

The red seaweed *Asparagopsis taxiformis* genome and integrative -omics analysis

M Zhao (✉ mzhao@usc.edu.au)

University of the Sunshine Coast

AH Campbell

University of the Sunshine Coast

Patwary ZP

University of the Sunshine Coast

T Wang

University of the Sunshine Coast

T Lang

University of the Sunshine Coast

J Webb

University of the Sunshine Coast

Zuccarello GC

Victoria University of Wellington

A Wegner

University of the Sunshine Coast

D Heyne

University of the Sunshine Coast

L McKinnie

University of the Sunshine Coast

C Pascelli

University of the Sunshine Coast

N Satoh

Okinawa Institute of Science and Technology Graduate University

E Shoguchi

Okinawa Institute of Science and Technology Graduate University

Paul NA

University of the Sunshine Coast

Cummins SF

University of the Sunshine Coast

Keywords: Asparagopsis taxiformis, whole genome, peroxidase transcriptome, multi-omics, database

Posted Date: November 9th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2232367/v1>

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

[Read Full License](#)

Abstract

Seaweeds (marine macroalgae) are crucial to the functioning of healthy coastal ecosystems and global biogeochemical cycles, and sometimes provide novel solutions to help mitigate climate change. The red seaweed *Asparagopsis taxiformis* (Bonnemaisoniaceae, Rhodophyta) produces bioactive natural products that, when fed to cattle and sheep, can eradicate methane emissions from these livestock. However, in order to cultivate enough *A. taxiformis* to have a meaningful impact on global greenhouse gas emissions, we need to improve our understanding of the biology of this new crop. In this study, we used a domesticated diploid sporophyte (> 1.5 years in culture, with relatively low microbial diversity) to establish a high-quality draft nuclear genome for *A. taxiformis* from Queensland, Australia. The *A. taxiformis* lineage was confirmed as Lineage 6 (L6) based upon phylogenetic analysis (Cox2-3 spacer). The genome of *A. taxiformis* (L6) was 142 Mb in size with approximately 11,000 protein-coding genes, including those associated with secondary metabolism, photosynthesis and defence, and the assembly contained 70.67% repeat regions. Based on protein domain analysis, the most prominent lineage-specific duplications belonged to those containing WD repeat proteins, as well as bestrophin and N6_N4_Mtase domain proteins. Cultured (domesticated) *A. taxiformis* (L6) sporophytes contained 4-times more bromoform (the key anti-methanogenic natural product) compared to wild sporophytes. To obtain information regarding associated molecular differences, the genome was used as a reference to explore differential gene expression related to environment. Cultured sporophytes demonstrated an enrichment of regulatory factors (kinases, transcription factors), whereas wild sporophytes were enriched with defence and stress-related genes, including those involved in protein folding (heat shock proteins) and halogenated metabolite production. Wild sporophytes also expressed a relatively high level of novel secreted proteins, with similarity to collagen-alpha proteins (termed rhodophyte collagen-alpha-like proteins, RCAPs). Proteomic investigation of the genome of cultured sporophytes, resulting in the identification of over 400 proteins, including RCAPs, as well as numerous enzymes and phycobiliproteins, which will facilitate future functional characterisation. In summary, as the most comprehensive genomic resource for any *Asparagopsis* species, this resource provides a gateway for seaweed researchers to fast-track the development and production of *Asparagopsis* to meet demand by agriculture and do so with economic and environmental agility.

Introduction

Our ability to identify and harness the full potential of biodiversity to meet the environmental challenges of agriculture is critical for future food security and mitigating climate change [1, 2]. This notion is particularly pertinent for ruminant livestock, including cattle and sheep, that collectively belch the methane produced by their gut microbes that is responsible for more than 6% of global greenhouse gas emissions [3]. Methane has a global warming potential 28-times that of carbon dioxide over 100 years and, in Australia, around 15% of total greenhouse gas emissions come from agriculture, of which two-thirds is methane from ruminants [4]. Therefore, reducing methane emissions from agriculture represents one of the most significant challenges for the sustainability of extant agricultural systems.

Recently, Australian researchers investigated the use of seaweed (marine macroalgae) as feed additives for cattle and sheep and discovered that one species of tropical red seaweed, *Asparagopsis taxiformis* (hereafter '*Asparagopsis*'), knocks out enteric methanogenesis in cattle rumen fluid when added in dried form at 2% of the organic matter [5]. This effect was almost exclusively related to one halogenated natural product known as bromoform or tribromomethane [6]. The research was based on the natural product knowledge of the genus *Asparagopsis* [7] and its evaluation against other marine and freshwater macroalgae [5]. All seaweeds and microalgae release *de novo* bromoform as part of their primary metabolism [12]. However, *Asparagopsis* is unique in that it also stores bromoform in specialised gland cells and the dry weight content in the seaweed can reach up to 4% [7]. This natural source of bromoform is so concentrated that the glands are a prominent feature of the cells under light microscopy. No other seaweed or plant localises or accumulates bromoform in its tissue to this extent.

With regulations against the manufacture, import and export of some of the most effective synthetic additives for mitigating methane emissions [8], industry and researchers have instead heavily invested in finding natural alternatives [9]. Many seaweeds other than species of *Asparagopsis*, and more specifically the natural products contained within them, have antimicrobial activity [10–12]. Moreover, while some seaweed additives can alter the ruminant microbiome [13] and others can also reduce methane [5, 14], none are comparable to *Asparagopsis* with activity at < 2% feed dry matter leading to undetectable methane production [6]. *Asparagopsis* is the only seaweed that can be used as a high-efficacy feed additive for ruminants because it specifically targets the archaea portion of the gut microbiome without impacting fermentation [6].

International interest in the use of a natural feed additive in feedlots, including in New Zealand, California and Europe, has shifted the question from “what additive?” to “how can we produce *Asparagopsis* at scale?”. There needs to be significant investment in infrastructure and knowledge to meet the global demand for this seaweed feed additive and to retain the efficacy of the biomass under mass production. Controlled, large-scale aquaculture is the only means to deliver this consistent quality and quantity of product (i.e., a minimum content of bromoform and free from impurities). However, we know that the content of the halogenated natural products in the environment is variable, with many direct and indirect effects at play, including biosynthetic pathway variation [15], genetic and environmental components [16], through to ecological interactions with microbes [10, 17]. For seaweeds such as *Asparagopsis* with alternation of generations, i.e., free-living diploid (sporophyte) and haploid (gametophyte) stages, there are also differences across the life cycle [18]. The biological and ecological variation in natural products needs to be understood and addressed before companies can sensibly invest in commercialising production of *Asparagopsis*.

By understanding the biosynthetic pathways of bromoform and other halogenated natural products in *Asparagopsis*, and evaluating differential gene expression across the life cycle, we can gain critical information regarding the biology and ecology of the production of bromoform in the seaweed under different culture conditions and environments. This information is also critical for fast-tracking the development of this new aquaculture species and for setting the parameters for what is possible within

this biological system and guiding investment in research and development. The availability of a high-quality genome can provide the foundation resource to support such multi-level investigations [12], the benefits of which have been demonstrated in other red seaweeds including *Chondrus crispus*, *Gracilariopsis chorda*, *Porphyra umbilicalis* and *Pyropia haitanensis* [2, 12–14]. For *Asparagopsis*, genomic sequencing has already been useful for the elucidation of a bromoform biosynthesis locus [13], although limited genomic features were reported in this instance. Seaweed genomes and transcriptomes are also resources to facilitate high-throughput proteomics investigations with applied or environmental outcomes. For example, a transcriptomic-derived protein database in the red seaweed *Eucheuma denticulatum* enabled the proteomic characterisation of extracellular proteins following soluble protein extraction [19]. In the red seaweed *Neoporphyra haitanensis*, a genome-derived protein database was used to assess the proteins associated with molecular responses to environmental stresses [20].

In this study, we present an *Asparagopsis taxiformis* (Lineage 6) genome, sequenced from a cultured sporophyte that was collected and domesticated from Queensland, Australia, where the first demonstration of the anti-methanogenic effects of the seaweed were conducted. We report the genome features of *A. taxiformis* in comparison to other red seaweed genomes, which may assist researchers in fundamental and applied investigations related to climate change mitigation and also agricultural and aquaculture production. The usefulness of this resource was supported by our own comparative RNA-seq analysis of cultured and wild sporophytes, as well as a proteomics analysis, confirming the presence of a novel family of secreted proteins.

Materials And Methods

Asparagopsis taxiformis tissue collection and analysis of microbial contamination

The wild seaweeds were collected between February and April 2021 from rocky reefs at Moffat Beach, Queensland, Australia (26°47'21.0"S 153°08'37.5"E) between 1–3 m depth. Samples were collected with nitrile disposable gloves, rinsed with sterile seawater (SSW), placed in sterile centrifuge tubes, snap-frozen in a Taylor-Wharton CP300 Cryogenic Liquid Nitrogen Shipper and kept in – 80°C prior to DNA extractions. Additionally, a fresh portion of the samples were fixed in 4% paraformaldehyde in SSW for microscopic analysis. The microbial community were compared between *A. taxiformis* sporophyte individuals sampled from the culture and wild.

