comprehensive investigation of stress-related genes in *Chlorella vulgaris* (*Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, and *Coccomyxa subellipsoidea*) using method explained in [16]. These species' selected stress-related genes were known for their ability to control many types of stress, including oxidative stress, heavy metal stress, and others. The homologous FASTA sequences of these genes were found in NCBI (https://www.ncbi.nlm.nih.gov*)* using a local BLASTp algorithm search [16]. All gene sequences were examined for domains in the PFAM database (http://pfam.xfam.org*)* to ensure the accuracy of the results and each anticipated sequence [18]. The functional domains of the genes associated with stress were manually identified [19]. In addition to the additional in *silico* study, 15 genes in *C. vulgaris* were classified as stress-related genes. Using the ExPASy translate program (https://web.expasy.org/translate/*)*, open reading frames of genes associated with stress were determined [20]. Then, using the primary sequences of the genes, ProtParam (https://web.expasy.org/protparam/*)*, a tool accessible at ExPASy, predicted the chemical and biophysical parameters of 15 Stress-related genes in *Chlorella vulgaris* [20]. For example, the coding sequence (CDS), protein length (aa), gene length (bp), grand average of hydropathicity (GRAVY), molecular weight (MW), isoelectric point (pI), aliphatic index (AI), and instability index were predicted to be explored by these characteristics.

Sequence alignment and phylogeny inference

The sequences of the genes involved in stress were obtained from the NCBI database (www.ncbi.nlm.nih.gov/*)*, and the MEGA11 tool (https://www.megasoftware.net/*)* was used to align several full-length gene sequences. By applying "align by ClustalW," the parameters were at their default state. Additionally, a neighbor-joining (NJ) algorithm-based comparative phylogenetic tree was created using the MEGA 11 program. This method has computational efficiency for large datasets of amino acid sequences, suitable for this exploratory study. The tree was displayed, and its reliability was tested using 1000 bootstrap replicates with 50% cut-off values. This threshold excluded poorly supported branches while retaining strong evolutionary relationships for interpretation [20].

Gene structure and conserved motif analysis

An online program called Gene Structure Display Server (http://gsds.gaolab.org*)* was used to forecast the introns and exons patterns of the structural information of the genes associated with stress [21]. Coding sections (CDS) of genes in BED file format were used to estimate gene structure. To predict major functional and conserved protein motifs, the MEME program (https://meme-suite.org/meme/*)* was employed. To conduct this study, parameters like the ideal motif width (3 residues), the number of distinct motifs (20), and the distribution of motifs (any number of repetitions) were adjusted [20].

Response of *Chlorella vulgaris* under cd and pH stress

Chlorella vulgaris was imported from the Fisheries Research Institute Penang, Malaysia. The selected microalgal species (10% (v/v) were inoculated in BG11 media with the provision of optimized environmental conditions [22]. The sub-culturing was done after 21 days. 500ppm stock solution of Cd heavy metal was prepared [23]. With the help of stock solution BG11 media was prepared with 5ppm Cd stress. Cd levels as low as 0.001–0.05 ppm can significantly affect sensitive microalgae, while more tolerant species exhibit survival and stress-response adaptations at higher concentrations. 5ppm concentration allows for assessing stress tolerance mechanisms and predicting genes contributing to cadmium detoxification. 5 culture mediums were prepared with varying pH of 5, 6, 7, 8, and 9 [23]. One culture medium was prepared without Cd stress and with normal pH. 15% (v/v) inoculum of microalgae was added into all culture mediums. All other required nutrients, light, and standard conditions were provided. 3 replicates were prepared for each sample. Samples were collected after intervals of 24, 48, 72, and 96 h under sterile conditions. The samples were collected and used for further analysis [24].

Cell counting was performed for microalgal growth evaluation on alternate days by using a hemocytometer and Olympus microscope [25]. From each sample, 5 readings were taken for cell count i.e. 1 from the central part, 2 from the lower parts, and 2 from the upper parts. The calculation for each sample was done as described by [26]:

$$Total\ cells = \frac{\{a\,(2\,boxes\ upper) + b\,(2\,boxes\ lower) + c\,(1\,box\ center)}{5\frac{cells}{ml} = total\ cells\ \times\ \left(\frac{1}{4}\right) \times\ 106}$$

Biomass yield was inspected after every 24 h for 4 days of culturing. For biomass yield measurement, 20 ml of microalgae culture was filtered with pre-dried microfiber filter papers and then dried in an oven at 80 °C for 4 h. As a control group 100 ml of distilled water was also filtered similarly. Algal dry weight was calculated by the formula:

$$\text{Algal dry weight}\ \left(\text{gL}^{-1}\right) = DWA - DWC \times\ V$$

Where DWA = average dry weight on the algal filter, DWC = dry weight on the control filter, V = volume of algal culture, and Biomass expressed in (g L^1) [26].

Results

Analysis and selection of genes

Fifteen stress-responsive genes out of 65 were selected for examination based on their alignment characteristics with the *C. vulgaris* genome. TRHX demonstrated the highest query coverage, significant percentage identity, and E-value, confirming its relevance to stress responses. CRD1, DLAT, NADH, and CTH1 exhibited more than 90 per cent query coverage, along with significant identities that validated their selection as key candidates. FOX1 demonstrated moderate identity but was selected for its excellent query coverage. Pcy1, Gpxh, petF and APX stood out for their robust alignment and significance in response. PETJ and ATX1 were included for their functional roles, while PCS1 and PCS2 were closely related homologs for their notable query coverages and identities. GSTF10 demonstrated a lower identity but was included for its alignment and potential significance in stress responses. These selected genes provide a comprehensive foundation for exploring stress response mechanisms in *C. vulgaris*. Table 1 summarizes the selection process for stress-related genes in *Chlorella vulgaris* based on query coverage, percentage identity, E-value, and sequence length.

