diversity

MDPI

Article

The Implication Inferred from the Expression of Small Heat-Shock Protein Genes in Dinoflagellate Resting Cysts Buried in Marine Sediment

Yunyan Deng [1,2,3], Fengting Li [1,4], Zhangxi Hu [1,2,3], Caixia Yue [1,4] and Ying Zhong Tang [1,2,3,*]

1 CAS Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; yunyandeng@qdio.ac.cn (Y.D.); lifengting@qdio.ac.cn (F.L.); zhu@qdio.ac.cn (Z.H.); yuecaixia@qdio.ac.cn (C.Y.)
2 Laboratory of Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and Technology, Qingdao 266237, China
3 Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao 266071, China
4 University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: yingzhong.tang@qdio.ac.cn; Tel.: +86-532-8289-6098

Abstract: Dinoflagellates are unicellular eukaryotic microalgae, occupying pivotal niches in aquatic ecosystems with great ecological, biological, and economic significance. Small heat shock proteins (sHsps) are the most omnipresent, but the least conserved, family of molecular chaperones found in all domains of life. Although their common name (small Hsp) implies to exclusively stress their heat shock-responsive function, many sHsps in fact engage in a variety of physiological processes, from cell growth and proliferation to embryogenesis, development, differentiation, apoptosis, and even to human disease prevention. Recent years have greatly expanded our understanding of sHsps in higher plants; however, comprehensive study aiming to delineate the composition and expression pattern of dinoflagellate sHsp gene family has not yet been performed. In this study, we constructed dinoflagellate-specific environmental cDNA library from marine sediment and sequenced using the third-generation sequencing technique. Screening of sHsp genes from the library returned 13 entries with complete coding regions, which were considered to be transcriptionally activated in the natural community of dinoflagellate resting cysts. All the 13 dinoflagellate sHsps consisted of a solely characteristic α-crystallin domain, covering 88–123 amino acid residues with the typical A-X-X-X-N-G-V-L motif, flanked by variable N- and C-terminal extensions. Multiple alignment revealed considerable amino acid divergence (~26.7% average similarity) among them. An unexpected close relationship was revealed between dinoflagellate and green algal sHsps in the phylogenetic tree, seemingly reflecting a close evolutionary relationship of these sHsps themselves. We confirmed that sHsp mRNAs are expressed during dormancy of the resting cyst assemblages of dinoflagellates that were buried in marine sediment, which raised the possibility that the sHsp expression is part of the machinery of maintaining the dormancy or/and the adaptation to ambient conditions of dinoflagellate resting cysts. Our results, although preliminary, gained an important glance on the universal presence of sHsps in dinoflagellates and their active expressions in the assemblage of resting cysts that were buried in the marine sediment. The essentiality of sHsps functioning in resting cysts necessitate more intensive and extensive investigations on all possible functions of Hsps in dinoflagellates, a group of protists with vital ecological and biological importance.

Keywords: dinoflagellate; dormancy; environmental cDNA library; harmful algal blooms; marine sediment; resting cysts; resting stage persistence; small heat shock protein (sHsp)

Citation: Deng, Y.; Li, F.; Hu, Z.; Yue, C.; Tang, Y.Z. The Implication Inferred from the Expression of Small Heat-Shock Protein Genes in Dinoflagellate Resting Cysts Buried in Marine Sediment. *Diversity* **2021**, *13*, 471. https://doi.org/10.3390/d13100471

Academic Editors: Michael Wink, Savvas Genitsaris and Maria Moustaka-Gouni

Received: 16 August 2021
Accepted: 24 September 2021
Published: 27 September 2021

1. Introduction

Dinoflagellates are an ancient and diverse group of eukaryotic plankton, occupying crucial niches in aquatic ecosystems. As the second largest group of eukaryotic photoautotrophs, they are pivotal contributors to primary productivity and food webs of coastal

marine ecosystems and engage in the global carbon fixation, element cycles, and oxygen production [1]. A number of photosynthetic species in this lineage have been extensively studied due to their mutualism relationship with other living organisms, such as reef-building corals and invertebrates [2]. Many other members in this group receive great attention for being the most common causative agents of harmful algal blooms (HABs), which account for up to roughly 40% of the species forming HABs worldwide [3]. This algal group also includes the highest number of toxin-producing species among marine phytoplankton, which directly poison animals or indirectly affect human health through consumption of seafood [4]. The dinoflagellate HAB events threaten coastal ecosystems, fisheries, and human health through releasing toxins and causing local oxygen depletion. Many species of phytoplankton include a relatively long-term resting stage in their life history, which are proven to be fundamental to their ecology and evolution [5,6]. The resting stage of dinoflagellates is well-known as resting cysts, which can endure harsh environmental conditions after being settled in the marine sediment and survive an elongated period of time in dormancy (decades to as long as a century, in the sediment; [7]). They are proposed to play greatly vital roles in the ecology of dinoflagellates, particular for those HABs-causative species, as they have been well documented to be linked with genetic recombination, initiation and termination of blooms, resistance to adverse environments, protection from viruses, grazers, or parasite attacks, geographic expansion of populations, and historical records and paleoecological indicators for environmental changes [8–10].

Organisms across kingdoms in nature are constantly subjected to acutely and chronically environmental stress, which might cause deleterious effects on proteome infrastructure and cellular homeostasis. The response to stresses at molecular level is found in all kingdoms of life, which always induces activation of gene sets involved in cell survival and/or cell death, especially the sudden changes in genotypic expression leading to increase in the synthesis of protein groups called "heat-shock proteins (Hsps)" [11–14]. Hsp members are widespread molecular chaperones that play a fundamental role in assisting the correct folding of nascent proteins and preventing aggregation of stress-accumulated misfolded proteins, thus contributing to modulate the cellular proteostasis (protein quality control and protein homeostasis) [13,14]. Almost all kinds of stressing agents could trigger gene expression and/or protein synthesis of Hsps [11,12]. Although their name implies exclusively to stress relevant function, many Hsps constitutively express under normal conditions and engage in a broad range of physiological processes [11].