The cultured individuals were grown vegetatively from sporophytes collected at Moffat Beach (July 30th 2019) and maintained indoors (16:8 h light:dark cycle, 22°C) in a 2 L flask with fortnightly seawater exchange, and addition of F/8 enrichment medium. Healthy filaments were isolated at a small scale (5–30 ml Petri dishes) and maintained with regular SSW exchanges (addition of F/8) until April 2020. At this point they were upscaled to 280 ml inverted culture vessels ('hatchers') at 1 g/L. Once upscaled to hatchers, cultures were maintained with weekly water exchanges (filtered SSW, autoclaved), with addition of F/4 (half the concentration of F2 enrichment medium) and 448 µg/L of germanium dioxide (GeO) to prevent diatom growth. Cultured sporophytes were then sub-sampled for microbial analysis in February

2021 prior to processing for genomic analysis. Sub-samples were snap-frozen in liquid nitrogen prior to microbial DNA extractions, then fixation of fresh fragments in 4% paraformaldehyde in SSW prior to microscopic analysis.

In addition to the molecular analyses of the microbial communities, confocal microscopy was performed to visualize and compare the epiphytic community associated with wild and cultured individuals to provide further insights on the selection of sporophyte cultures for genomic analysis. The fixed filaments of sporophytes were stained with 3 mM DAPI (4',6-diamidino-2-phenylindole) (Thermo Fisher Scientific) and 1 mM SYTO 9 Green Fluorescent Nucleic Acid Stain (Thermo Fisher Scientific) by immersing the filaments for 10 min in each of the stain solutions, prepared according to the manufacturer's instructions. Samples were analyzed on a Nikon C2 TI Eclipse Microscope equipped with confocal laser scanning and CCD camera.

Amplicon sequencing analysis was further performed on *A. taxiformis* individuals to compare microbial richness associated with cultured and wild sporophytes. This is part of a more comprehensive research on the *Asparagopsis* microbiome underway at the University of the Sunshine Coast (Pascelli C, personal communication). We confirmed that the 2019 culture was of relatively low microbial content. In brief, total DNA was extracted using DNeasy PowerSoil Pro Kit (Qiagen) according to manufactures' instructions. The 16S rRNA V4 amplicon were sequenced using Miseq kit v2 250 bp fragments paired-end NANO sequencing (Illumina) at the Ramaciotti Centre for Genomics (Australia). Quality filtering, data denoising and Operational Taxonomic Units (OTUs) clustering were based on an open-source Galaxy-based workflow of 16S Microbial Analysis with Mothur [21]. Statistical analyses were carried out on RStudio.

Library construction and sequencing

An *Asparagopsis taxiformis* cultured sporophyte (2019 winter collection) was selected for genome sequencing based on minimal microbial contamination (see section above). This sample of clonally propagated filaments was subsequently analysed using DNA barcoding with the Cox2-3 spacer to confirm the cultured sample was *A. taxiformis* Lineage 6 (L6), GBR strain [22], as were additional cultures collected and domesticated from the same site in 2018 and 2021. DNA extraction, amplification and phylogenetic analysis followed procedures described previously [22, 23].

For genomic DNA sample preparation, ~ 20 g fresh weight of sporophyte was harvested from hatcher vessels (28/01/2021) and snap frozen with liquid nitrogen. An approximately 5 g sample was shipped in dry ice to BGI, Hong Kong. A hybrid sequencing strategy was used for sequencing the genome, by combining both long reads from PacBio sequencing and short reads from BGI DNBSeg sequencing. Isolated genomic DNA was subjected to short-read library preparation with standard paired-end protocols. Briefly, 1 µg genomic DNA was randomly fragmented by Covaris, followed by fragment selection by Agencourt AMPure XP-Medium kit to an average size of 300–400 base pairs (bp). Selected fragments were repaired and 3'adenylated, then the adaptors were ligated to the ends of these 3'adenylated

fragments. The products were amplified by PCR and purified by the Agencourt AMPure XP-Medium kit. The purified double stranded PCR products were heat denatured to single strand, and then circularized by the splint oligo sequence. The single strand circle DNA (ssCir DNA) were formatted as the final library and qualified by quality-control. Subsequently, libraries with insert lengths of 370 bp were constructed and sequenced by MGISEQ-2000. ssCir DNA molecule formed a DNA nanoball (DNB) containing more than 300 copies through rolling-cycle replication. The DNBs were loaded into the patterned nanoarray by using high density DNA nanochip technology. Finally, 150 bp paired-end reads were obtained by combinatorial Probe-Anchor Synthesis.

To sequence longer reads, high molecular weight genomic DNA was sheared into fragments of approximately 20 kb using a Covaris g-tube (part No.520079; KBiosciences). The preparation and sequencing of the SMRTbell DNA library were performed according to the manufacturer's protocols (Pacific Biosciences, CA, USA). The SMRTbell library with 20 kb insert size was generated with the BluePippin system (Sage Scientific, MA, USA). One lane was loaded and run on the Sequel System, which utilizes single molecule, real-time (SMRT) sequencing with fluorescently labeled nucleotides.

Genome assembly and functional evaluation

The DNBSeg generated about 4.6 Gb data with 31,146,070 raw reads. By filtering out low-quality reads, unknown nucleotides, adapter, and duplicated reads, we harvested a total of 29,530,652 clean reads with a GC content about 44%. For pair-end reads, the k-mer frequencies ($k = 17$) were computed based on Jellyfish v2.2.10 (Marçais and Kingsford, 2011). Then we estimated the genome size based on the 17-mer depth distribution. In total, PacBio sequencing generated 11.8 Gb data with 1,137,468 polymerase reads. With > 80X PacBio data, PacBio reads were firstly corrected and assembled using Canu (v1.7.1) [24]. Contigs were then polished using Pilon based on the short reads from DNBSeg [25]. The raw data of all sequencing was submitted to NCBI Bioproject and SRA database under a Bioproject accession number PRJNA809757. By using sequence-based BLAST search, we compared a published *A. taxiformis* mitochondria genome (NCBI accession number KJ398158.1) with our assembly. The completed mitochondrial genome and the genes were predicted. By running the same procedure on a complete *A. taxiformis* plastid genome (NCBI accession number NC_031148.1), we identified the putative plastid genes in our assembly.

Based on the assembly, overall statistics were calculated for GC content and median contig length. A Benchmarking Universal Single-Copy Orthologs (BUSCO 5.2.2) analysis was then performed to evaluate the accuracy and completeness of our assembly [26]. In general, BUSCO focused on core genes based on evolutionarily informed expectations of gene content of those conserved sequences. BUSCO relied on the OrthoDB database, which was used to construct single-copy gene sets of several large evolutionary branches. In this way, BUSCO metric (% complete genes) could be functionally complementary to technical metrics, such as median contig length N50.

Repeat masking and transposable elements

Before we explored the functional fragments in the genome, we examined the repetitive regions present in the assembly. By using RepeatModeler2 [27], a *de novo* repeat library was built. The repeat library, including known self-mobilizing transposable elements from other species, was then subjected to RepeatMasker (v 4.1.2-p1) [28] to find and mask homologous repeats in the assembled genome using RMBLAST (v 2.10.0+) as the default search engine. The total length of repetitive sequences was 100,684,772 bp, accounting for 70.67% of the assembled genome.

Analysis of microbial DNA contamination

Even with the pre-checks for microbial contamination, there was still potential for microbial contamination as the cultures were not axenic. The residual microbial contamination (bacteria, archaea and virus) was therefore partitioned from the seaweed genome. The Kraken software in the Galaxy platform [29] was used to assign taxonomic labels to the *Asparagopsis* DNA by using a k-mer based approach to compare the reads to the MiniKraken database, constructed from complete bacterial, archaeal, and viral genomes in RefSeq (**Fig. S1** for parameters). The classified reads were translated to the full representation of NCBI taxonomy by using the convert Kraken tool and its output was visualized on the Krona tool.

Gene prediction and functional annotation

Since the BUSCO output is a set of highly conserved genes in our assembly that are single copy in most eukaryotic species, these sequences were used as training dataset for the gene prediction tool AUGUSTUS [30]. Briefly, AUGUSTUS (3.4.0) used the summarized features, such as the length distribution of single exons and the distribution of the number of exons per gene, to optimize the gene prediction. By inputting the assembly after repeat masking, we are able to predict the protein-coding regions [30]. In total, the AUGUSTUS prediction obtained 10,867 genes. Based on the predicted gene sequences and the translated protein sequences, we then ran the BLAST2GO pipeline to annotate the gene functions [31]. In brief, BLASTP was used to search the NCBI non-redundant protein databases [32] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [33] with a threshold of E-value less than $1E^{-5}$. The top five BLAST alignment results were used to extract the functional annotations. In addition to BLAST, InterProScan were also utilized to search all the possible functional domains in the 10,867 proteins. To annotate the protein family information, HMMer was used to search all the Pfam protein family profiles [34].

Comparative genome analysis

To explore the common and unique gene function in our assembly, we first downloaded complete genomes and proteomes of seven red algae from NCBI (**File S1a**). A comparative proteome analysis was then performed using OrthoFinder2. This provided a genome-wide comparison by grouping orthologous proteins. In brief, eight proteome sequences were used as input for OrthoFinder2 [35] to predict the orthologous relationship. The output from OrthoFinder2 include the phylogenetic trees for thousands of detected orthologous groups. In OrthoFinder, these sequences were concatenated into a supergene and used for multiple sequence alignment by MAFFT [36]. Then, the consensus phylogenetic information

from the orthologous groups was inferred by STAG algorithm [37]. The final unrooted species tree was built by FastTree using the maximum likelihood method [38]. Based on the species tree, the gene duplication events that were satisfied by the Duplication-Loss-Coalescent model were extracted at each branch of the species tree. To obtain the basic functional information, we ran HMMer and extracted all the Pfam domain information for all the seven proteomes [34].