Identification and annotation of Stress-responsive genes in *Chlorella vulgaris*

In algae, several genes related to stress have been identified and annotated. Genome-wide analyses of stress-related genes were conducted in a few algal species by analyzing publicly available data. The recent investigations used a homology study (Supplementary Table S1) to uncover a total of 15 stress-related genes in *Chlorella vulgaris*. The open reading frames (ORF) of stress-related genes were determined using the ExPASy program (Supplementary Table S2). The study used PFAM analysis to evaluate whether each potential gene contained the conserved domain, and it ultimately validated the existence of linked domains in the PFAM database. Table 2 lists the genes' symbols, accession numbers, lengths of coding sequences, and properties of these proteins.

Stress-related genes in *C. vulgaris* have physical and metabolic characteristics that were predicted from the main sequences of the genes by ProtParam, a program available at ExPASy. These features, such as protein length (aa), coding sequences (CDS), molecular weight (MW), isoelectric point (pI), aliphatic index (AI), extinction coefficients (EC), and grand average of hydropathicity (GRAVY) and their anticipated half-life, were predicted to be explored by these qualities. The putative proteins produced by the genes related to stress ranged in size from 115 to 2254 amino acid (aa) residues, with an average of about 509 aa. Their coding sequences ranged from 345 base pairs (bp) (TRXH) to 6762 bp (NADH). Theoretical molecular weights ranged from 12499.38 (TRHX) to 240939.2, while the theoretical pI ranged from 4.6 (FOX1) to 9.71 (Gpxh). The genome-wide analyses found significant variation in the aliphatic index, extinction coefficients, and GRAVY, which is consistent with the characteristics and suggests functional diversification and a high level of complexity among the Stress-related genes of *C. vulgaris* (Table 3).

Comparative evolutionary analysis of stress-related genes

The sequences of all the identified stress-related genes were aligned to gain access to the comparative evolutionary link between the stress-related genes in *Chlorella vulgaris* and model species such as *Arabidopsis thaliana*,

Table 1 Gene selection process

Sr. No	Gene	Selected Match	Query Coverage (%)	% Identity	E-value	Length (bp)
1	TRXH	D9Q98_007762	100	56.64	3.00E-40	115
2	DLAT	KAI3433633.1	98	50.16	0	642
3	CRD1	D9Q98_008510	98	65.01	0	398
4	CTH1	D9Q98_008510	94	69.09	0	398
5	FOX1	D9Q98_004511	92	41.02	0	1113
6	NADH	D9Q98_008339	96	59.22	2.00E-41	107
7	Pcy1	KAI3431670.1	65	72.63	4.00E-51	147
8	Gpxh	D9Q98_008038	95	47.1	6.00E-48	226
9	petF	D9Q98_004942	80	82.18	1.00E-58	127
10	APX	D9Q98_009390	72	70.5	2.00E-127	352
11	PETJ	BAC54100.1	68	69.91	1.00E-47	139
12	ATX1	KAI3438126.1	66	57.35	7.00E-89	269
13	PCS1	KAI3424114.1	50	55.74	4.00E-95	607
14	PCS2	KAI3424114.1	53	54.51	2.00E-90	607
15	GSTF10	KAI3426897.1	88	27.45	3.00E-15	242

Table 2 Identification of Stress-related genes in *Chlorella vulgaris*

Sr. No	Gene Symbol	Accession numbers	Nucleotides CDS (bp)	Length (aa)	II	pI	Mw	AI	EC	GRAVY	Half life
1	CRD1	KAI3426131.1	1194	398	37.67	8.2	45876.5	78.07	53,080	−0.288	> 10 h
2	Gpxh	KAI3426070.1	678	226	41.01	9.71	24137.49	74.29	21,555	−0.18	> 10 h
3	petF	KAI3430347.1	381	127	48.81	4.62	13470.01	76.14	7700	−0.204	> 10 h
4	TRXH	KAI3425787.1	345	115	24.14	4.83	12499.38	82.26	19,605	0.042	> 10 h
5	APX	KAI3430987.1	1056	352	33.09	6.02	38343.18	65.57	70,360	−0.467	> 10 h
6	PETJ	BAC54100.1	417	139	39.43	5.31	14326.25	85.18	14,105	0.072	> 10 h
7	Pcy1	KAI3431670.1	441	147	28.81	5	15018.94	88.37	4470	0.227	> 10 h
8	FOX1	KAI3431459.1	3339	1113	35.92	4.6	119236.87	79.79	145,705	−0.139	> 10 h
9	CTH1	KAI3426131.1	1194	398	37.67	8.2	45876.5	78.07	53,080	−0.288	> 10 h
10	ATX1	KAI3438126.1	807	269	47.66	7.74	30062.14	71.15	46,535	−0.428	> 10 h
11	PCS1	KAI3424114.1	1821	607	54.71	5.15	65152.21	73.18	66,890	−0.286	> 10 h
12	GSTF10	KAI3426897.1	726	242	57.59	6.99	26783.02	99.59	35,450	−0.068	> 10 h
13	PCS2	KAI3424114.1	1821	607	54.71	5.15	65152.21	73.18	66,890	−0.286	> 10 h
14	NADH	KAI3429699.1	6762	2254	31.92	6.19	240939.22	83.41	194,770	−0.168	> 10 h
15	DLAT	KAI3433633.1	1926	642	51.12	5.29	65543.14	82.46	49,180	−0.003	> 10 h

Table 3 Conserved sequences, sites, width, and e-values of Stress-related gene motifs in *C. Vulgaris*