The Hsps are classified on the basis of sequence homology and approximate subunit molecular weight into Hsp100, Hsp90, Hsp70, Hsp60, and small Hsp (sHsp) family [12]. The sHsps are a large and ancient family of proteins, which have been found in all domains of life, Achaea, Eukarya, and Bacteria, and even in viruses [14–17]. Their monomers are characterized by low molecular weights of 12–43 kDa and a highly conserved α-crystallin domain (ACD) (Pfam PF00011), which represents their identification label [14–16,18,19]. The ACDs are flanked by N- and C-terminal extensions variable in both sequence and length between orthologues that might be pertaining to functional specificity and/or preferential chaperone activity [13]. Proteins in sHsp family generally present in the form of monomers or dimers, or assemble into large multimeric complexes varying in size of approximately 200–800 kDa and including up to 24–40 subunits [13,14,18]. The oligomerization and dissociation of sHsps are regulated by post-translational modification, which in turn influence their binding affinity to specific clients and, therefore, modulate their functions [19]. In contrast to chaperones with ATPase activity (such as Hsp70s, Hsp90s, and Hsp100s) classified as chaperone "foldases", sHsps without ATPase activity are classified as "holdases" chaperone, since they can recognize a large variety of misfolded and/or unfolded substrates and bind them to avoid irreversible aggregation [18,19]. Therefore, sHsp members act as the first line of the cellular proteostasis maintenance, and the sHsp-bound substrates are further processed by downstream ATP-dependent chaperones for renaturation and reactivation [18]. The sHsp family is the most omnipresent, but the least conserved, family of molecular chaperones, which covers particularly heterogeneous

members [15,20,21]. The abundance and heterogeneity imply that sHsps may play multiple and unique physiological roles [20]. For stress resistance, sHsps protect cells against a broad range of hostile environmental insults, such as temperature shock, dehydration, oxidative stress, UV exposure, osmotic shock, pathogen challenge [13,14,19,22]. Apart from primary chaperone function, publications on sHsps have recorded that they are implicated in diversely essential processes, involving cell growth and proliferation, embryogenesis, differentiation, apoptosis, actin and intermediate filament dynamics, insect development, cytoskeletal organization, membrane fluidity, and human disease prevention [13,19,22].

In contrast to the extensive characterization of other Hsp members, sHSP chaperones generally have received little attention [17]. In dinoflagellates, the growing world of Hsps are advanced primarily upon investigations on individuals of Hsp70 and Hsp90, information concerning sHsp members is quite rare and lags significantly behind [22]. Differential expressions were found for *Hsp20*-annotated genes in *Symbiodinium* species exposure to thermal stress from transcriptomic-based resources [23,24]. One identified Hsp20 protein from dinoflagellate *Karenia mikimotoi* was merely detected in nitrate sufficient cells and proposed to be involved in growth and cell proliferation [25]. Based on full-length cDNA isolation, our recent work on an *Hsp20* gene from the HAB-forming dinoflagellate *Scrippsiella trochoidea* observed totally different transcriptional profiles between thermal and cold fluctuations, implying its function as in most cases leading to protection from cellular heat-induced aggregation [22]. Developmental regulation of sHsps are widely reported in terrestrial plants during pollination, embryogenesis, seed germination, and fruit ripening, although their precise roles and underlying mechanism are still far from being fully elucidated [12–15,26]. In our previous transcriptomic work on *S. trochoidea*, we intriguingly detected resting cysts significantly down-regulated expressions in three genes encoding Hsp70s [27]. Markedly prominent transcription of an *Hsp40* gene was observed in newly formed resting cysts of this species, relative to those in vegetative cells and mature cysts [22]. Similar data showing in dinoflagellate *Akashiwo sanguinea*, mRNA abundance of an *Hsp70* peaked in newly formed resting cysts and then declined to lower levels in more mature cysts [28]. Another two Hsp genes, *Hsp60* and *Hsp10*, also displayed a significant increase from vegetative cells to newly formed and long-stored cysts [29]. All the results above together imply a probable involvement of Hsp members in the life stage alternation of dinoflagellates [22].

In this study, using the unique dinoflagellate mRNA-specific spliced leader (DinoSL; [30]) as a selective primer, we constructed DinoSL-selected environmental cDNA (e-cDNA) library from marine sediment and sequenced using the third-generation sequencing technique. We then identified sHsp genes from the DinoSL-based full-length e-cDNA library and thus all the retrieved sHsp sequences are considered to be from natural community of dinoflagellate resting cysts. Evaluation of sHsp members' expression in field highlight the putative role they might play in dormancy maintenance of dinoflagellate in natural scenarios. Our results enhance the current knowledge about the sHsps properties in dinoflagellates and form concrete basis for further functional validation in this ecological important lineage.

2. Materials and Methods

2.1. Sediment Collection and Resting Cysts Separation

Sediment sample was taken from Jiaozhou Bay, Qingdao, Shandong Province, China (36.159° N, 120.229° E) on 24 May 2016. The upper 10 cm of sediment was collected by using a grab sampler, transferred to sterile plastic bags, and kept in an ice-contained cooler to maintain darkness and low temperature. After being transported to the laboratory, sediment sample was immediately used for resting cysts separation. More than 200 g (wet weight) surface sediment (at layer of 3–6 cm depth) was used following the standard protocols described in Bolch et al. (1997) [31] by using sodium polytungstate (SPT). The harvested resting cysts were washed with fresh sterile filtered seawater several times, concentrated by centrifugation, and immediately used for subsequent RNA extraction.

2.2. RNA Extraction, cDNA Synthesis, and cDNA Amplification

Total RNA was extracted from all the harvested resting cysts of each replicate sediment sample using RNeasy Plant Mini Kit (Qiagen, Hilden, Germany), and digested with Qiagen's RNase-Free DNase Set, according to the manufacturer's instructions. RNA quality and quantities were measured by NanoDrop^{TM} 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The single-strand cDNAs were synthesized from ~1 µg total RNA and made with SMARTer PCR cDNA Synthesis Kit (Clontech, Mountain View, CA, USA) by using the anchor primer (5′-GCTGTCAACGATACGCTACGTAACGGCATGACA GTGTTTTTTTTTTTTTTTTT-3′). Resultant cDNAs were then purified with Zymo DNA Clean and Concentrator Kit (Zymo Research, Orange, CA, USA) and used as templates in the PCR amplifications. To minimize PCR bias, PrimeSTAR GXL DNA High Fidelity Polymerase (Takara, Tokyo, Japan) was used. PCR was performed with the specific primers DinoSL (5′-TCCGTAGCCATTTTGGCTCAAG-3′) and Racer3 (5′-TGTCAACGATACGCTA CGTAACG-3′) to amplify the dinoflagellate-specific cDNAs [32,33] under the program: 94 °C for 3 min, followed by 95 °C 15 s, 68 °C 3 min 30 s for 5 cycles; 95 °C 15 s, 62 °C 30 s, 72 °C 3 min for 15 cycles, and a final extension at 72 °C for 10 min [34].