For data analysis and visualization, all statistics analyses were generated by R language. The basic bar plots were generated using the Excel chart function. The overlapping relationship of Pfam annotations and the heatmaps of the comparative genomics analyses were generated using in-house R scripts.

Targeted gene annotation

To predict transcription factors, we downloaded all known transcription factors from plant genomes in PlantTFDB version 4 [39] and then mapped our proteome to these. Finally, all the conserved transcription factor families were assigned based on the best match from BLAST output. Genome annotations were then used to target investigation of genes associated with: (1) cell structure and regulation, (2) defence-related and stress genes, and (3) secreted proteins. The identification of genes in these groups was performed by using homologs identified in *C. crispus* [40] as BLASTp searches in the *A. taxiformis* (L6) genome (e-value cut-off 10^{-3}). The *mbb* locus (involved in halogenated natural product synthesis) was identified through BLASTp for each *mbb* gene (GenBank MN966723). ExPASy Translate tool was used to assess presence of open reading frames and introns across the *mbb* locus. The SignalP (Version 5.0) [41] and TMHMM (Version 2.0c) [42] were used to predict secreted peptides and proteins with a transmembrane domain. Multiple sequence alignments of the protein and nucleotide sequences were generated using Clustal Omega [43] in a ClustalW output format. NCBI Conserved Domains [44] were used to locate the domain and active site residues. Clustal Omega was utilised to determine nucleotide identity. Predicted structure homology modelling was performed using the SWISS-MODEL workspace [45].

Cultured and wild sporophyte samples for comparative bromoform analysis and RNA-seq gene quantification

Cultured sporophyte (filamentous, non-reproductive) of *A. taxiformis* (L6) was collected as described above. Wild sporophytes (filamentous, non-reproductive) of *A. taxiformis* (L6) were also collected from Moffat Beach, Queensland, from the sub-tidal from ~ 0.5-3 m depth (July 2020) for comparisons to the cultured strains. Wild sporophyte samples were transported in seawater to the laboratory of the University of the Sunshine Coast. Within 1 h, the samples were cleaned with sterile seawater to remove visible mud, epiphytes and epibionts. All samples were frozen (in 1.5 ml tubes) in liquid nitrogen before storage at -80°C until use.

Gas chromatography-mass spectrometry (GC-MS) quantification of bromoform

Bromoform was extracted from cultured and wild sporophyte ($n = 3$) by solid-liquid extraction of 0.02 g of dried, milled biomass into 4 mL of methanol containing naphthalene internal standard, using a method previously described [46]. Samples were stored for 3 days at 4°C, prior to filtration and GC-MS analysis. Chromatography was performed on a Perkin Elmer Clarus 580 GC tandem Perkin Elmer Clarus SQ8S MS using a DB-5 column (Perkin Elmer Elite-5MS, 30.0 m x 0.25 mm x 0.25 μ m). Injections (1.0 μ l) were introduced to a 280°C injection port with a 50:1 split ratio. The gas chromatography parameters were as follows: 40°C for 1 min, increased to 250 °C at a rate of 20 °C/min, immediately followed by a 30 s equilibration time. The carrier gas used was helium, with a flow rate of 1 mL/min. The mass spectrometer analysed a mass range from 50 to 340 (m/z), from 3.0 to 12.0 min at 70 eV. Quantitation was conducted by referencing the TIC integrated by TurboMass software against analytical standard calibration curves. Final bromoform concentration did not account for the moisture content of biomass (typically ~ 3–5% under laboratory conditions).

Genome-guided RNA-seq quantification for cultured and wild sporophyte

Total RNA was extracted from the experimental samples (3 cultured and 5 wild), using a RNeasy Plant Mini Kit (QIAGEN, Germany) according to the manufacturer's protocol with the following specifics: 82 ± 8 mg of tissue was ground in liquid nitrogen before further lysis using Buffer RLT with added β -mercaptoethanol (1%). Extracted RNA was treated with a DNA-free™ DNA Removal Kit (Invitrogen, USA) to reduce genomic DNA contamination for cultured sporophytes. RNA quality was measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific, USA) and the concentration and integrity verified using an Agilent RNA 6000 Nano assay run on an Agilent 2100 Bioanalyzer (Agilent, USA).

RNA samples ($n = 8$) were either sent on dry ice or stabilised for transport at ambient temperature by drying in GenTegra RNA tubes (GenTegra, USA) on a SpeedVac concentrator (Thermo Fisher Scientific, USA), according to the manufacturer's protocol. Whole RNA-seq was performed by Novogene (Singapore) using Next-Generation Illumina sequencing, according to the standardised workflow. This process involved construction of cDNA libraries, before sequencing of cDNA fragments using an Illumina NovaSeq 6000 System (raw sequence reads have been deposited in the NCBI GenBank PRJNA809757). Differential gene expression between cultured and wild sporophyte RNA-seq data was mapped against the *A. taxiformis* (L6) genome using CLC Genomics Workbench (Ver. 21.0.3) with default parameters. Transcript abundances were converted to transcripts per million (TPM) for each gene. Genes with significant upregulation or downregulation were identified using a significance cut-off set at an FDR P-value of < 0.05 . Differentially expressed genes (DEGs) underwent Gene Ontology (GO) annotation obtained using Blast2Go ($e\text{-value} < 1.0E^{-3}$), InterProScan, and GO Mapping [47] performed on OmicsBox software. Venn diagrams comparing total expressed genes between each replicate for the cultured and wild samples, were constructed using Venn Diagram Maker (Good Calculators). A heatmap of DEGs from all of the cultured and wild replicates was constructed using ClustVis, utilising z-scores as input data [48].

Asparagopsis sporophyte proteomics

Cultured sporophyte samples (as described above) were snap frozen and manually powdered with a mortar and pestle in liquid nitrogen. Proteins were initially extracted using Tris-HCl, pH 7.4 buffer, with and without acetone precipitation. For a single extraction, approximately 100 mg of the milled sample was homogenised in 20 mM Tris-HCl pH 7.4 buffer containing protease inhibitor (Merck), then briefly vortexed and centrifuged at 4,000 $\times g$ for 25 min at 4°C, then supernatant and pellets were extracted separately.

Initial supernatant was further extracted by non-acetone and acetone procedures. For extractions without acetone, the supernatant was collected and processed through an Amicon Ultra-4 centrifugal filter unit (3 kDa cut-off; Millipore Corporation, Billerica, MA, USA) at 4,000 $\times g$ for 40 min (4°C), then > 3 kDa and < 3 kDa samples were lyophilized in a vacuum concentrator (Express SpeedVac Concentrator SC250, Thermo Fisher Scientific). For extractions with acetone, the filtered supernatants were mixed with a pre-chilled (-20°C) 100% acetone (volume = 4-times that of each protein sample) and stored overnight at -20°C to allow for precipitation. Following centrifugation at 4,000 $\times g$ for 25 min at 4°C, supernatant was decanted, and the protein pellets were air-dried to evaporate the acetone. Protein pellets were resuspended in lysis buffer (6 M urea), briefly vortexed and centrifuged at 4,000 $\times g$ for 25 min (4°C). As above, the supernatant was collected, centrifuged through an Amicon Ultra-4 centrifugal filter unit (3 kDa cut-off), then > 3 kDa and < 3 kDa samples were lyophilized in a vacuum concentrator. Similarly, acetone precipitation of the supernatant was carried out, as above.

Lyophilized protein samples were resuspended in 0.1% trifluoroacetic acid (TFA) in Milli-Q, then processed by solid-phase extraction using a C₁₈ Sep-Pak cartridges (500 mg sorbent per cartridge, 55–105 μm - particle size; Waters), following the manufacturer's instructions, to ensure complete removal of salts from all samples. Eluates were lyophilized in a vacuum concentrator (Express SpeedVac Concentrator SC250, Thermo Fisher Scientific,) and stored at -20°C prior to further analysis. Samples extracted with acetone were trypsin-digested as previously described [49], following resuspension in 6 M urea.

Samples were analysed by ultra-High-Performance Liquid Chromatography (uHPLC) followed by Mass Spectrometry (MS) to identify proteins following a procedure previously described [49]. A 6 μl aliquot of each sample was injected onto a 50 mm $\times$ 300 μm C18 trap column (Agilent Technologies, Australia) and de-salted for 5 min using 0.1% formic acid (aq) at 30 $\mu l/min$. The trap column was then placed in-line with a 150 mm $\times$ 75 μm 300SBC18, 3.5 μm nano HPLC column (Agilent Technologies, Australia) for LC-MS analysis. Peptides were eluted at a flow rate of 300 nL/min using a linear gradient of 1–40% solvent B (solvent A = 0.1% formic acid in milliQ water; solvent B = acetonitrile: 0.1% formic acid (aq) 90:10) over 35 min followed by a steeper gradient from 40–80% solvent B over 5 min. The gradient was held at 80% solvent B for 5 min to wash the column and returned to 1% solvent B for equilibration prior to the next sample injection. The ion-spray voltage was set to 2400V, declustering potential (DP) 100V, curtain gas flow 25, nebuliser gas 1 (GS1) 12 and interface heater at 150°C. The mass spectrometer acquired 500 ms full scan TOF-MS data followed by 20 $\times$ 50 ms full scan product ion data in an Information Dependent Acquisition (IDA) mode. Full scan TOF-MS data was acquired over the mass range 350–1800 m/z and for product ion MS/MS 100–1800 m/z . Ions observed in the TOF-MS scan exceeding a threshold of 100 counts and a charge state of + 2 to + 5 were set to trigger the acquisition of product ion MS/MS spectra

of the resultant 20 most intense ions. The data was acquired and processed using Analyst TF 1.5.1 software (ABSCIEX, Canada).