Motif	Conserved amino acid sequences	e-values	Sites	Width
1	EPNLKHGINGYLYCNLPGLSFTQGDS-VRWHLVALGSEVDMHSPNLVGQT	5.2e-021	3	49
2	PGGTFTYRWYVPESAGPGPADPSTJL-WLYRSSVD LVGDPNAGL	4.3e-011	3	43
3	SPTMNFGNIIEWKKKEGDEVAPGDILCE-VETDKA SIEWEAQEEGFIAKI	1.2e-008	3	49
4	MMPGAARSVDVAMDNPGKWLVH-CRVNDHINA GM	6.6e-008	2	33
5	TYYIAAEPVDWDYVPAGGDFC	9.0e-007	3	21
6	IGSKYEKAQYQQYTDATFTE	2.8e-003	3	20
7	HHGJLGPIIRAEVGQVIQVVFKNBLDF-PANJHPDG	6.2e-003	3	35
8	QQQTQQQ	2.9e + 000	3	7
9	DPRRNWKGPWGWFAEELDGCC	1.1e + 001	5	21
10	PDKQKFPWFEYFENWC	7.3e + 000	2	16
11	FFQPKNEINALYADEKIGEWGFNSIQ	7.7e + 000	2	26
12	WQGGKLMSDEKWDFYKF	1.4e + 001	2	17
13	WCAPCRMIGHN	9.0e + 000	2	11
14	WFPEFJGRYCSADC	1.6e + 001	4	14
15	IMQRKCHT	7.7e + 000	3	8
16	RFYTPDFDEMTEJYNKKKKWN	3.9e + 001	2	21
17	LWWYNK	9.4e + 001	3	6
18	DYLNDHS	9.9e + 001	2	7
19	CGGDYC	1.0e + 002	2	6
20	NPWPGGKGEFKPDY	1.8e + 002	2	14

Chlamydomonas reinhardtii, and *Coccomyxa subellipsoidea*. The neighbor-joining method was used to create a phylogenetic tree (Fig. 1). All relationships among the stress-related genes were determined by full-gene multiple sequence alignment, and each sequence pair's unclear positions were eliminated (pairwise deletion option). The well-resolved tree produced by the comparative phylogenetic analysis of these gene sequences reveals that the stress-related genes are divided into three significant sub-families. These three gene subfamilies all show evidence of strong similarity with a common ancestor. *C. vulgaris* and other model species exhibit close homology of stress-related genes according to comparative analysis (*A. thaliana*, *C. reinhardtii*, and *C. subellipsoidea*).

Gene structure and conserved motif analysis and intron/exon organization of stress-related genes in *Chlorella vulgaris*

To understand more about the evolution of these genes, structural diversity and functional characterization of stress-related genes are important. The exon/intron structure of genes associated with stress was therefore analyzed using BED files; the results of the analysis indicated that all of the genes are full-length, meaning that the stress-responsive domains are located inside the CDS regions. Upstream of the CDS sections in these genes, nevertheless, exist intronic regions. Certain genes lacked intronic regions (Fig. 2). Blue boxes and pink lines, respectively, are used to display the coding sequence and the intronic regions. The genes are arranged according to the alphabet of their symbol. These findings suggested that each gene exhibits the highest level of common gene structure conservation.

Using an online MEME tool, conserved motifs analysis of stress-related genes was predicted (Fig. 3). To better comprehend the structural composition and diversity of genes, the conserved motif prediction is important. This study presents detailed information on the 15 stress-responsive genes, including name, width, and best hits. Twenty conserved motifs with various amino acids were identified in the current study presented in Table 3. Number of conserved domains, motifs and predicted subcellular localizations of selected genes are summarized in