2.3. PacBio Iso-Seq Sequencing and Data Processing

Sequencing library was constructed generally according to the Isoform Sequencing (Iso-Seq) protocol and the BluePippin Size Selection System protocol as described by Pacific Biosciences (PN 100-092-800-03). Size fractionation and selection were performed using the BluePippin with the following bins: 0.5–1.8 kb and >1.8 kb. The DinoSL-based e-cDNA library underwent single-molecule real-time (SMRT) sequencing using 6 SMRT cells on PacBio Sequel platform. Sequencing was performed by BGI Genomics Co., Ltd. (Shenzhen, China).

The raw SMRT sequencing reads were subjected to SMRT Analysis Server v2.3.0 supported by PacBio to filter out low-quality polymerase reads (read-length < 50 bp and read-score < 0.75). The CCSs (circular consensus sequences) were filtered from the subreads with the full pass threshold set to $\geq$0 and the predicted unique accuracy set to $\geq$0.75. Based on examining for cDNA primers and poly (A) signal, CCS reads were classified into full-length non-chimeric (FLNC), non-full-length (nFL), chimeras, and short reads. Short reads less than 500 bp were discarded. The CCSs with all three elements (5′-cDNA primer, a 3′-cDNA primer, and a poly (A) tail) and not containing any additional copies of the adapter sequence within the DNA fragment were selected and classified as FLNC (full-length non-chimeric) sequences. Then 5′- and 3′-cDNA primers and poly (A) tail were removed from FLNCs according to the PacBio recommended procedure. The obtained FLNC reads were subsequently clustered by Iterative Clustering for Error Correction (ICE) algorithm to generate the cluster consensus transcripts. Then, nFL reads were used to polish the above cluster consensus transcripts to yield the full-length polished high quality consensus sequences (accuracy $\geq$ 99%). The final full-length cDNA sequences were yielded by removing the redundant sequences with software CD-HIT using a threshold of 0.98 identities [35].

2.4. Identification and Sequence Analysis of sHsp Genes from the Dinoflagellate-Specific cDNA Library

The sequence homology and predicted functions of the generated SMRT genes were annotated using BLAST2GO under the translated nucleotide BLAST (BLASTX) algorithm. Then, the transcript pool was searched by keyword alpha crystallin protein, small heat shock protein, sHsp, and heat shock protein to collect potential sHsp candidates. The retrieved entries were further confirmed by conducting BLASTX searches against public databases, including the NCBI non-redundant protein database (Nr) database, NCBI non-redundant nucleotide database (Nt) database, SwissProt database, InterPro database, with a 10^{-6} E-value cutoff. The potential protein encoding segments in the identified nucleotide sequences were searched via Open Reading Frame Finder [36]. Simple Modular

Architecture Research Tool (SMART) tool was used to predict the conserved domains in the identified protein sequences [37]. The conserved domain (ACDs) were further confirmed by Pfam [38] and aligned using MAFFT v.7 with default settings [39]. Predicted molecular weight and isoelectric point were calculated by ProtParam tool [40].

2.5. Phylogenetic Analysis

The 13 newly generated sequences from the DinoSL-based full-length e-cDNA library, together with the other 47 sHsp homologues from archaea, bacteria, fungi, higher plants, animals, and algae (Supplementary Table S1) downloaded from GenBank database, were aligned by using the MAFFT v.7 with the G-INS-i algorithm [39]. The alignment was further refined manually with BioEdit 7.0.9.1 [41]. Bayesian inference (BI) analysis was performed with MrBayes 3.2.6 [42]. Posterior probability was estimated using four chains running 1,000,000 generations and trees were sampled every 100 generations. The first 25% of sampled trees were discarded as burn-in prior to generating a 50% majority-rule consensus tree. The posterior probabilities for all branches were calculated using a majority-rule consensus approach. Phylogenetic trees were visualized using FigTree v1.4.4.

3. Results

3.1. Screening of Nuclear Dinoflagellate sHsp Sequences from Dinoflagellate-Specific e-cDNA Library

The characteristic feature of the sHsp members is the presence of ACD, which is widely used as the criterion for assigning a sequence to the sHsp family [14–16,18,43]. The preliminary keywords search (alpha crystallin protein, small heat shock protein, sHsp, and heat shock protein) returned 22 entries containing at least one instance of ACD. Then, the 22 hits were manually queried in public databases for more accurate identification. After excluding sequences with very short coding region and/or without diagnostic sHsp conserved amino acids [13,14,16,18], a total of 13 sHsp genes with complete ORF (open reading frame) were identified from the dinoflagellate-specific full-length e-cDNA library. Their ORFs spanned 459–840 bp in length coding for proteins of 152–279 amino acid residues. The generated sHsp proteins were named on the basis of their molecular weight. Large variation in predicted molecular weight and isoelectric point were shown among these 13 members (Table 1). The new yielded sequences were deposited in GenBank with the accession numbers MZ485793-MZ485805 (Table 1).

Table 1. Characterizations of identified sHsp sequences from dinoflagellate-specific e-cDNA library.

Protein Name	Deduced Amino Acid Residues (aa)	Molecular Weight (KDa)	Isoelectric Point [1]	ACD Location (Position) [2]	Accession Number
Hsp16.0	152	16.04	6.73	55–144	MZ485793
Hsp17.6	157	17.60	9.62	15–128	MZ485794
Hsp24.6	219	24.57	9.65	79–191	MZ485795
Hsp18.1	172	18.13	4.90	51–138	MZ485796
Hsp19.9	185	19.94	5.34	40–128	MZ485797
Hsp20.3	183	20.25	9.33	37–152	MZ485798
Hsp19.7	175	19.69	9.22	18–124	MZ485799
Hsp17.7	160	17.68	10.26	18–120	MZ485800
Hsp21.2	194	21.23	10.14	55–162	MZ485801
Hsp18.4	178	18.42	5.66	47–133	MZ485802
Hsp22.3	208	22.30	5.29	55–142	MZ485803
Hsp30.4	274	30.40	6.62	99–221	MZ485804
Hsp30.1	279	30.14	6.60	107–194	MZ485805

[1] The isoelectric point of each sHsp was predicted via ProtParam tool [40] based on the deduced amino acid sequence. [2] It shows the location of the conserved ACD in deduced amino acid sequence of each sHsp.

3.2. Sequences Comparison of Dinoflagellates Nuclear sHsps

Sequences comparison and homologous analysis were conducted with the 13 sHsp entries obtained in current study and one sHsp gene containing complete coding region from dinoflagellate *Scrippsiella trochoidea* (GenBank accession: MN698834; [22]). All the 14 members consisted of a solely characteristic ACD flanked by variable N- and C-terminal extensions. The ACDs covered 88–123 amino acid residues with the typical A-X-X-X-N-G-V-L motif (Figure 1). The N- and C-terminal extensions were highly variable both in length and amino acid sequence. An average amino acid identities of about 26.7% was detected among 14 sHsp members from dinoflagellate species (Supplementary Table S2). Two entries coding for Hsp18.1 (MZ485796) and Hsp16.0 (MZ485793) exhibited the highest similarity of 79.5%. Merely 12.5% similarity was found between MZ485803 coding for Hsp22.3 and MZ485800 coding for a sHsp protein with predicted molecular mass of 17.7 kDa.