Proteins were analysed using PEAKS v7.0 (BSI, Canada) against the protein database built from the reference *A. taxiformis* (L6) genome-derived protein database. *De novo* sequencing of proteins, database searches and characterisation of specific post-translational modifications will be used to analyse the raw data; false discovery rate will be set to $\leq 1\%$, and $[-10 \cdot \log(p)]$ will be calculated accordingly where p is the probability that an observed match is a random event. The PEAKS used the following parameters: (i) precursor ion mass tolerance, 0.1 Da; (ii) fragment ion mass tolerance, 0.1 Da (the error tolerance); (iii) tryptic enzyme specificity with two missed cleavages allowed; (iv) monoisotopic precursor mass and fragment ion mass; (v) a fixed modification of cysteine carbamidomethylation; and (vi) variable modifications including lysine acetylation, deamidation on asparagine and glutamine, oxidation of methionine and conversion of glutamic acid and glutamine to pyroglutamate. MS raw data was deposited into the PRIDE database under accession number PXD035669.

Results And Discussion

The quality of the genome was paramount and facilitated using domesticated sporophyte samples maintained in culture with relatively minimal microbial contamination. Genomic contaminants are a recurrent concern when creating any draft genome for an organism [50]. Microbes are a primary source of genomic contaminant in marine ecosystems, as they represent about 90% of the biomass in the ocean [51] and can establish intimate associations with multicellular hosts [52]. Seaweeds are classic examples of marine holobionts as they host dense communities of microbes that play key roles to host health and defense [10, 53]. In this regard, a complete separation between a seaweed and their microbial genome is often unfeasible, even in culture. However, minimizing microbial contamination, as well as acknowledging their genomic signatures should be considered prior to confirming a draft genome. Our microscopy images revealed extensive biofouling in wild samples of *A. taxiformis* (Fig. 1A,B), in comparison with the same species kept in cultures (Fig. 1C,D). The epiphytic aggregations in wild *A. taxiformis* morphologically resembled microalgae and prokaryotes, which are common epiphytes across seaweed groups [54].

The rich microbial community associated with wild *A. taxiformis* were reinforced by the genomic analysis, as there was a significantly higher abundance of microbial OTUs associated with wild individuals in comparison with cultured *Asparagopsis* (Table 1, **Fig. S2**). Both microscopy and genomic analysis indicated that the microbial community was less abundant on cultured *A. taxiformis* and, therefore, these were selected for establishing the *A. taxiformis* genome. Details of the microbiomes associated with different life history stages of this seaweed maintained under natural and cultured conditions, is of interest for future research.

Table 1

One-way ANOVA results for microbial OTU richness associated with wild versus cultured *Asparagopsis taxiformis* individuals. SS = sum of squares, df = degrees of freedom, MS = mean square.

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1484185.34	1	1484185.34	40.5938681	0.00012957	5.11735503
Within Groups	329056.3	9	36561.8111			
Total	1813241.64	10				

Phylogeny of *Asparagopsis* and sample used for genomic analysis

Our phylogenetic analysis of the *cox2-cox3* spacer, commonly used in intra-specific diversity studies in *Asparagopsis*, indicated that the sporophyte sample used in the genome sequencing (2019; submitted as NCBI GenBank accession number OP779373), is a member of lineage 6 of *A. taxiformis* [22] (**Fig. S3**). This lineage is common in tropical Australia, Queensland and Western Australia. Our sample is part of a strain designated the Great Barrier Reef strain (GBR) [22] and used in feeding trials in Australia (41–44). The genetic diversity of *Asparagopsis*, both *A. taxiformis* and *A. armata*, indicates that multiple distinct lineages are found in both species, that suggest there may be many cryptic species within these morphological forms, although formal taxonomic changes have not been made [22, 23]. While this tropical Australian lineage is the best studied so far, the monophyly of all these other taxonomically related '*A. taxiformis*', and including *A. armata*, lineages suggested that the genomic information presented here will serve as an important resource, and reference for future comparative studies on these other lineages.

Genomic features of *Asparagopsis taxiformis*, Lineage 6 (L6)

We present the first draft genome for the red seaweed *A. taxiformis* (L6) at 142 Mb in size. This high-quality nuclear genome for *Asparagopsis*, with its large total assembly size (total base), average contig length and median contig length, provided the basis for the following multi-omics exploration of this seaweed. As shown in Table 2, the assembled genome of *A. taxiformis* (L6) contained 3,308 contigs with a median length of 55,006 bp. By comparison to the only other reported *A. taxiformis* assemblies [55], this assembly had a larger total assembly size (total base), higher average contig length and higher median contig length (N50). Based on the 255 Eukaryotic BUSCO groups, we found 208 complete BUSCOs (81.6% complete genes), among which there were 183 complete single copy BUSCOs. In addition, there were only 11 fragmented BUSCOs and 36 absent BUSCOs. Therefore, the *A. taxiformis* (L6) had a better overall assembly, compared to *A. taxiformis* Guam (GCA_018397955.1) and *A. taxiformis* California (GCA_018397975.1) (**File S1b**).

Table 2
Overall statistics for *A. taxiformis* Sunshine Coast
genome assembly.

Assembly statistics	
Total base	142,472,235
Number of contigs	3,308
Average length	43,069
Largest contig length	523,412
N50	55,006
Gaps	0
BUSCO overview	
Overall coverage (C/Total)	81.60%
Complete BUSCOs (C)	208
Complete and single-copy BUSCOs (S)	183
Complete and duplicated BUSCOs (D)	25
Fragmented BUSCOs (F)	11
Missing BUSCOs (M)	36
Total BUSCO groups searched	255

Approximately 70.67% of the *A. taxiformis* (L6) assembly was interspersed repeats, including 20.71% retroelements, 33.48% DNA transposons, 0.21% rolling circles and 16.47% unclassified repeats. The conserved gene structures collected from BUSCO were used for training in AUGUSTUS and predicted 10,867 putative genes in *A. taxiformis* (L6) genome (**File S1c**). In addition, there are full-length genomes for the mitochondria (26,034 bp) and plastid (176,936 bp). By using sequence-based searching against our mitochondrial and plastid genomes, we predicted 50 mitochondrial genes and 227 plastid genes, respectively (**File S1d**). Among the 50 mitochondrial genes, we found two ribosomal RNA genes and 22 different tRNA genes.

We explored the comparative nuclear genome size and the number of protein-coding genes across 8 red seaweed whole genomes (Fig. 2A). The *A. taxiformis* (L6) had the largest genome at 142.472 Mb, yet ranked third in terms of number of protein-coding genes. *Porphyra umbilicalis* contains a predicted 13,125 genes, and a genome size of 87.889 Mb [1]. In total, the *A. taxiformis* (L6) genome contained 6,858 monoexonic genes (63.13%). The ratio between average and median intergenic gene distances (3.78) indicated the gene organization was not as compact as for *C. crispus* (8.625), whose genes had exceptionally low gene density and were highly clustered [40].

Metagenomic analysis of the genome presented strong evidence to support our microscopic analysis, in that *A. taxiformis* (L6) sporophytes cultured in semi-sterile conditions harbored a reduced microbial community, compared to wild sporophytes. Despite this, a total of 258 reads were taxonomically assigned to prokaryotes or viral groups (**File S1e**), which corresponded to 0.7% of the *A. taxiformis* (L6) sequences. A previous study using Kraken screening to investigate genomic contaminants in eukaryotic draft genomes demonstrated that an average of 10% of their pseudo-reads (reads reconstructed from assembled genomes) were identified as foreign species, mostly matching bacteria [50]. Thus, microbial sequences were present in low proportions in the *A. taxiformis* (L6) genome, primarily consisting of Alphaproteobacteria groups, in which Rhodobacterales were the most abundant order followed by Rhizobiales. The dominance of these groups in *A. taxiformis* (L6) genome are in accordance with other red seaweed microbial profiles [56]. Members of these orders are noted for often establishing mutualistic and symbiotic relationships with marine hosts, playing key roles in biogeochemical cycling and metabolic pathways [57–59].

Comparative and functional features of the *A. taxiformis* (L6) genome

To provide a well-annotated genomics resource, we utilized a BLAST2GO pipeline, which is the most efficient approach to associate genes to known proteins with function. Of the 10,867 genes, there were 8,856 genes (81.52%) with homologous hits against the NCBI non-redundant (NR) database based on an E-value of 0.00001. The similarity for each BLAST hit was defined by the division of exact number of matches over the length of the hit. The majority of matching hits had over 60% similarity (**Fig. S4**) and the species distribution showed that the highest identity was with *Chondrus crispus* (2,583 genes) followed by *Gracilariopsis chorda* (2,182 genes) (Fig. 2B), reflecting their evolutionary history [60]. Among the genes with NR homologs, 6,786 genes (62.47%) were further mapped to gene ontology (GO). By examining the top 10 GO, we found the most abundant categories consisted of cytosol and membrane genes (cellular components), peptide transport and protein phosphorylation (biological processes) and protein and ATP binding genes (molecular function) (**Fig. S5**).