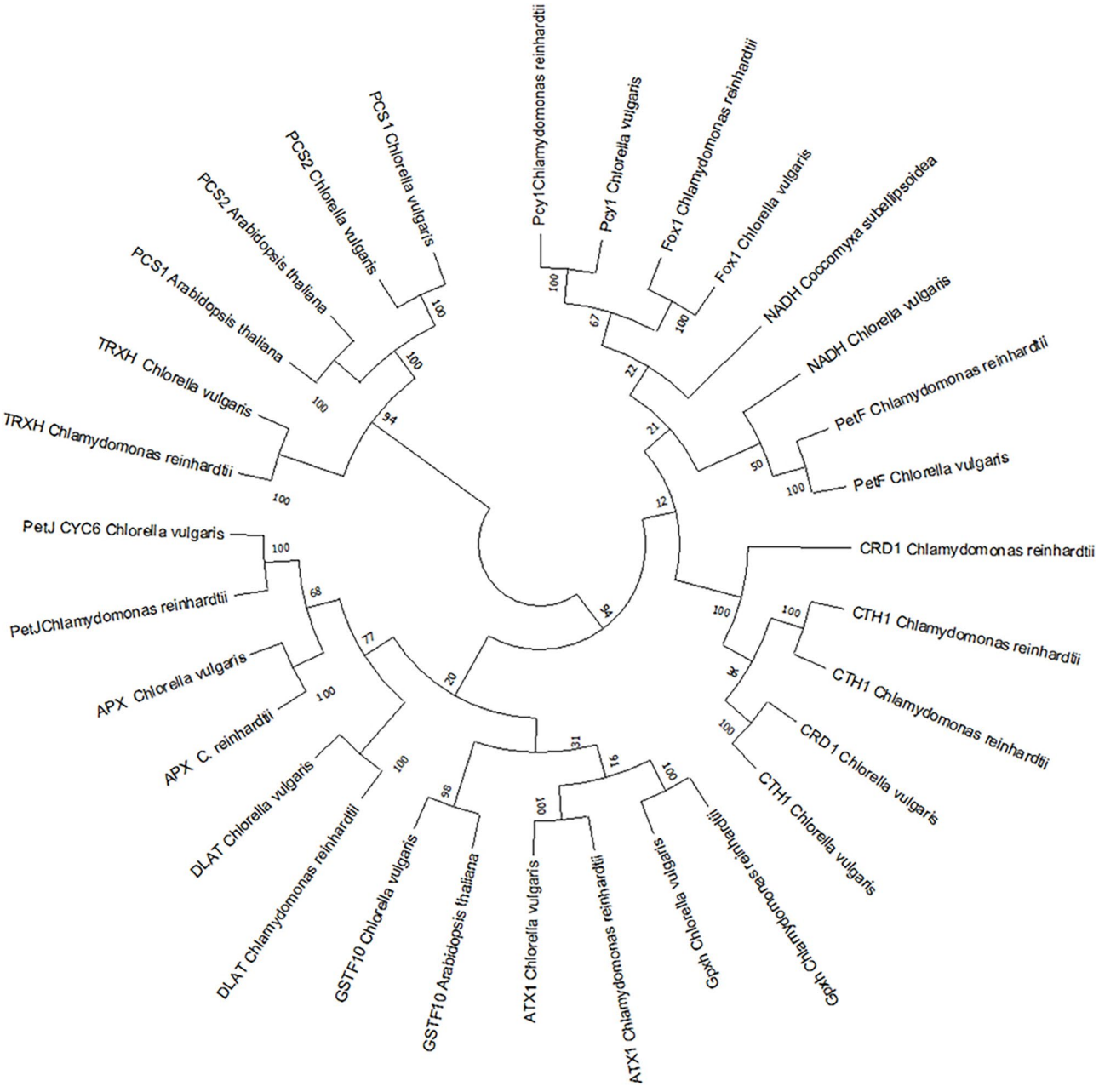

Fig. 1 The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates are collapsed. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. This analysis involved 31 amino acid sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 2257 positions in the final dataset. Evolutionary analyses were conducted in MEGA11

Table 4. It predicts that these genes exhibit diverse cellular functions. For instance, thioredoxin (TRXH) is localized in cytoplasm, predicts its role in redox reactions, 265 while glutathione peroxidase (Gpxh) localized in nucleus focusing its activity in oxidative stress. Phytochelatin synthases have multiple roles, localized in chloroplast, nucleus and peroxisome, predicting their role in heavy metal detoxification. These assessments provide insight into functional diversity of stress-responsive genes in *Chlorella vulgaris*.

Cell density and dry weight

The culture of *Chlorella vulgaris* was grown under 5ppm Cd stress and varying pH (5, 6, 7, 8, 9 and control group). The samples were collected for four consecutive days. Cell density and dry weight were measured to obtain

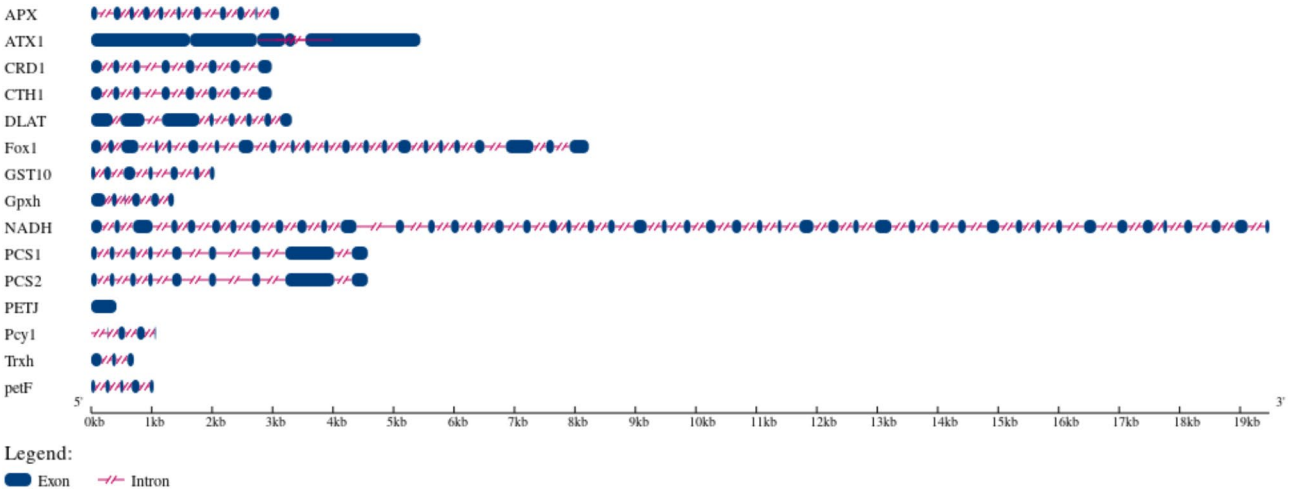

Fig. 2 The gene structure of Stress-related genes in *Chlorella vulgaris*. Blue boxes and pink lines, respectively, are used to display the coding sequence and the intronic regions. The symbols for the genes are arranged in alphabetical order

the growth curve. The growth curve is represented in graphs (Fig. 4). The growth curve indicates that *Chlorella vulgaris* showed excellent survival under given stressed conditions.