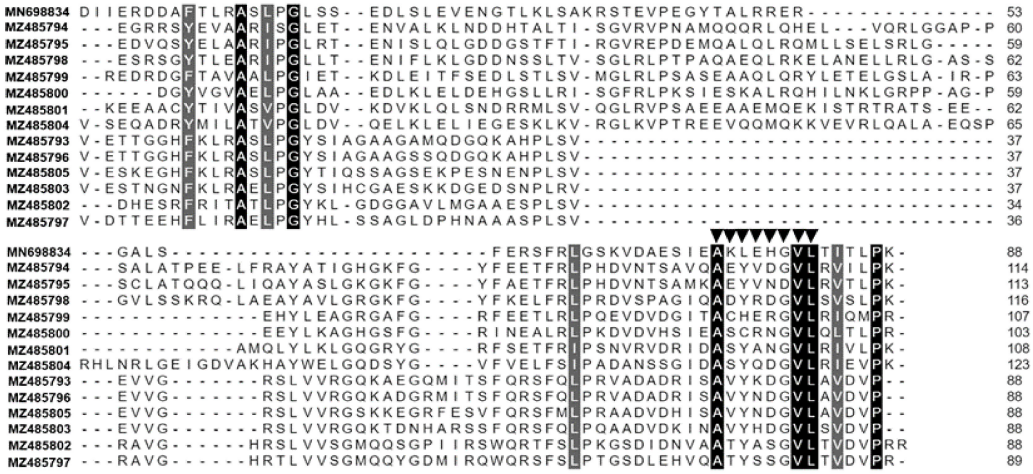

Figure 1. Alignment and comparison of the ACD domains of 14 dinoflagellate sHsps. GenBank accession numbers of these sequences are noted on the left. Similar and identical amino acid residues are gray and black shaded, respectively. The typical A-X-X-X-N-G-V-L motif inside the ACD region are marked by triangles above alignment. Deletions are shown by dashes.

3.3. Phylogeny Analysis

The Bayesian Inference (BI) distance tree formulated from alignment of the full amino acid sequences of complete coding regions of 60 sHsp sequences was evaluated by bootstrap. Bacterial sequences were used as outgroups. Generally, entities from archaea and higher plants grouped together to form sister clades with high node support (BI 0.98), which were basal to other homologues (Figure 2). The entities from members of the alveolates clustered together and also branched early. The 14 dinoflagellate sequences (13 newly identified sequences and one sequence reported in our previous study) were spread into sister branches with a medium node support of 0.77, which then formed a clade with the green algal sequences (Figure 2). A close relationship was found between animal and fungal sHsps with a BI node support of 0.87, which paralleled with the clades including those of algal species (Figure 2).

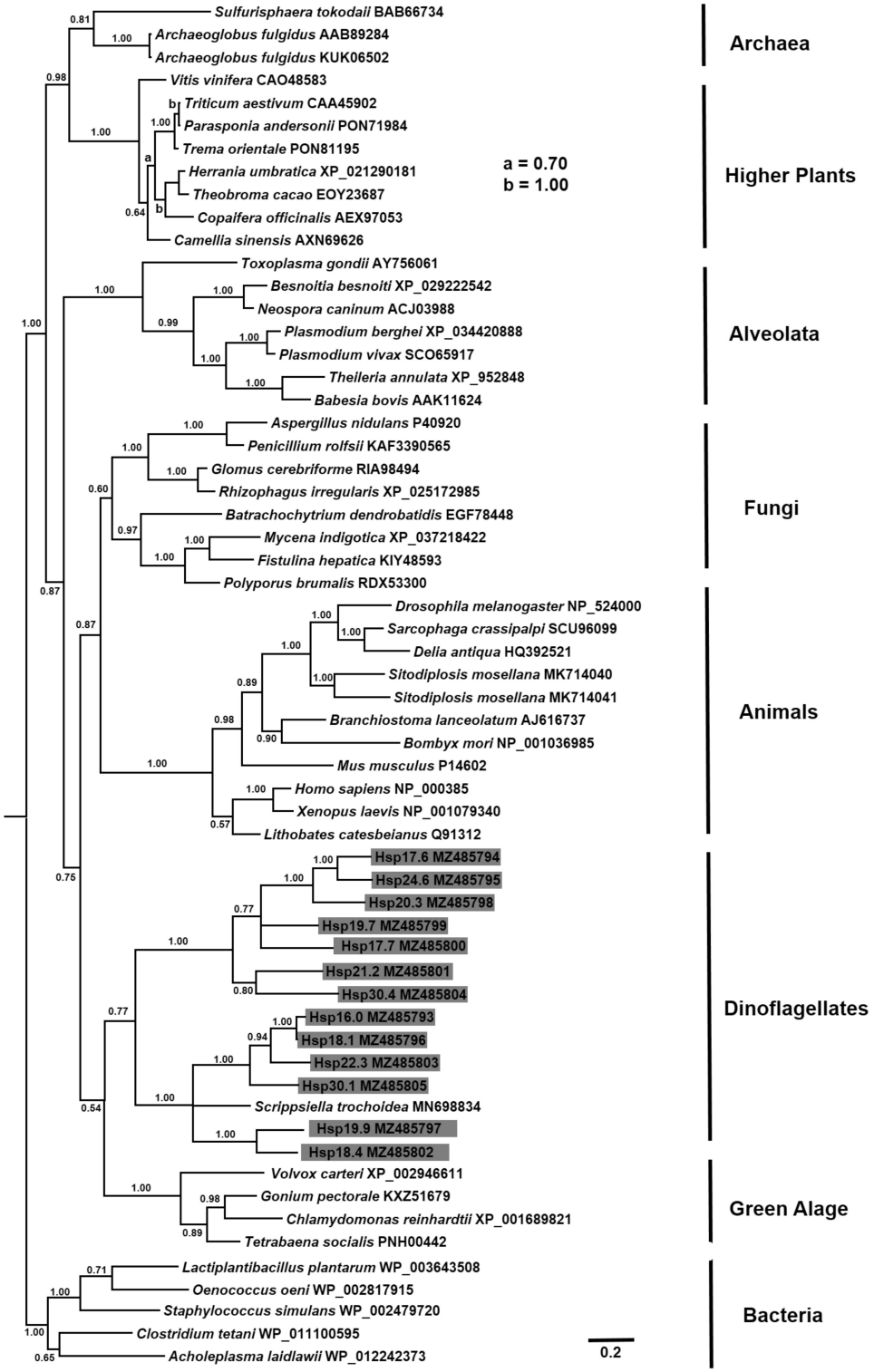

Figure 2. Phylogenetic reconstruction inferred from sHsp amino acid sequences using the Bayesian-Inference (BI) method. Numbers at the nodes represent BI posterior probabilities. GenBank accession number of each sequence is noted following the species name. The 13 newly yielded sequences are present as gene name following GenBank accession number and highlighted in gray background.