All 10,867 *A. taxiformis* (L6) genes contained a protein functional site and domain (assigned by InterPro annotations), while 6,300 genes (57.99%) had Pfam protein family/domain mapping. By comparing the Pfam annotation among 7 other red algae genomes (same species as Fig. 2A), we found that *A. taxiformis* shared 2,471 Pfam domains with *Chondrus crispus* (supporting our BLAST analysis) and 2,938 with other red algae species (Fig. 2C). Of interest, 381 *A. taxiformis* unique domains could be placed into categories of functional domains, protein families and repeats, based on InterPro annotations (**Fig. S6**). The most prominent of these (186 genes) were associated with the WD40 repeat, which encompass members of the beta-propeller domain family that are known to form interaction scaffolds in protein complexes [61]. Mapping *A. taxiformis* genes to the KEGG database, showed that 5,368 genes (49.42%) had KEGG orthologs, of which 1790 genes were categorized into “metabolism”. Of those with 50 or more genes, we found purine metabolism and pyruvate metabolism contain over 80 genes in the *A. taxiformis* genome (Fig. 2D).

As the key regulator in gene expression, transcription factors (TFs) are critical to understand the potential developmental processes of organisms. To explore the presence of TF families in the *A. taxiformis* genome compared to other red algae, we identified 145 TFs in the genome (**File S1f**) and performed a comparative genome-wide analysis of TFs in seaweeds (Fig. 2E). Similar to other seaweeds, TFs belonging to C2H2, bZIP, MYP-related and bHLH have the highest representation in *A. taxiformis*. Some TFs are located close to each other in the genome, where for example, two MYB-related TFs are gene neighbours. MYB TFs are generally associated with plant development and, biotic and abiotic stress responses in the plants [62]. A single Trihelix TF (TTF) was identified in *A. taxiformis* (L6), while no TTF was identified in other red seaweeds. The TTF has three-helix structure (helix-loop-helix-loop-helix) and is sensitive to light [63].

Ortholog analysis was performed on 8 red seaweed genome-derived proteome datasets (72,277 proteins) that had each been subjected to all-against-all BLASTp (E-value of 0.00001). We found that 56,813 proteins (78.6%) were clustered into 9,841 orthologous groups, which were used to construct a species tree (Fig. 3A,B). Among the orthologous groups, 2,430 (11,578 proteins) were species-specific. The average orthologous group size was 5.8 genes, while there were 604 single-copy orthologous groups (4,832 proteins). In *A. taxiformis* (L6), 8,898 proteins (81.9%) were mapped into 5,047 orthologous groups. Interestingly, 15,148 putative gene duplication events were identified among the eight genomes, all of which are traced back to the node of the species tree on which these duplications occurred. For *A. taxiformis* (L6), there were 1,359 non-terminal duplications, which may provide clues to the phylogenetic relationships of gene families. These were mapped to 1,233 Pfam domains and the number of genes in each protein family from each species was aligned with the species tree.

To better understand lineage-specific gene duplications, we mapped 4,829 *A. taxiformis*-specific duplicated genes to Pfam annotations, which could provide evidence of genome-wide duplications (e.g., polyploidy). Then we prioritized the Pfam domains by the number of duplicated genes (Fig. 3C). The gene duplications often generate new copies, which may have relatively less selection pressure. Therefore, these duplicated genes may change and be one of the major driving forces in genome evolution [64]. For instance, a relatively abundant number of WD40 repeat proteins were identified, supporting previously observed expansion of these protein domains. Potentially the most striking feature were the 66 genes associated with a N6_N4_Mtase domain in *A. taxiformis* (L6) compared to *P. umbilicalis* (5 genes), *G. chorda* (4 genes) and *C. crispus* (6 genes). In terms of metabolic enzyme duplication, we identified 25 glucose/sorbose dehydrogenases (GSDHs) in the *A. taxiformis* (L6) genome, which was expanded from GSDHs in *P. purpureum* (1 gene), *P. umbilicalis* (5 genes), *G. chorda* (6 genes), *C. crispus* (5 genes). In plants, GSDH are important for ascorbic acid metabolism to generate ascorbate precursors [65]. The phosphatidylserine decarboxylases (EC:4.1.1.65, PS_Dcarbxyase), which are known to catalyse the reaction: phosphatidyl-L-serine $\rightleftharpoons$ phosphatidylethanolamine + CO₂ [66] are prominent in *A. taxiformis* (L6) (15 genes), compared to the other seaweeds that have a combined 9 phosphatidylserine decarboxylase genes. In summary, these duplicated enzymes may provide clues to explore gene innovations for the biosynthesis of ascorbate precursors and aminophospholipids in *A. taxiformis* (L6).

A. taxiformis (L6) cultured and wild sporophyte: an integrative omics analysis

The availability of a high-quality draft genome provided the resource for a first opportunity for an integrative *Asparagopsis*-omics investigation, to explore how environmental conditions alter sporophyte bromoform content and gene expression. Cultured *A. taxiformis* (L6) sporophytes were maintained in relatively sterile water at constant temperature, salinity and light cycles, whereas wild sporophytes are exposed to changing environmental conditions (e.g. temperature, salinity) and various predators (e.g. rabbitfish [67]), so genes associated with defence and stress-resistance are likely of significant importance. Also, cultured sporophytes (under the conditions described in this study) have never transitioned into reproductive sporophytes, suggesting a lack of specific cues, and control of reproduction might enable potential trade-offs between reproduction and growth to be managed [68].

Differences in gene expression were investigated between cultured and wild sporophytes of *A. taxiformis*. Overall, cultured sporophyte total gene expression exhibited stronger clustering based on principal component analysis, highlighting greater variation in gene transcripts of wild samples (**Fig. S7**). Of those genes expressed, 399 and 257 genes were identified as significantly differentially expressed (FDR $P < 0.05$, $>\pm 4 \log_2$ fold-change) between cultured and wild sporophytes, respectively (Fig. 4A). In cultured sporophytes, there was a relative enrichment of genes that annotated to enzymes (e.g., kinases, serine carboxypeptidase, cytochrome P450) and regulatory factors (e.g., zinc fingers, translation initiation factors), as well as those with an unknown function (e.g. unknown, unnamed, hypothetical; Fig. 4B). In addition, *cyclin* (*AtaL6_9753*), which encodes a protein well known for complexing with cyclin dependent kinase (CDK-cyclin complexes) to regulate progression of the cell cycle [69], was notably absent, yet highly expressed in wild sporophytes (**File S1g**). This could help explain the observed suppression of cultured sporophyte progression through to the reproductive stage. We expect that more in-depth molecular comparative analysis of non-reproductive and reproductive (with sporangia) sporophytes, including changes in CDK-cyclin complex coupling, could provide the necessary knowledge to manipulate sporophyte transitions in seaweeds, including *A. taxiformis* (L6).

In wild sporophytes, there was a relative enrichment of genes that annotated to haloperoxidases, animal heme peroxidase-like homologs and heat shock proteins (HSPs), which fit within designated categories (based on *C. crispus* gene annotations [40]) related to defence and stress-resistance genes (Table 3 and Fig. 4C), and therefore, their relative abundance in wild sporophytes was of interest. The HSP family (including, but not limited to, HSP20, HSP70, HSP100) are well known for their role in facilitating protein folding, assembly, translocation and degradation during normal cellular homeostasis [70]. However, they have been of most interest due to their additional ability to stabilize proteins and assist refolding under stress conditions [71]. If not just by virtue of their name, the HSPs will remain targets for the development of stress resilience in organisms, including seaweeds. We identified that several stress HSPs were upregulated in the wild sporophyte, and specifically those within the HSP20 family (molecular weights ~ 20 kDa, so also known as small HSPs). This aligns with research in the red seaweed *Pyropia yezoensis*, where small HSPs were upregulated during high temperature, oxidative stress or copper exposure [72].

Table 3

Summary of *A. taxiformis* (L6) genes involved in defence-related and stress. Number of corresponding genes significantly upregulated (> 4-fold) in cultured or wild, is shown. For more details, see **File S2**.

Category	Number	Examples	Cultured	Wild
Halogen metabolism	49	PAP2/vanadium haloperoxidases, non-animal heme peroxidases	1	5
Potential pathogen receptors	391	WD40-repeat containing proteins, LRR and TRR domain proteins, Sel1 repeat containing proteins	8	6
Potential defence effectors	8	SGT1 ortholog, apoptosis inducing factor, exportin	0	0
Stress genes ROS scavenging	27	Ascorbate peroxidase, Cu/Zn superoxide dismutase, glutathione reductase, peroxiredoxin	4	1
Stress genes heat shock proteins	59	HSP20, HSP40, HSP90, HSP100, HSP70 binding protein, heat shock transcription factor	4	7
Stress related genes	59	T-complex alpha/beta/gamma, tubulin binding complex B/C/D, peptidyl-prolyl cis-trans isomerase	4	5
Cytochrome P450 (CYP450)	13	CYP51G, CYP97G, CYP80B	3	0