Discussions

Algae play a significant role on the planet Earth by providing more than half of the global production [27]. Algae have to face multiple stress conditions in their environment for their survival [28]. Some species of the algae show poor survival in stress conditions as they have no tolerance against stress conditions whereas some species have excellent survival and growth in stress conditions [29]. *Chlorella vulgaris* is one of the algal species with amazing tolerance against stressed environments such as varying pH and temperature, oxidative stress, toxic heavy metals stress, and many others [9]. Stress-related genes play a significant role in the survival of the species under stressed environments including *C. vulgaris* [30]. Species such as *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, and *Coccomyxa subellipsoidea* have reported stress-related genes that enable these species to survive in stressed conditions [31]. For instance, *CRD1* and *Gpxh* genes in *C. reinhardtii* are involved in the management of oxidative stress and protect against free radicals [32]. Moreover, the *PCS* gene in *Arabidopsis thaliana is* a metal-binding protein-encoding gene that helps to mediate the toxicity of heavy metals [33]. *C. vulgaris* is a potential algal species that plays its part in the bioremediation of environmental pollutants, especially heavy metals [34]. This study was conducted to explore the bioremediation mechanism of *C. vulgaris* mediated by stress-related genes. Identification and comparative genome-wide analysis were performed to explore the stress-related genes in *C. vulgaris*. The conducted study comprised comparative phylogenetic analysis, conserved motif detection, determination of gene structure, and their subcellular localization identification. 15 stress-related genes (*CRD1*, *Gpxh*, *PetF*, *PetJ*, *TRXH*, *APX*, *APX*, *Pcy1*, *Fox1*, *CTH1*, *Atx1*, *PCS1*, *GSTF10*, *PCS2*, *NADH*, and *DLAT*) were annotated and characterized in *C. vulgaris*. The analysis revealed that these genes play a significant role in stress management in *C. vulgaris*. The phylogenetic analysis revealed that selected stress-related genes represent evolutionary divergence and independent evolution relative to *Chlorella vulgaris*. The subfamilies suggest evolutionary convergence on stress response mechanisms, potentially indicating conserved functional domains originating from a shared ancestral lineage. The clustering patterns highlight close homology between stress related genes in *C. vulgaris* and other model species. It suggests similar evolutionary pressures have preserved critical functions in adaptation pathways across species [35]. However, certain branches with lower bootstrap values indicate weaker relationships, likely reflecting diverse functional origins or limitations in sequence conservation. The exploration underscores the potential relevance of these subfamilies to stress adaptations, as clusters such as PCS1/PCS2 and TRXH align with redox and detoxification regulation pathways [36]. The selected genes are not part of a single gene family, and their functional domains vary significantly. Future work will incorporate functional validation to assess gene expression changes under specific stress conditions, providing a clearer link between phylogenetic relationships and biological function. Moreover, including more stress-responsive genes would enhance the reliability and scope of evolutionary inferences. Selected cadmium (Cd) concentration and pH range were opted to mimic real-world stress conditions encountered in polluted

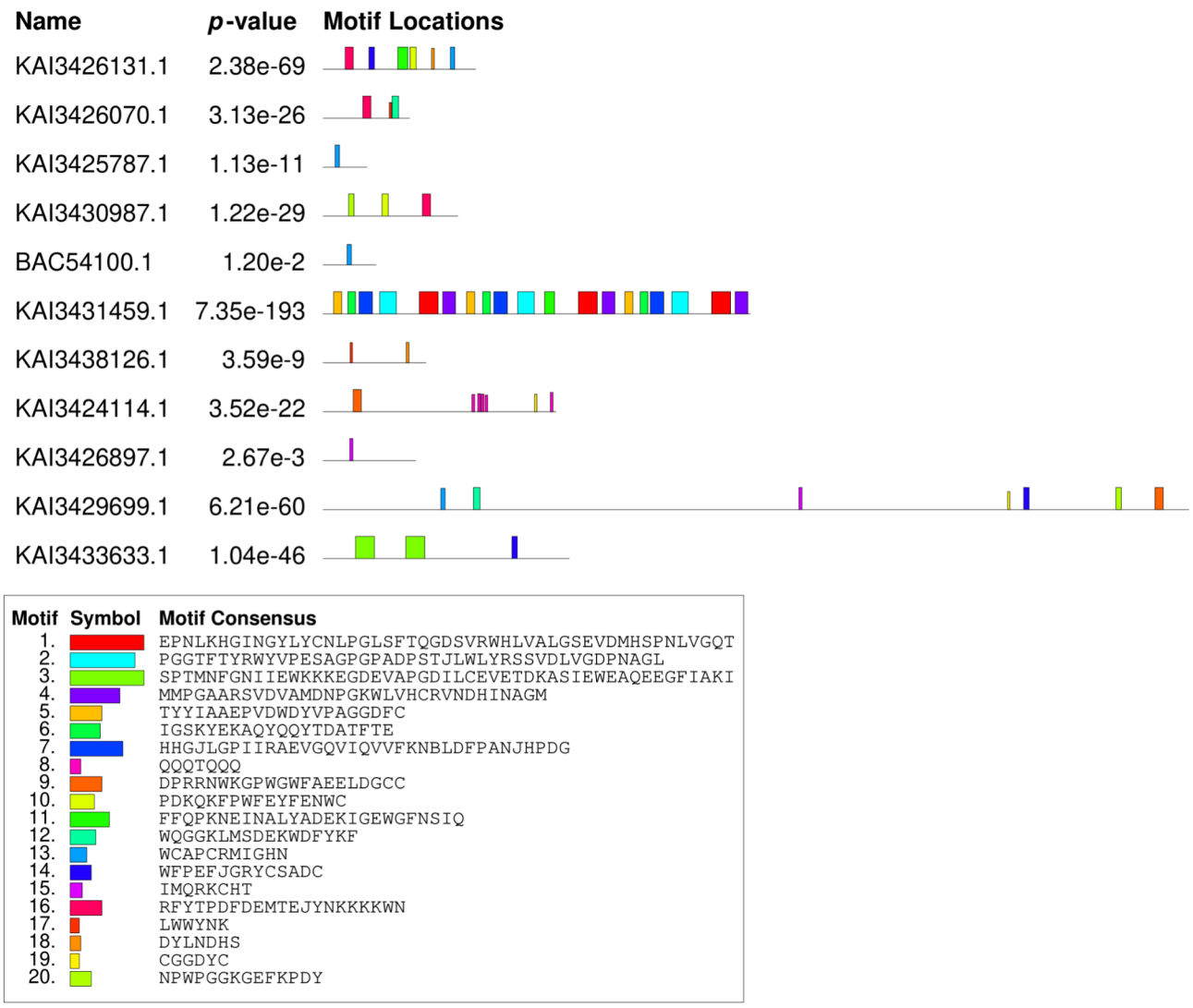

Fig. 3 Colored boxes represent the protein motifs of the genes associated with stress. Twenty predicted motifs were displayed in distinct colored boxes and grouped schematically by the phylogenetic tree's pattern