4. Discussion

Spliced leader (SL) *trans*-splicing is an mRNA processing mechanism, in which a short and conserved non-coding RNA fragment (SL; generally 15–50 nt) is *trans*-spliced to the splice acceptor site in the 5′-untranslated region of an independently transcribed pre-mRNAs [44]. This phenomenon was initially described in trypanosomes, and later was recorded in a phylogenetically disjointed lineage of eukaryotes [45,46]. While the functions of SL *trans*-splicing are far from entirely understood, the underlying mechanism is presumed to be similar among different groups and the currently known SL sequences are verified to be highly lineage specific [30,44,45,47]. Dinoflagellate mRNAs experience *trans*-splicing with a 22 nt SL sequence DinoSL, which is proofed to be conserved among all dinoflagellate species examined to date and show no similarity to counterparts in other organisms [30,45]. Due to being unique to dinoflagellate nuclear-encoded mRNAs and ubiquitous in the dinoflagellate transcriptome, DinoSL could be used as specific hook to separate dinoflagellate nucleus-encoded gene transcripts from those of other organisms, as has been demonstrated for mixed microbial samples [34] and natural samples of aquatic ecosystems [32,33]. Full-length cDNA sequences with complete coding regions, allowing for more accurate gene identification and/or functional prediction, have shown to be informative, particularly for non-model species or field-source resources in which a majority portion of the yielded sequences fail to align well with the known sequences in public databases [46,47]. In this study, we used DinoSL as the selective primer combined with a common 3′-end sequence incorporated in a modified oligo (dT) applied in cDNA synthesis to generate full-length e-cDNA library from marine sediment sample and sequenced using the third-generation sequencing technique. Thus, all the retrieved sHsp genes from transcript pool of the e-cDNA library were considered to be dinoflagellate-sourced and maintained expressing at transcriptional level during dormancy maintenance of dinoflagellate in natural sediment.

Although found in all forms of life [14–17,43], sHsp is the least conserved family of molecular chaperones [15,20,21]. Higher plant sHsps are encoded by nuclear multigene families and are particularly diverse and numerous [12,13]. In sharp contrast to prokaryotes usually containing only one or two sHsps [17,21], some plant species have up to 40 individual sHsps [14]. Detailed studies have revealed that there are 19, 23, and 51 sHsp-encoding genes in genomes of *Arabidopsis*, rice, and soybean, respectively [12,14]. Higher diversification of plant sHsps seems to reflect unique acclimation to stressful conditions that are specific to sessile plants [15,18,43]. However, whether different sHsps have varying substrate specificities is still an open question. Works done in many organisms also demonstrate that the number and diversity of sHsps are not linked with adaptation to extreme conditions [14,19,21]. Although recent years have greatly expanded our understanding of plant sHsps diversity, comprehensive analysis attempting to delineate the composition and expression pattern of dinoflagellate sHsp gene family has not yet been performed. In this study, a series of 13 nuclear dinoflagellate sHsp genes were screened from the full-length e-cDNA library. All of them shared a conserved ACD, flanked by non-conserved N- and C-terminal extensions of variable length (Figure 1). The ACD, the conserved hallmark of sHsp monomer, which is vital for chaperone activity maintenance; in addition, the N-terminus contributes to denatured clients binding and substrate interactions and the C-terminus participates in homo-oligomerization and heat stress granule formation [15,18,19,21]. Multiple alignment revealed considerable amino acid divergence among dinoflagellate sHsp members (Supplementary Table S2), which is in accordance with previous documents of the lower similarity about 30% among algal sHsps and the overall similarity below 50% at protein level among members in this heterogeneous family [15,20,21,43].

Due to lack of high similarity, it is difficult to trace the evolutionary trajectory of sHsp members [15,19,21]. Considering the divergent N- and C-terminal extensions also contributing to preferential and/or functional specificity of sHsps [15,18,19], our phylogeny reconstruction was plotted on the basis of the full amino acid sequences of complete

coding regions. The yielded tree clearly showed that sHsps from one phylum clustered together as a separate branch, but the clustering among different phyla were not in well accordance with their taxonomic relationships (Figure 2). Previous evidence favored the notion that sHsps had already existed in the last common ancestor of prokaryotes and eukaryotes, and thus diverged early in their respective evolutionary histories [13,15]. Therefore, sHsp members are extremely diverse among domains of life because they have gone through independent evolutions in all lineages of life (e.g., bacteria, fungi, higher plants, and metazoans) [13,17,21]. The plant sHsps have very unique evolutionary routes compared to other eukaryotes, in which the gene families have arisen by duplication and divergence [14]. Algae, akin to land plants, were confirmed to include many more sHsps than other eukaryotes do [43]. However, it was interesting to note that none of the algal sHsps were found in the lineage of higher plants in our tree. Therefore, it was apparent that the phylogeny observed from our sHsps tree was not consistent with the phylogenetic affinity of eukaryotic lineages reconstructed with other molecular markers or via traditional taxonomic features, but might reflect evolutionary pathways of sHsps independent from those determining the phylogenetic relationships of different lineages. Evolutionarily, dinoflagellate nuclear genomes are highly dynamic. They have acquired targeted genes via successive horizontal gene transfer from the peridinin plastid, the tertiary replacement plastid and its host nucleus, cyanobacteria, red algae, green algae, haptophyte, and even bacteria, giving rise to a highly chimeric nuclear genome [48,49]. In the phylogenetic tree, an unexpected close relationship was found between our dinoflagellate sHsp sequences and four sHsps from the Volvocales of green algae. It was also unexpected that all these sHsp from dinoflagellates were not clustered like that of Alveolates, a group of protists systematically (i.e., taxonomically) closer to dinoflagellates than green algae. These findings seem to raise the interesting possibility that dinoflagellate sHsps might originate from the endosymbionts of plastids and/or green algae. However, this hypothesis needs to be verified by employing sHsp sequences from more algal species and/or green algal organelle-localized sHsps (whose cellular locations are determined experimentally). Therefore, at this moment, due to the scarce entries of sHsp members in limited algal groups and the lack of clearly identified organelle-localized sHsps in algae, we are not allowed to make any further speculation with certainty.