Halogenating peroxidases, or haloperoxidases (HPOs), are enzymes that oxidatively activate halides to the corresponding hypohalites at the expense of peroxides that belong to two groups: Heme-dependent or vanadium-dependent, which differ with respect to the prosthetic group and consequently in their catalytic mechanisms [73]. In seaweeds, numerous vanadium haloperoxidases (VHPOs) exist, while vanadium-dependent bromoperoxidases (VBPOs) have been of most interest due to their production of brominated compounds (e.g. bromine), which are assumed to be used in chemical defences in organisms [7]. Knowledge of how brominated compounds are biosynthesized could be used to improve their supply. We found that *A. taxiformis* (L6) sporophytes maintained in culture produced higher levels of bromoform (18.02 +/-2.3 mg/g dried algae) when compared to those in the wild (4.4 +/-0.96 mg/g dried algae) (Fig. 5A). Twenty-one genes identified in the *A. taxiformis* (L6) genome demonstrated significant (e-value > 10⁻³) similarities with PAP2/VHPOs. Phylogenetic analysis supported 3 predominant clusters, of which 15 had at least some expression in the cultured or wild sporophytes (Fig. 5B). Only the VHPO gene *AtaL6 10688* showed a significant difference, being higher in cultured sporophytes than in wild sporophytes. The *AtaL6 9003* demonstrated highest overall expression, and with comparable expression in both cultured and wild experimental samples. *AtaL6 9003* corresponds to a VBPO previously described as *mbb 1* of the *mbb* locus [55]. The *mbb* locus, known to consist of 3 VBPOs and 1 NADPH-oxidase (NOX), was identified in *A. taxiformis* (L6), with VBPO *AtaL6 8990* identified in close proximity (Fig. 5C). A 5th gene within the *mbb* locus, denoted as *mbb2.5* (*AtaL6 9001*), was predicted and further confirmed to be present in the California and Guam genomes of *A. taxiformis* (GenBank MN966723 and MN893468, respectively) (Fig. 5D) [55]. Conservation of *mbb* locus genes between the known *A. taxiformis* genomes

indicated significant intra-specific variability for these genes (e.g., *mbb1*: 96.56–97.74%). It is possible that the variability of the *mbb* locus genes could be because neither the California or Guam derived genomes were from Lineage 6 of *Asparagopsis taxiformis*, although this can't be ascertained from the data presented in [6]. Lineage 6 of *A. taxiformis* has not been confirmed from California or Guam. In comparison, sequence analysis of the 3 VBPOs demonstrated that these proteins are moderately conserved and have similar lengths, between 581 and 590 amino acids long. The *AtaL6* VBPOs all featured a PAP2 domain, which is situated between the early-300 to the mid-500 amino acid range and covered short regions of both high and low conservation. In addition, all 3 VBPOs contained 8 highly conserved PAP2 active site residues, including a histidine (H), suggested to covalently bond to the vanadate cofactor [74].

An *in silico* secretome analysis of the *A. taxiformis* (L6) genome-derived proteins, using conventional protein secretion pathway predictions [75], found that 345 proteins were likely secreted from cells (**File S1h**). This included an expanded family of proteins annotated as 'collagen-alpha-like' proteins (~ 43; 12.5% of total proteins secreted), several of which were enriched (Fig. 6) and demonstrated significantly higher expression in the wild sporophyte (Fig. 6A). These proteins all shared the features of a signal peptide and at least one von Willebrand factor type A (VWA) region with spatially conserved cysteine residues; structure models indicated similar sequence identity (16.29%) to a proximal thread matrix protein-1 (PDB # 4cn9.2.A) found in the mussel byssus (Fig. 6B; **Fig. S8**) [76]. The proximal thread matrix protein 1 can bind collagen, thereby influencing the thread mechanical characteristics (e.g., flexibility) [77]. VWA-containing proteins are often associated with protein-protein interactions [78]. BLASTp similarity searches demonstrated some conservation (up to 62%) with predicted proteins from the red seaweeds *G. chorda* and *C. crispus* (Fig. 6C). Based on the aforementioned evidence, we suggest that these secreted proteins in *A. taxiformis* (L6), herein named rhodophyte collagen-alpha-like proteins (RCAPs), may help regulate the seaweeds flexibility and/or substrate attachment; cultured sporophytes used in this study were not attached to a substrate. A secreted liquid mucilage is critical for seaweed bioadhesion, allowing for cell wall adhesion to a substrate, then irreversibly hardening [79].

Asparagopsis proteomics

The comparative gene expression analysis provided protein targets for further investigation and their isolation (as native proteins) from seaweeds would enable further analysis, including structural and functional characterisation. For example, the *A. taxiformis* (L6) PAP2/VHPOs, animal heme peroxidase-like and RCAPs were of interest, so their isolation would help for future functional characterisation to provide a full understanding into their role in brominated compound biosynthesis and defence. The *A. taxiformis* (L6) genome provided an excellent opportunity for proteomic investigation of this seaweed, with specific interest to isolate the secretome proteins reported above. Of the proteins predicted from the genome, 421 proteins were identified following sporophyte extraction (with Tris and Urea) and proteomic analysis (Fig. 7 and **File S3**), of which, 25 were predicted to be secreted including a VHPO (encoded by *AtaL6* 1934), an animal heme peroxidase-like (*AtaL6* 3164) and an RCAP (*AtaL6* 10422). Besides putative secreted proteins, 45% of the extracted proteins consisted of enzymes (e.g., lipoxigenase, fructose-1,6-

biphosphate aldolase, peroxiredoxin), 4% phycoproteins (e.g., allophycocyanin beta 18 subunit, phycoerythrin gamma 31 kDa subunit, phycobilisome linker polypeptides) and 44% that were either hypothetical or unknown.

This proteomic investigation of *A. taxiformis* (L6) will complement other experimental approaches to protein-based production in *Asparagopsis* (and other seaweeds), such as production of recombinant proteins (e.g., [55]) and gene/protein knockdowns. In addition, proteins that annotated with enzymes were highly represented in the proteome; seaweed enzymes show prebiotic and nutraceutical potential, primarily through applications that develop bioactive compounds from seaweeds, as well as assist in the extraction and hydrolysis of macromolecules [80]. Also, phycobiliproteins (water-soluble pigments) are highly sought after as natural pigments of food, cosmetics, dyes and other industries [81], so their purification and identification via a genomic resource database will assist the development of approaches for large-scale purification.

Conclusions

Genome sequencing of *A. taxiformis* (L6) has greatly enhanced the genetic resources available for this much lauded seaweed, most well-known for its potential in mitigating methane production in ruminants. However, the potential applications of *Asparagopsis* in agriculture and aquaculture are diversifying, with new research and development relating to enhancing fish production and boosting immunity [82], novel antimicrobial and biotechnology uses [83], as well as environmental applications with respect to understanding and utilizing *Asparagopsis* (L6) where it has become a pest and invasive species [84]. To facilitate precision genetic manipulation of this seaweed for large-scale aquaculture production, we have shared our high-quality draft genome assembly, gene and protein sequences, which can be used for in-depth research. This could include CRISPR (for controlled heritability/land-based production/selection) and, because of the significant variability between lineages of *Asparagopsis* (L6), paving the way for more accessible genomics through tools like Nanopore. At present, there is significant investment in the aquaculture and farm-level demonstration of *Asparagopsis* collected and cultured throughout the globe. The *Asparagopsis* (L6) work to date has focused on utilizing wild strains and natural features of the seaweed that allow for simple vegetative production of explants or semi-controlled release of reproductive propagules. This timely database could help deliver further updates to the *Asparagopsis* (L6), by characterizing genes associated with bromoform synthesis, stress, reproduction and much more, for examples leading to avenues for marker-assisted seaweed breeding that can select for high-yielding individuals. Finally, the *A. taxiformis* (L6) genome allows for more robust comparative evolutionary comparisons and assessment of genome innovations in seaweed.

Declarations

Acknowledgements

This work has been supported by the Australian Research Council Discovery Project (DP200103013). This research was also supported by computational resources provided by the Australian Government through QCIF under the National Computational Merit Allocation Scheme. This storage in this research was supported by the use of the Nectar Research Cloud, a collaborative Australian research platform supported by the NCRIS-funded Australian Research Data Commons (ARDC). We acknowledge the Traditional Owners of the land and sea in which this work was done, the Kabi Kabi/Gubbi Gubbi people in the area now known as the Sunshine Coast, Queensland, Australia.

Author contributions

MZ carried out the analyses and developed the database. ZPP, JW and AW undertook sample collection and preparation for sequencing. ZPP and TW performed proteomics analysis. TL carried out *in silico* secretome analysis of samples. LM helped to characterize gene families. CP performed genomic contamination analysis. DH performed bromoform concentration analysis. GCZ performed analysis of strain. MZ, AHC, NS, ES, GCZ, NAP and SFC conceived the idea and helped write the manuscript.

Data availability

Raw data of all genome sequencing was submitted to NCBI Bioproject and SRA database under a Bioproject accession number PRJNA809757. Phylogenetic identification data was submitted under GenBank accession number OP779373. Mass spectrometry peptide raw data was deposit into the PRIDE database under accession number PXD035669.

References

1. Brawley SH, Blouin NA, Ficko-Blean E, Wheeler GL, Lohr M, Goodson HV, et al. Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis* (Bangiophyceae, Rhodophyta). *Proc Natl Acad Sci U S A*. 2017;114(31):E6361-E70. doi: 10.1073/pnas.1703088114. Epub 2017 Jul 17.
2. Barrows G, Sexton S, Zilberman D. Agricultural biotechnology: The promise and prospects of genetically modified crops. *Journal of Economic Perspectives*. 2014;28(1):99–120.
3. Climate change 2014 - Synthesis report. 2015.
4. National Greenhouse Gas Inventory Kyoto Protocol classifications: ageis.climatechange.gov.au; 2015 [cited 2022].
5. Machado L, Magnusson M, Paul NA, de Nys R, Tomkins N. Effects of marine and freshwater macroalgae on in vitro total gas and methane production. *PLoS One*. 2014;9(1):e85289.
6. Machado L, Magnusson M, Paul NA, Kinley R, de Nys R, Tomkins N. Dose-response effects of *Asparagopsis taxiformis* and *Oedogonium* sp. on in vitro fermentation and methane production. *Journal of Applied Phycology*. 2016;28(2):1443–52.