aquatic environments. This setup links the experimental setup and genome-wide analysis. These sub-lethal levels were designed to induce stress responses without completely inhibiting growth to observe the adaptive potential in *C. vulgaris*. The observed growth patterns under stressed conditions, such as reduced growth rates and dry weight, align with the functional roles of identified stress responsive genes. Many of the annotated genes contain conserved motifs, such as rubrerythrin, glutathione peroxidase, redoxin, thioredoxin, cytochrome C, peroxidase, phytochelatin synthases, and glutathione S-transferase, known for their role in stress tolerance mechanisms. Rubrerythrin protects cellular components from oxidative stress by mitigating oxidative stress by reducing hydrogen peroxide [37]. Glutathione peroxidase has a role in detoxifying reactive oxygen species (ROS) and maintaining cellular homeostasis [38]. Redoxin and thioredoxin motifs play their role in reducing disulfide bonds in proteins, helping protein repair and proper folding under stress conditions [39]. Likewise, cytochrome c is crucial for electron transport and cellular respiration, enabling energy production even under suboptimal environmental conditions [40]. Phytochelatin synthases chelate toxic heavy metal ions. Glutathione S-transferase motifs are crucial in detoxifying xenobiotics and ROS [41]. These motifs exhibit their likely involvement in conferring stress tolerance by modulating the physiological and biochemical responses of *Chlorella vulgaris*. The subcellular localization of stress-related genes in *C. vulgaris* focuses on their functional diversity, with TRXH in cytoplasm regulating redox balance and Gpxh in nucleus protects against oxidative stress [42]. FOX1 in cell membrane and Phytochelatins in chloroplast, nucleus and peroxisome underscore their role in electron transport

Table 4 Associated domains, motifs, and subcellular localization of Stress-related genes identified in *Chlorella vulgaris*

Sr. No	Gene	Selected Match	Domains	Motif	Location
1	TRXH	D9Q98_007762	3	Thioredoxin	cytoplasm
2	DLAT	KAI3433633.1	8	2-oxoacid dehydrogenases acyltransferase	Mitochondrion
3	CRD1	D9Q98_008510	2	Rubrerythrin	chloroplast
4	CTH1	D9Q98_008510	2	Rubrerythrin	chloroplast
5	FOX1	D9Q98_004511	4	Multicopper oxidase	Cell membrane, Cell wall
6	NADH	D9Q98_008339	15	Amino_oxidase Flavin containing amine oxidoreductase	chloroplast
7	Pcy1	KAI3431670.1	2	Copper binding proteins	chloroplast
8	Gpxh	D9Q98_008038	2	Glutathione peroxidase	Nucleus
9	petF	D9Q98_004942	4	Thioredoxin	chloroplast
10	APX	D9Q98_009390	4	peroxidase	Mitochondrion
11	PETJ	BAC54100.1	3	Cytochrom_C	chloroplast
12	ATX1	KAI3438126.1	2	AhpC-TSA-AhpC/TSA family	Nucleus
13	PCS1	KAI3424114.1	3	Phytochelatin synthase	Chloroplast. Nucleus. Peroxisome.
14	PCS2	KAI3424114.1	3	Phytochelatin synthase	Chloroplast. Nucleus. Peroxisome
15	GSTF10	KAI3426897.1	5	GST_N Glutathione S-transferase, N-terminal domain	cytoplasm

and metal detoxification [43]. The correlation between the functional roles of stress-responsive genes 346 reinforces the biological relevance of the identified genes and highlights their contributions to stress-responsive mechanisms in *C. vulgaris*. The roles of these genes are to be experimentally validated. While their homologs in model species are known to be involved in stress responses, such as heavy metal detoxification and oxidative stress regulation, it remains to be confirmed whether they exhibit identical functions or novel properties unique to *C. vulgaris*. Functional assessment involving RNA extraction and expression profiling will be conducted under a stressed environment for functional validation. RNA extraction followed by Quantitative PCR will be