In addition to the molecular insights as described, it was equally interesting to note the physiological implications of our evidences. Intensive works have established that some sHsps are produced in response to a broad array of hostile environmental insults and their production could reach up to 1% of the total proteins upon thermal stress [13,14,19,22]. Apart from primary chaperone activity, accumulated data continue to imply subtle regulatory and signaling functions of sHsps during development under normal physiological status or non-stressful environmental scenarios [50–52]. In the plant kingdom, seed dormancy is present throughout and has a profound impact on the structure and development of plant communities. The mandatory dormancy of dinoflagellate resting cysts (a period of temporary block of germination even exposure to favorable conditions; [53]) functionally resembles the dormancy of plant seeds [6,9,54]. It has been noticed for over two decades that multiple sHsps mRNA are expressed in dry seeds and disappear from the onset of germination [26,49,52]. Small Hsps were proposed to be correlated with increased seed longevity and germination vigor [14,51]. A recent work on tobacco (*Nicotiana tabacum*) revealed that a mitochondrial sHsp24.7 positively modulated seed germination via temperature-dependent cytochrome production and ROS generation [52]. The sHsp-silenced rice line was detected to delay seed germination compared to the wild-type plant, implying essential roles played by sHsp genes in seed physiology [26]. Many insects undergo a developmental process of seasonal dormancy termed “diapause”, which offers an adaptive advantage to survive adverse environments and allows life cycle synchronization with suitable periods for growth, development, and reproduction [50]. Several studies have shown that insects’ sHsps are associated with the diapause process and their implication during diapause might be quite different amongst different insect species [50,55–58].

In protozoan parasites, the natural life cycle is complex, involving differentiation through various stages of development and transition through hosts [59]. Although the transformation process (from one stage to another) is not entirely clear, stage-specific expressions of sHsps have been determined in many species of *Plasmodium berghei*, *Neospora caninum*, *Toxoplasma gondii*, *Nippostrongylus brasiliensis*, *Leishmania donovani*, *Ostertagia* ostertagi, *Fasciola hepatica*, *Brugia pahangi*, *B. malayi*, *Strongyloides ratti*, *Schistosoma mansoni* [17,59–62], suggesting their productions play roles in the transformation program and promote the acclimation of parasites to new hostile environments [17,59]. The fact that sHsps are found across all domains of life, including those lineages with greatly reduced genomes, implies their particularly early origins and essential functions among organisms of evolutionarily distant groups [43]. The 13 sHsp genes identified from dinoflagellate-specific e-cDNA library in this study were obviously transcribed actively in the resting cyst assemblages of dinoflagellates buried in marine sediment. Since encystment (i.e., formation of resting cysts) of dinoflagellates is generally considered to be an adaptive or survival strategy to adverse environments and the resting cysts buried in natural settings usually face harsh conditions, such as darkness, anoxia, and lower temperature [8,9,54], these sHsps can be reasonably hypothesized to be essentially functioning during the dormancy of resting cysts, possibly in handling the harsh conditions and/or maintaining the homeostasis of cysts. Therefore, this work raised the possibility that the sHsp expression is part of the machinery of maintaining the dormancy or/and the adaptation to ambient conditions of dinoflagellate resting cysts. Our results, although preliminary, gained an important glance on the universe presence of sHsps in dinoflagellates and their active expressions in the assemblage of resting cysts that were buried in the marine sediment. The essentiality of sHsps functioning in resting cysts necessitate more intensive and extensive investigations on all possible functions of Hsps in dinoflagellates, a group of protists with vital ecological and biological importance.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/d13100471/s1, Table S1. Details for the sHsp sequences used in the phylogenetic analysis. Table S2. Amino acid similarities among 14 sHsps with complete coding region from dinoflagellates.

Author Contributions: Conceptualization, Y.D. and Y.Z.T.; Methodology, Y.D. and Z.H.; Software, F.L. and C.Y.; Validation, Y.D., Z.H., and Y.Z.T.; Formal analysis, F.L.; Investigation, F.L. and C.Y.; Resources, Y.D. and F.L.; Data curation, Z.H.; Writing—original draft preparation, Y.D.; Writing—review and editing, Y.D. and Y.Z.T.; Visualization, Y.Z.T.; Supervision, Y.Z.T.; Project administration, Y.Z.T.; Funding acquisition, Y.Z.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Deployment Project of Centre for Ocean Mega-Research of Science, Chinese Academy of Sciences (Grant No. COMS2019Q09) and the National Science Foundation of China (Grant No. 41776125).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article or Supplementary Materials.

Conflicts of Interest: All the authors declare that they have no conflict of interest.

References

1. Taylor, F.J.R.; Hoppenrath, M.; Saldarriaga, J.F. Dinoflagellate diversity and distribution. *Biodivers. Conserv.* **2008**, *17*, 407–418. [CrossRef]
2. Lin, S.; Cheng, S.; Song, B.; Zhong, X.; Lin, X.; Li, W.; Li, L.; Zhang, Y.; Zhang, H.; Ji, Z.; et al. The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* **2015**, *350*, 691–694. [CrossRef] [PubMed]
3. Jeong, H.J.; Kang, H.C.; Lim, A.S.; Jang, S.H.; Lee, K.; Lee, S.Y.; Ok, J.H.; You, J.H.; Kim, J.H.; Lee, K.H.; et al. Feeding diverse prey as an excellent strategy of mixotrophic dinoflagellates for global dominance. *Sci. Adv.* **2021**, *7*, eabe4214. [CrossRef] [PubMed]