7. Paul NA, de Nys R, Steinberg PD. Chemical defence against bacteria in the red alga *Asparagopsis armata*: linking structure with function. *Molecular Ecology Progress Series*. 2006;306:87–101.
8. Tomkins N, Colegate S, Hunter R. A bromochloromethane formulation reduces enteric methanogenesis in cattle fed grain-based diets. *Animal Production Science*. 2009;49(12):1053–8.
9. Davison T, Hill J, Savage S. Reducing Emissions from Livestock Research Program. Meat and Livestock Australia Limited. 2012.
10. Egan S, Harder T, Burke C, Steinberg P, Kjelleberg S, Thomas T. The seaweed holobiont: understanding seaweed-bacteria interactions. *FEMS Microbiol Rev*. 2013;37(3):462–76. doi: 10.1111/574-6976.12011. Epub 2012 Dec 10.
11. Fenical W. Natural products chemistry in the marine environment. *Science*. 1982;215(4535):923–8.
12. Hollants J, Leliaert F, De Clerck O, Willems A. What we can learn from sushi: a review on seaweed–bacterial associations. *FEMS microbiology ecology*. 2013;83(1):1–16.
13. Zhou M, Hünnerberg M, Chen Y, Reuter T, McAllister TA, Evans F, et al. Air-dried brown seaweed, *Ascophyllum nodosum*, alters the rumen microbiome in a manner that changes rumen fermentation profiles and lowers the prevalence of foodborne pathogens. *MSphere*. 2018;3(1):e00017-18.
14. Kinley R, Fredeen A. In vitro evaluation of feeding North Atlantic stormtoss seaweeds on ruminal digestion. *Journal of Applied Phycology*. 2015;27(6):2387–93.
15. Manley SL. Phytogenesis of halomethanes: A product of selection or a metabolic accident? *Biogeochemistry*. 2002;60(2):163–80.
16. Mata L, Lawton RJ, Magnusson M, Andreakis N, de Nys R, Paul NA. Within-species and temperature-related variation in the growth and natural products of the red alga *Asparagopsis taxiformis*. *Journal of Applied Phycology*. 2017;29(3):1437–47.
17. Wahl M, Goecke F, Labes A, Dobretsov S, Weinberger F. The second skin: ecological role of epibiotic biofilms on marine organisms. *Frontiers in Microbiology*. 2012;3:292.
18. Vergés A, Paul NA, Steinberg PD. Sex and life-history stage alter herbivore responses to a chemically defended red alga. *J Ecology*. 2008;89(5):1334–43.
19. Gregersen S, Pertseva M, Marcatili P, LøvstadHoldt S, Jacobsen C, García-Moreno PJ, et al. Proteomic characterization of pilot scale hot-water extracts from the industrial carrageenan red seaweed *Eucheuma denticulatum*. *Algal Research*. 2022;62.
20. Wang D, You W, Chen N, Cao M, Tang X, Guan X, et al. Comparative Quantitative Proteomics Reveals the Desiccation Stress Responses of the Intertidal Seaweed *Neoporphyra haitanensis*. *J Phycol*. 2020;56(6):1664–75. doi: 10.1111/jpy.13052. Epub 2020 Oct 22.
21. Batut B, Hiltemann S, Bagnacani A, Baker D, Bhardwaj V, Blank C, et al. Community-Driven Data Analysis Training for Biology. *Cell Syst*. 2018;6(6):752-8.e1. doi: 10.1016/j.cels.2018.05.012.
22. Andreakis N, Costello P, Zanolli M, Saunders GW, Mata L. Endemic or introduced? Phylogeography of *Asparagopsis* (Florideophyceae) in Australia reveals multiple introductions and a new mitochondrial

- lineage. *J Phycol.* 2016;52(1):141–7. Epub 2016/03/18. doi: 10.1111/jpy.12373. PubMed PMID: 26987096.
23. Preuss M, Nelson WA, D'Archino R. Cryptic diversity and phylogeographic patterns in the *Asparagopsis armata* species complex (Bonnemaisoniales, Rhodophyta) from New Zealand. *Phycologia.* 2022;61(1):89–96. doi: 10.1080/00318884.2021.2015223.
24. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27(5):722–36. Epub 2017/03/17. doi: 10.1101/gr.215087.116. PubMed PMID: 28298431; PubMed Central PMCID: PMC5411767.
25. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 2014;9(11):e112963. Epub 2014/11/20. doi: 10.1371/journal.pone.0112963. PubMed PMID: 25409509; PubMed Central PMCID: PMC4237348.
26. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–2. Epub 2015/06/11. doi: 10.1093/bioinformatics/btv351. PubMed PMID: 26059717.
27. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A.* 2020;117(17):9451–7. Epub 2020/04/18. doi: 10.1073/pnas.1921046117. PubMed PMID: 32300014; PubMed Central PMCID: PMC7196820.
28. Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics.* 2004;Chap. 4:Unit 4 10. Epub 2008/04/23. doi: 10.1002/0471250953.bi0410s05. PubMed PMID: 18428725.
29. Afgan E, Jalili V, Goonasekera N, Taylor J, Goecks J. Federated Galaxy: Biomedical Computing at the Frontier. *IEEE Int Conf Cloud Comput.* 2018;2018:10.1109/cloud.2018.00124. (doi):10.1109/cloud.2018.00124. Epub 2018 Sep 10.
30. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 2006;34(Web Server issue):W435-9. Epub 2006/07/18. doi: 10.1093/nar/gkl200. PubMed PMID: 16845043; PubMed Central PMCID: PMC1538822.
31. Conesa A, Gotz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International journal of plant genomics.* 2008;2008:619832. doi: 10.1155/2008/619832. PubMed PMID: 18483572; PubMed Central PMCID: PMC2375974.
32. Coordinators NR. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2016;44(D1):D7-19. doi: 10.1093/nar/gkv1290. PubMed PMID: 26615191; PubMed Central PMCID: PMC4702911.
33. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44(D1):D457-62. doi: 10.1093/nar/gkv1070.

PubMed PMID: 26476454; PubMed Central PMCID: PMCPMC4702792.

34. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 2013;41(12):e121. doi: 10.1093/nar/gkt263. PubMed PMID: 23598997; PubMed Central PMCID: PMCPMC3695513.
35. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238. Epub 2019/11/16. doi: 10.1186/s13059-019-1832-y. PubMed PMID: 31727128; PubMed Central PMCID: PMCPMC6857279.
36. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30(14):3059–66. Epub 2002/07/24. doi: 10.1093/nar/gkf436. PubMed PMID: 12136088; PubMed Central PMCID: PMCPMC135756.
37. Emms DM, Kelly S. STAG: Species Tree Inference from All Genes. *bioRxiv*; 2018.
38. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol.* 2009;26(7):1641–50. Epub 2009/04/21. doi: 10.1093/molbev/msp077. PubMed PMID: 19377059; PubMed Central PMCID: PMCPMC2693737.
39. Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 2017;45(D1):D1040-D5. Epub 2016/12/08. doi: 10.1093/nar/gkw982. PubMed PMID: 27924042; PubMed Central PMCID: PMCPMC5210657.
40. Collén J, Porcel B, Carré W, Ball SG, Chaparro C, Tonon T, et al. Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc Natl Acad Sci U S A.* 2013;110(13):5247–52. doi: 10.1073/pnas.1221259110. Epub 2013 Mar 15.
41. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol.* 2019;37(4):420-3. Epub 2019/02/20. doi: 10.1038/s41587-019-0036-z. PubMed PMID: 30778233.
42. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305(3):567–80. Epub 2001/01/12. doi: 10.1006/jmbi.2000.4315. PubMed PMID: 11152613.
43. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research.* 2019;47(W1):W636-W41. doi: 10.1093/nar/gkz268. PubMed PMID: 30976793.
44. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic acids research.* 2017;45(D1):D200-D3. Epub 2016/11/29. doi: 10.1093/nar/gkw1129. PubMed PMID: 27899674.
45. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018;46(W1):W296-W303. doi: 10.1093/nar/gky427.