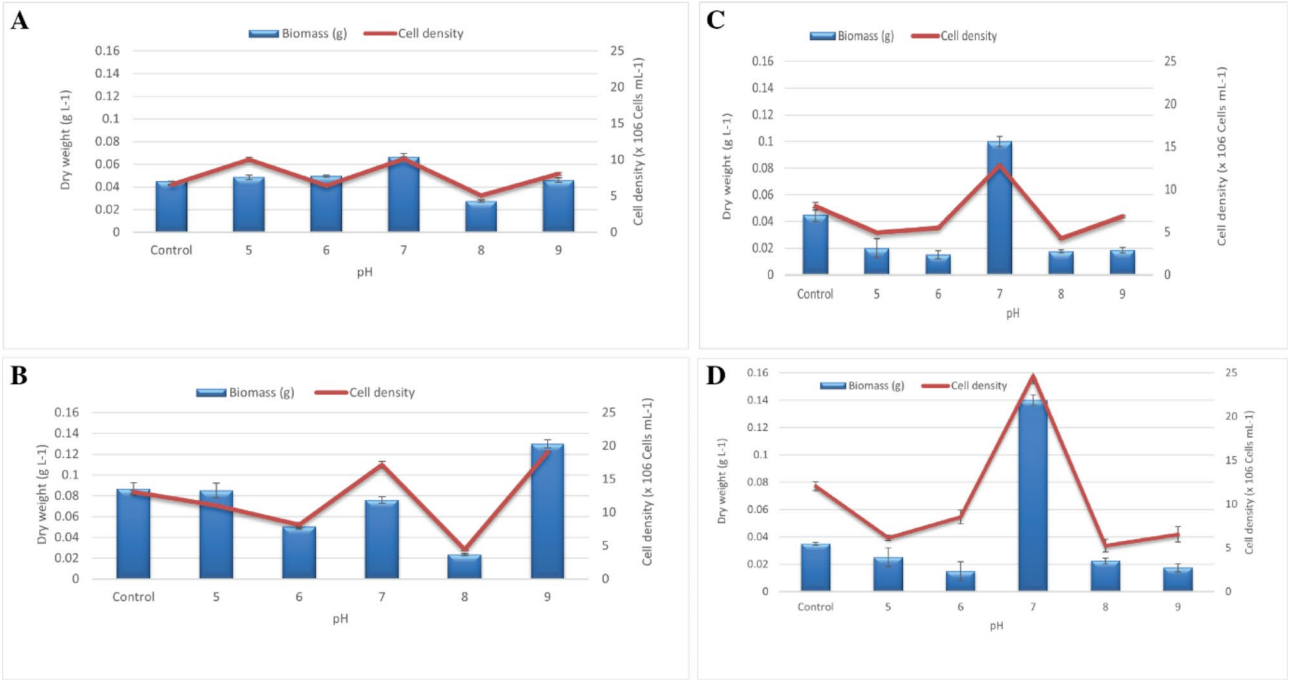

Fig. 4 Biomass and Cell density of *Chlorella vulgaris* under Cd (5ppm) and pH stress; Day 1 (**a**), Day 2 (**b**), Day 3 (**c**), and Day 4 (**d**)

employed to assess the expression levels of target genes under different stress treatments, such as cadmium exposure to pH variation [44]. Differential expression patterns will then be employed to assess the expression levels of target genes. By correlating expression data with phenotypic stress tolerance, the functional significance of key genes will be validated, providing experimental evidence to complement the bioinformatics predictions.

Conclusions

While the study focuses on 15 stress-related genes, we acknowledge that a more comprehensive study involving additional genes would provide a broader understanding of the genome-wide stress response mechanism in *Chlorella vulgaris*. Future studies will address this limitation by incorporating transcriptomic or proteomic approaches to identify and validate the roles of a more extensive set of stress-responsive genes under varying stress conditions, such as heavy metals, pH, and oxidative stress. Experimental validation will open the doors for advanced approaches such as CRISPR-based genetic modifications, and large-scale studies can be pursued to explore further and enhance the roles of stress-responsive genes and mechanisms in *C. vulgaris.*

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12863-025-01307-4.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
The authors extend their appreciation to the Researchers Supporting Project number (RSP-2025R-369), King Saud University, Riyadh, Saudi Arabia.

Authors' contributions
YK, UF, and MU designed the study. YK, UF, MU and OR perform the experiments, data collection, data analysis, and writing of the original draft. KAA, AMA, RI and UZ provided technical expertise to improve the revision of the article. UF and MU supervised the research. All authors review and edit the manuscript.

Funding
Researchers Supporting Project number (RSP-2025R-369), King Saud University, Riyadh, Saudi Arabia.

Data availability
All the relevant data are within the paper.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Biosciences, COMSATS University Islamabad, Islamabad, Pakistan
[2]National Institute for Genomics and Advanced Biotechnology (NIGAB), Park Road, Islamabad 45500, Pakistan
[3]Department of Biochemistry, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia
[4]Biotechnology School, 26th of July Corridor, Nile University, Sheikh Zayed City, Giza 12588, Egypt
[5]Department of Genetics, Agriculture College, Ain Shams University, Cairo, Egypt
[6]Institute of Molecular Biology and Biotechnology, The University of Lahore, Lahore 54590, Pakistan
[7]School of Food Science and Biological Engineering, Jiangsu University, Zhenjiang 212013, PR China
[8]Department of Agronomy, Faculty of Agriculture and Environment, The Islamia University of Bahawalpur, Bahawalpur, Pakistan
[9]Department of Life Sciences, Western Caspian University, Baku, Azerbaijan

Received: 17 October 2024 / Accepted: 27 February 2025
Published online: 21 March 2025

References

1. Lauritano C et al. Physiological and molecular responses to main environmental stressors of microalgae and bacteria in polar marine environments. Microorganisms. 2020;8(12):1957.
2. Spain O, Plöhn M, Funk C. The cell wall of green microalgae and its role in heavy metal removal. Physiol Plant. 2021;173(2):526–35.
3. Ru ITK, et al. Chlorella vulgaris: A perspective on its potential for combining high biomass with high value bioproducts. Appl Phycol. 2020;1(1):2–11.
4. Bolaños-Martínez OC, et al. Current status and perspective on the use of viral-based vectors in eukaryotic microalgae. Mar Drugs. 2022;20(7):434.
5. Manikandan A, et al. Emerging role of microalgae in heavy metal bioremediation. J Basic Microbiol. 2022;62(3–4):330–47.
6. Genchi G, et al. The effects of cadmium toxicity. Int J Environ Res Public Health. 2020;17(11):3782.
7. Expósito N, et al. Performance of Chlorella vulgaris exposed to heavy metal mixtures: linking measured endpoints and mechanisms. Int J Environ Res Public Health. 2021;18(3):1037.
8. Ahmad MT, et al. Applications of microalga Chlorella vulgaris in aquaculture. Reviews Aquaculture. 2020;12(1):328–46.
9. Geng W, et al. Response and tolerance ability of Chlorella vulgaris to cadmium pollution stress. Environ Technol. 2022;43(27):4391–401.
10. Soto-Ramírez R, et al. Effect of growth conditions on cell wall composition and cadmium adsorption in Chlorella vulgaris: A new approach to biosorption research. J Hazard Mater. 2021;411:125059.
11. Barati B, et al. Green algal molecular responses to temperature stress. Acta Physiol Plant. 2019;41:1–19.
12. Ma X, et al. Glutathione peroxidase 5 deficiency induces lipid metabolism regulated by reactive oxygen species in Chlamydomonas reinhardtii. Microb Pathog. 2020;147:104358.
13. Yoo H. Chaperones remodel Stress-Induced biomolecular condensates. The University of Chicago; 2021.
14. Jaiswal D, Pandey-Rai S, Agrawal SB. Untangling the UV-B radiation-induced transcriptional network regulating plant morphogenesis and secondary metabolite production. Environ Exp Bot. 2021;192:104655.
15. Mathur P, Roy S. Insights into the plant responses to drought and decoding the potential of root associated Microbiome for inducing drought tolerance. Physiol Plant. 2021;172(2):1016–29.
16. Zhu F, et al. Bringing more players into play: leveraging stress in genome wide association studies. J Plant Physiol. 2022;271:153657.
17. Gilchrist CL, et al. Cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters. Bioinf Adv. 2021;1(1):vbab016.
18. Mistry J, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412–9.
19. Liu M, et al. The draft genome of a wild barley genotype reveals its enrichment in genes related to biotic and abiotic stresses compared to cultivated barley. Plant Biotechnol J. 2020;18(2):443–56.