4. Hallegraeff, G.M.; Anderson, D.M.; Belin, C.; Bottein, M.-Y.D.; Bresnan, E.; Chinain, M.; Enevoldsen, H.; Iwataki, M.; Karlson, B.; McKenzie, C.H.; et al. Perceived global increase in algal blooms is attributable to intensified monitoring and emerging bloom impacts. *Commun. Earth Environ.* **2021**, *2*, 1–10. [CrossRef]
5. Von Dassow, P.; Montresor, M. Unveiling the mysteries of phytoplankton life cycles: Patterns and opportunities behind complexity. *J. Plankton Res.* **2011**, *33*, 3–12. [CrossRef]
6. Ellegaard, M.; Ribeiro, S. The long-term persistence of phytoplankton resting stages in aquatic 'seed banks'. *Biol. Rev.* **2018**, *93*, 166–183. [CrossRef]
7. Delebecq, G.; Schmidt, S.; Ehrhold, A.; Latimier, M.; Siano, R. Revival of ancient marine dinoflagellates using molecular biostimulatio. *J. Phycol.* **2020**, *56*, 1077–1089. [CrossRef]
8. Bravo, I.; Figueroa, R.I. Towards an ecological understanding of dinoflagellate cyst functions. *Microorganisms* **2014**, *2*, 11–32. [CrossRef]
9. Tang, Y.Z.; Hu, Z.X.; Deng, Y.Y. Characteristical life history (resting cyst) provides a mechanism for recurrence and geographic expansion of harmful algal blooms of dinoflagellates: A Review. *Studia Mar. Sin.* **2016**, *51*, 132–154.
10. Tang, Y.Z.; Gu, H.; Wang, Z.; Liu, D.; Wang, Y.; Lu, D.; Hu, Z.; Deng, Y.; Shang, L.; Qi, Y. Exploration of resting cysts (stages) and their relevance for possibly HABs-causing species in China. *Harmful Algae* **2021**, *107*, 102050. [CrossRef]
11. Feder, M.E.; Hofmann, G.E. Heat-shock proteins, molecular chaperones, and the stress response: Evolutionary and ecological physiology. *Annu. Rev. Physiol.* **1999**, *61*, 243–282. [CrossRef]
12. Al-Whaibi, M.H. Plant heat-shock proteins: A mini review. *J. King Saud Univ. Sci.* **2011**, *23*, 139–150. [CrossRef]
13. Tanguay, R.M.; Hightower, L.E. *The Big Book on Small Heat Shock Proteins*; Springer International Publishing: New York, NY, USA, 2015.
14. Waters, E.R.; Vierling, E. Plant small heat shock proteins–evolutionary and functional diversity. *New Phytol.* **2020**, *227*, 24–37. [CrossRef]
15. De Jong, W.W.; Caspers, G.-J.; Leunissen, J.A. Genealogy of the α-crystallin—small heat-shock protein superfamily. *Int. J. Biol. Macromol.* **1998**, *22*, 151–162. [CrossRef]
16. Maaroufi, H.; Tanguay, R.M. Analysis and phylogeny of small heat shock proteins from marine viruses and their cyanobacteria host. *PLoS ONE* **2013**, *8*, e81207. [CrossRef]
17. Pérez-Morales, D.; Espinoza, B. The role of small heat shock proteins in parasites. *Cell Stress Chaperones* **2015**, *20*, 767–780. [CrossRef]
18. Haslbeck, M.; Vierling, E. A first line of stress defense: Small heat shock proteins and their function in protein homeostasis. *J. Mol. Biol.* **2015**, *427*, 1537–1548. [CrossRef]
19. Carra, S.; Alberti, S.; Arrigo, P.A.; Benesch, J.L.; Benjamin, I.J.; Boelens, W.; Bartelt-Kirbach, B.; Brundel, B.J.J.M.; Buchner, J.; Bukau, B.; et al. The growing world of small heat shock proteins: From structure to functions. *Cell Stress Chaperones* **2017**, *22*, 601–611. [CrossRef]
20. Vierling, R.A.; Nguyen, H.T. Heat-shock protein gene expression in diploid wheat genotypes differing in thermal tolerance. *Crop Sci.* **1992**, *32*, 370–377. [CrossRef]
21. Strauch, A.; Haslbeck, M. The function of small heat-shock proteins and their implication in proteostasis. *Essays Biochem.* **2016**, *60*, 163–172. [CrossRef]
22. Deng, Y.; Hu, Z.; Shang, L.; Chai, Z.; Tang, Y.Z. Transcriptional responses of the heat shock protein 20 (Hsp20) and 40 (Hsp40) genes to temperature stress and alteration of life cycle stages in the harmful alga *Scrippsiella trochoidea* (Dinophyceae). *Biology* **2020**, *9*, 408. [CrossRef]
23. Levin, R.A.; Beltran, V.H.; Hill, R.; Kjelleberg, S.; McDougald, D.; Steinberg, P.D.; van Oppen, M. Sex, scavengers, and chaperones: Transcriptome secrets of *Divergentsymbiodiniumthermal* tolerances. *Mol. Biol. Evol.* **2016**, *33*, 2201–2215. [CrossRef]
24. Gierz, S.L.; Forêt, S.; Leggat, W. Transcriptomic analysis of thermally stressed symbiodinium reveals differential expression of stress and metabolism genes. *Front. Plant Sci.* **2017**, *8*, 271. [CrossRef]
25. Lei, Q.-Y.; Lü, S.-H. Molecular ecological responses of dinoflagellate, *Karenia mikimotoi* to environmental nitrate stress. *Mar. Pollut. Bull.* **2011**, *62*, 2692–2699. [CrossRef]
26. Sarkar, N.K.; Kotak, S.; Agarwal, M.; Kim, Y.-K.; Grover, A. Silencing of class I small heat shock proteins affects seed-related attributes and thermotolerance in rice seedlings. *Planta* **2019**, *251*, 26. [CrossRef]
27. Deng, Y.; Hu, Z.; Shang, L.; Peng, Q.; Tang, Y.Z. Transcriptomic Analyses of *Scrippsiella trochoidea* reveals processes regulating encystment and dormancy in the life cycle of a dinoflagellate, with a particular attention to the role of abscisic acid. *Front. Microbiol.* **2017**, *8*, 2450. [CrossRef]
28. Deng, Y.; Hu, Z.; Zhan, Z.; Ma, Z.; Tang, Y. Differential expressions of an Hsp70 gene in the dinoflagellate Akashiwo sanguinea in response to temperature stress and transition of life cycle and its implications. *Harmful Algae* **2015**, *50*, 57–64. [CrossRef]
29. Deng, Y.; Hu, Z.; Chai, Z.; Tang, Y.Z. Molecular cloning of heat shock protein 60 (Hsp60) and 10 (Hsp10) genes from the cosmopolitan and harmful dinoflagellate *Scrippsiella trochoidea* and their differential transcriptions responding to temperature stress and alteration of life cycle. *Mar. Biol.* **2018**, *166*, 7. [CrossRef]
30. Zhang, H.; Hou, Y.; Miranda, L.; Campbell, D.A.; Sturm, N.R.; Gaasterland, T.; Lin, S. Spliced leader RNA trans-splicing in dinoflagellates. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4618–4623. [CrossRef]