20. Khurshid K, et al. Genome wide identification and characterization of nodulation related genes in Arachis hypogaea. PLoS ONE. 2022;17(9):e0273768.
21. Shi Y, et al. Comprehensive analysis of glutamate Receptor-like genes in rice (Oryza sativa L.): Genome-Wide identification, characteristics, evolution, chromatin accessibility, GcHap diversity, population variation and expression analysis. Curr Issues Mol Biol. 2022;44(12):6404–27.
22. Dani N, et al. Isolation, screening and medium optimization of native microalgae for lipid production using nutritional starvation strategy and statistical design. Int J Environ Sci Technol. 2021;18:2997–3012.
23. Ajith M, Priyadarshini E, Rajamani P. Effective and selective removal of heavy metals from industrial effluents using sustainable Si–CD conjugate based column chromatography. Bioresour Technol. 2020;314:123786.
24. Abinandan S, et al. Potential of acid-tolerant microalgae, desmodesmus Sp. MAS1 and heterochlorella Sp. MAS3, in heavy metal removal and biodiesel production at acidic pH. Bioresour Technol. 2019;278:9–16.
25. Baroni ÉG, et al. The effect of nitrogen depletion on the cell size, shape, density and gravitational settling of Nannochloropsis Salina, Chlorella sp.(marine) and Haematococcus pluvialis. Algal Res. 2019;39:101454.
26. Uy KD. Algal analysis through a Chip-Based flow cytometer. University of Toronto (Canada); 2019.
27. Ullmann J, Grimm D. Algae and their potential for a future bioeconomy, landless food production, and the socio-economic impact of an algae industry. Org Agric. 2021;11(2):261–7.
28. de Carpentier F, Lemaire SD, Danon A. When unity is strength: the strategies used by Chlamydomonas to survive environmental stresses. Cells. 2019;8(11):1307.
29. Kharel HL, et al. Cyanidiales-Based Bioremediat Heavy Met BioTech. 2023;12(2):29.
30. Toghyani MA, et al. Two distinct time dependent strategic mechanisms used by Chlorella vulgaris in response to gamma radiation. J Appl Phycol. 2020;32:1677–95.
31. Colina F, et al. Genome-wide identification and characterization of CKIN/SnRK gene family in Chlamydomonas reinhardtii. Sci Rep. 2019;9(1):350.
32. Jan M, et al. Retrograde and anterograde signaling in the crosstalk between Chloroplast and nucleus. Front Plant Sci. 2022;13:980237.
33. Filiz E, et al. Comparative analyses of phytochelatin synthase (PCS) genes in higher plants. Biotechnol Biotechnol Equip. 2019;33(1):178–94.
34. León-Vaz A, et al. Impact of heavy metals in the microalga Chlorella sorokiniana and assessment of its potential use in cadmium bioremediation. Aquat Toxicol. 2021;239:105941.
35. La Rosa R, et al. Compensatory evolution of Pseudomonas aeruginosa's slow growth phenotype suggests mechanisms of adaptation in cystic fibrosis. Nat Commun. 2021;12(1):3186.
36. Agarwal P, et al. MYB4 transcription factor, a member of R2R3-subfamily of MYB domain protein, regulates cadmium tolerance via enhanced protection against oxidative damage and increases expression of PCS1 and MT1C in Arabidopsis. Plant Sci. 2020;297:110501.
37. Knop R et al. A rubrerythrin locus of Clostridioides difficile efficiently detoxifies reactive oxygen species. Anaerobe. 2025;92:102941.
38. Dorion S, Ouellet JC, Rivoal J. Glutathione metabolism in plants under stress: beyond reactive oxygen species detoxification. Metabolites. 2021;11(9):641.
39. Chakraborty S, et al. S-denitrosylation: a crosstalk between glutathione and Redoxin systems. Antioxidants. 2022;11(10):1921.
40. Nesci S, Lenaz G. The mitochondrial energy conversion involves cytochrome C diffusion into the respiratory supercomplexes. BIOCHIMICA ET BIOPHYSICA ACTA-BIOENERGETICS; 2021;1862:148394–6. 6.
41. Faizan M et al. Phytochelatins: A key regulator against heavy metal toxicity in plants. Plant Stress, 2024;11:100355.
42. Xu J, et al. Probiotic-inspired nanomedicine restores intestinal homeostasis in colitis by regulating redox balance, immune responses, and the gut Microbiome. Adv Mater. 2023;35(3):2207890.
43. Yuan H, et al. Gene activation of metazoan Fox transcription factors at the onset of metamorphosis in the marine demosponge Amphimedon Queenslandica. Development, Growth & Differentiation; 2022;64:455–68. 8.
44. Zhao F et al. An optimized protocol for Stepwise optimization of real-time RT-PCR analysis. Hortic Res. 2021;8:179.

Publisher's note