31. Bolch, C.J.S. The use of sodium polytungstate for the separation and concentration of living dinoflagellate cysts from marine sediments. *Phycologia* **1997**, *36*, 472–478. [CrossRef]
32. Lin, S.; Zhang, H.; Zhuang, Y.; Tran, B.; Gill, J. Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20033–20038. [CrossRef] [PubMed]
33. Zhuang, Y.; Zhang, H.; Hannick, L.; Lin, S. Metatranscriptome profiling reveals versatile N-nutrient utilization, CO_2 limitation, oxidative stress, and active toxin production in an *Alexandrium fundyense* bloom. *Harmful Algae* **2015**, *42*, 60–70. [CrossRef]
34. Zhang, H.; Zhuang, Y.; Gill, J.; Lin, S. Proof that dinoflagellate spliced leader (DinoSL) is a useful hook for fishing dinoflagellate transcripts from mixed microbial samples: *Symbiodinium kawagutii* as a case study. *Protist* **2013**, *164*, 510–527. [CrossRef] [PubMed]
35. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef]
36. Rombel, I.T.; Sykes, K.F.; Rayner, S.; Johnston, S.A. ORF-FINDER: A vector for high-throughput gene identification. *Gene* **2002**, *282*, 33–41. [CrossRef]
37. Letunic, I.; Doerks, T.; Bork, P. SMART 7: Recent updates to the proteindomain annotation resource. *Nucleic Acids Res.* **2012**, *40*, D302–D305. [CrossRef]
38. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2020**, *49*, D412–D419. [CrossRef]
39. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2019**, *20*, 1160–1166. [CrossRef]
40. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. *The Proteomics Protocols Handbook*; Humana Press: Totowa, NJ, USA, 2005.
41. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
42. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [CrossRef]
43. Waters, E.R.; Rioflorido, I. Evolutionary analysis of the small heat shock proteins in five complete algal genomes. *J. Mol. Evol.* **2007**, *65*, 162–174. [CrossRef]
44. Hastings, K.E. SL trans-splicing: Easy come or easy go? *Trends Genet.* **2005**, *21*, 240–247. [CrossRef]
45. Lidie, K.B.; Van Dolah, F.M. Spliced leader RNA-mediated trans-splicing in a dinoflagellate, Karenia brevis. *J. Eukaryot. Microbiol.* **2007**, *54*, 427–435. [CrossRef]
46. Islas-Flores, T.; Galán-Vásquez, E.; Villanueva, M. Screening a Spliced Leader-Based *Symbiodinium microadriaticum* cDNA library using the yeast-two hybrid system reveals a hemerythrin-like protein as a putative SmicRACK1 ligand. *Microorganisms* **2021**, *9*, 791. [CrossRef]
47. Yang, F.; Xu, N.; Zhuang, Y.; Yi, X.; Huang, Y.; Chen, H.; Lin, S.; Campbell, D.A.; Sturm, N.R.; Liu, G.; et al. Spliced leader RNA trans-splicing discovered in copepods. *Sci. Rep.* **2015**, *5*, 17411. [CrossRef]
48. Lin, S. Genomic understanding of dinoflagellates. *Res. Microbiol.* **2011**, *162*, 551–569. [CrossRef]
49. Yoon, H.S.; Hackett, J.D.; Van Dolah, F.M.; Nosenko, T.; Lidie, K.L.; Bhattacharya, D. Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol. Biol. Evol.* **2005**, *22*, 1299–1308. [CrossRef]
50. Macrae, T.H. Gene expression, metabolic regulation and stress tolerance during diapause. *Cell Mol. Life Sci.* **2010**, *67*, 2405–2424. [CrossRef]
51. Dirk, L.M.; Downie, A.B. An examination of Job's rule: Protection and repair of the proteins of the translational apparatus in seeds. *Seed Sci. Res.* **2018**, *28*, 168–181. [CrossRef]
52. Ma, W.; Guan, X.; Li, J.; Pan, R.; Wang, L.; Liu, F.; Ma, H.; Zhu, S.; Hu, J.; Ruan, Y.-L.; et al. Mitochondrial small heat shock protein mediates seed germination via thermal sensing. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4716–4721. [CrossRef]
53. Anderson, D.; Morel, F. The seeding of two red tide blooms by the germination of benthic *Gonyaulax tamarensis hypnocysts*. *Estuar. Coast. Mar. Sci.* **1979**, *8*, 279–293. [CrossRef]
54. Li, F.; Yang, A.; Hu, Z.; Lin, S.; Deng, Y.; Tang, Y. Probing the Energetic metabolism of resting cysts under different conditions from molecular and physiological perspectives in the harmful algal blooms-forming dinoflagellate *Scrippsiella trochoidea*. *Int. J. Mol. Sci.* **2021**, *22*, 7325. [CrossRef]
55. Rinehart, J.P.; Li, A.; Yocum, G.D.; Robich, R.M.; Hayward, S.A.L.; Denlinger, D.L. Up-regulation of heat shock proteins is essential for cold survival during insect diapause. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 11130–11137. [CrossRef]
56. Gkouvitsas, T.; Kontogiannatos, D.; Kourti, A. Differential expression of two small Hsps during diapause in the corn stalk borer *Sesamia nonagrioides* (Lef.). *J. Insect Physiol.* **2008**, *54*, 1503–1510. [CrossRef]
57. Lu, Y.-X.; Xu, W.-H. Proteomic and Phosphoproteomic analysis at diapause initiation in the Cotton Bollworm, *Helicoverpa armigera*. *J. Proteome Res.* **2010**, *9*, 5053–5064. [CrossRef]
58. Zhao, J.; Huang, Q.; Zhang, G.; Zhu-Salzman, K.; Cheng, W. Characterization of two small heat shock protein genes (*Hsp17.4* and *Hs20.3*) from *Sitodiplosis mosellana*, and their expression regulation during diapause. *Insects* **2021**, *12*, 119. [CrossRef]
59. Maresca, B.; Carratù, L. The biology of the heat shock response in parasites. *Parasitol. Today* **1992**, *8*, 260–266. [CrossRef]

60. Tweedie, S.; Grigg, M.E.; Ingram, L.; Murray, E.; Selkirk, M. The expression of a small heat shock protein homologue is developmentally regulated in *Nippostrongylus brasiliensis*. *Mol. Biochem. Parasitol.* **1993**, *61*, 149–153. [CrossRef]
61. Kobayashi, T.; Narabu, S.; Yanai, Y.; Hatano, Y.; Ito, A.; Imai, S.; Ike, K. Gene cloning and characterization of the protein encoded by *Theneospora Caninumbradyzoite*-specific antigen genebag. *J. Parasitol.* **2013**, *99*, 453–458. [CrossRef]
62. De Miguel, N.; Braun, N.; Bepperling, A.; Kriehuber, T.; Kastenmüller, A.; Buchner, J.; Angel, S.O.; Haslbeck, M. Structural and functional diversity in the family of small heat shock proteins from the parasite Toxoplasma gondii. *Biochim. Biophys. Acta* **2009**, *1793*, 1738–1748. [CrossRef]