46. Romanazzi D, Sanchez-Garcia C, Svenson J, Mata L, Pes K, Hayman CM, et al. Rapid Analytical Method for the Quantification of Bromoform in the Red Seaweeds *Asparagopsis armata* and *Asparagopsis taxiformis* Using Gas Chromatography–Mass Spectrometry. *ACS Agricultural Science & Technology*. 2021;1(5):436–42. doi: 10.1021/acsagscitech.1c00161.
47. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*. 2008;36(10):3420-35. doi: 10.1093/nar/gkn176%J Nucleic Acids Research.
48. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research*. 2015;43(W1):W566-W70. doi: 10.1093/nar/gkv468%J Nucleic Acids Research.
49. Hall MR, Kocot KM, Baughman KW, Fernandez-Valverde SL, Gauthier ME, Hatleberg WL, et al. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature*. 2017;544(7649):231–4.
50. Lu J, Salzberg SL. Removing contaminants from databases of draft genomes. *PLoS Comput Biol*. 2018;14(6):e1006277. doi: 10.1371/journal.pcbi.. eCollection 2018 Jun.
51. Worden AZ. Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquatic Microbial Ecology*. 2006;43:165–75. doi: DOI: 10.3354/ame043165.
52. Rosenberg E, Zilber-Rosenberg I. Microbes Drive Evolution of Animals and Plants: the Hologenome Concept. *mBio*. 2016;7(2):e01395. doi: 10.1128/mBio.-15.
53. Bengtsson MM, Sjøtun K, L. ø. Seasonal dynamics of bacterial biofilms on the kelp *Laminaria hyperborea*. *Aquatic Microbial Ecology*. 2010;(60):71–83. doi: DOI: 10.3354/ame01409.
54. Potin P. Intimate Associations Between Epiphytes, Endophytes, and Parasites of Seaweeds. In: Wiencke C, Bischof K, editors. *Seaweed Biology*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2012. p. 203–34.
55. Thapa HR, Lin Z, Yi D, Smith JE, Schmidt EW, Agarwal V. Genetic and Biochemical Reconstitution of Bromoform Biosynthesis in *Asparagopsis* Lends Insights into Seaweed Reactive Oxygen Species Enzymology. *ACS Chem Biol*. 2020;15(6):1662–70. Epub 2020/05/27. doi: 10.1021/acschembio.0c00299. PubMed PMID: 32453942; PubMed Central PMCID: PMCPMC7306000.
56. Selvarajan R, Sibanda T, Venkatachalam S, Ogola HJO, Christopher Obieze C, Msagati TA. Distribution, Interaction and Functional Profiles of Epiphytic Bacterial Communities from the Rocky Intertidal Seaweeds, South Africa. *Sci Rep*. 2019;9(1):19835. doi: 10.1038/s41598-019-56269-2.
57. Simon M, Scheuner C, Meier-Kolthoff JP, Brinkhoff T, Wagner-Döbler I, Ulbrich M, et al. Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J*. 2017;11(6):1483–99. doi: 10.038/ismej.2016.198. Epub 7 Jan 20.
58. Dang H, Li T, Chen M, Huang G. Cross-ocean distribution of Rhodobacterales bacteria as primary surface colonizers in temperate coastal marine waters. *Appl Environ Microbiol*. 2008;74(1):52–60. doi: 10.1128/AEM.01400-07. Epub 2007 Oct 26.

59. Zhao Y, Wang K, Budinoff C, Buchan A, Lang A, Jiao N, et al. Gene transfer agent (GTA) genes reveal diverse and dynamic *Roseobacter* and *Rhodobacter* populations in the Chesapeake Bay. *ISME J*. 2009;3(3):364–73. doi: 10.1038/ismej.2008.115. Epub Nov 20.
60. Verbruggen H, Maggs CA, Saunders GW, Le Gall L, Yoon HS, De Clerck O. Data mining approach identifies research priorities and data requirements for resolving the red algal tree of life. *BMC Evol Biol*. 2010;10:16. Epub 2010/01/22. doi: 10.1186/1471-2148-10-16. PubMed PMID: 20089168; PubMed Central PMCID: PMC2826327.
61. Xu C, Min J. Structure and function of WD40 domain proteins. *Protein Cell*. 2011;2(3):202 – 14. doi: 10.1007/s13238-011-1018-1. Epub 2011 Apr 6.
62. Ambawat S, Sharma P, Yadav NR, Yadav RC. MYB transcription factor genes as regulators for plant responses: an overview. *Physiol Mol Biol Plants*. 2013;19(3):307–21. Epub 2014/01/17. doi: 10.1007/s12298-013-0179-1. PubMed PMID: 24431500; PubMed Central PMCID: PMC3715649.
63. Cheng X, Xiong R, Yan H, Gao Y, Liu H, Wu M, et al. The trihelix family of transcription factors: functional and evolutionary analysis in Moso bamboo (*Phyllostachys edulis*). *BMC Plant Biol*. 2019;19(1):154. Epub 2019/04/27. doi: 10.1186/s12870-019-1744-8. PubMed PMID: 31023225; PubMed Central PMCID: PMC6482567.
64. Rensing SA. Gene duplication as a driver of plant morphogenetic evolution. *Curr Opin Plant Biol*. 2014;17:43 – 8. Epub 2014/02/11. doi: 10.1016/j.pbi.2013.11.002. PubMed PMID: 24507493.
65. Pallanca JE, Smirnoff N. Ascorbic acid metabolism in pea seedlings. A comparison of D-glucosone, L-sorbosone, and L-galactono-1,4-lactone as ascorbate precursors. *Plant Physiol*. 1999;120(2):453–62. Epub 1999/06/11. doi: 10.1104/pp.120.2.453. PubMed PMID: 10364396; PubMed Central PMCID: PMC59283.
66. Trotter PJ, Pedretti J, Yates R, Voelker DR. Phosphatidylserine decarboxylase 2 of *Saccharomyces cerevisiae*. Cloning and mapping of the gene, heterologous expression, and creation of the null allele. *J Biol Chem*. 1995;270(11):6071–80. Epub 1995/03/17. doi: 10.1074/jbc.270.11.6071. PubMed PMID: 7890740.
67. Thépot V, Campbell AH, Paul NA, Rimmer MA. Seaweed dietary supplements enhance the innate immune response of the mottled rabbitfish, *Siganus fuscescens*. *Fish Shellfish Immunol*. 2021;113:176–184.(doi):10.1016/j.fsi.2021.03.018. Epub Apr 3.
68. Liu X, Bogaert K, Engelen AH, Leliaert F, Roleda MY, De Clerck O. Seaweed reproductive biology: environmental and genetic controls. *Botanica Marina*. 2017;60(2):89–108.
69. Johnson DG, Walker CL. Cyclins and cell cycle checkpoints. *Annu Rev Pharmacol Toxicol*. 1999;39:295–312.(doi):10.1146/annurev.pharmtox.39.1.295.
70. Park CJ, Seo YS. Heat Shock Proteins: A Review of the Molecular Chaperones for Plant Immunity. *Plant Pathol J*. 2015;31(4):323–33. doi: 10.5423/PPJ.RW.08.2015.0150. Epub 2015 Dec 30.
71. Hüttner S, Strasser R. Endoplasmic reticulum-associated degradation of glycoproteins in plants. *Front Plant Sci*. 2012;3:67.(doi):10.3389/fpls.2012.00067. eCollection 2012.

72. Uji T, Gondaira Y, Fukuda S, Mizuta H, Saga N. Characterization and expression profiles of small heat shock proteins in the marine red alga *Pyropia yezoensis*. *Cell Stress Chaperones*. 2019;24(1):223–33. doi: 10.1007/s12192-018-00959-9. Epub 2019 Jan 10.
73. Höfler GT, But A, Hollmann F. Haloperoxidases as catalysts in organic synthesis. *Org Biomol Chem*. 2019;17(42):9267–74. doi: 10.1039/c9ob01884k.
74. Isupov MN, Dalby AR, Brindley AA, Izumi Y, Tanabe T, Murshudov GN, et al. Crystal structure of dodecameric vanadium-dependent bromoperoxidase from the red algae *Corallina officinalis*. *J Mol Biol*. 2000;299(4):1035–49. Epub 2000/06/14. doi: 10.1006/jmbi.2000.3806. PubMed PMID: 10843856.
75. Wang X, Chung KP, Lin W, Jiang L. Protein secretion in plants: conventional and unconventional pathways and new techniques. *J Exp Bot*. 2017;69(1):21–37. doi: 10.1093/jxb/erx262.
76. Suhre MH, Scheibel T, Steegborn C, Gertz M. Crystallization and preliminary X-ray diffraction analysis of proximal thread matrix protein 1 (PTMP1) from *Mytilus galloprovincialis*. *Acta Crystallogr F Struct Biol Commun*. 2014;70(Pt 6):769–72. doi: 10.1107/S2053230X14006165. Epub 2014 May 10.
77. Waite JH, Lichtenegger HC, Stucky GD, Hansma P. Exploring molecular and mechanical gradients in structural bioscaffolds. *Biochemistry*. 2004;43(24):7653–62. doi: 10.1021/bi049380h.
78. Whittaker CA, Hynes RO. Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol Biol Cell*. 2002;13(10):3369–87. doi: 10.1091/mbc.e02-05-0259.
79. Kerrison PD, Stanley MS, De Smet D, Buyle G, Hughes AD. Holding (not so) fast: surface chemistry constrains kelp bioadhesion. *European Journal of Phycology*. 2019;54(3):291–9.
80. Charoensiddhi S, Abraham RE, Su P, Zhang W. Seaweed and seaweed-derived metabolites as prebiotics. *Adv Food Nutr Res*. 2020;91:97–156.(doi):10.1016/bs.afnr.2019.10.001. Epub Dec 10.
81. Sonani RR, Rastogi RP, Patel R, Madamwar D. Recent advances in production, purification and applications of phycobiliproteins. *World J Biol Chem*. 2016;7(1):100–9. doi: 10.4331/wjbc.v7.i1.100.
82. Thépot V, Campbell AH, Rimmer MA, Jelocnik M, Johnston C, Evans B, et al. Dietary inclusion of the red seaweed *Asparagopsis taxiformis* boosts production, stimulates immune response and modulates gut microbiota in Atlantic salmon, *Salmo salar*. *Aquaculture*. 2022;546:737286. doi: <https://doi.org/10.1016/j.aquaculture.2021.737286>.
83. Félix R, Dias P, Félix C, Cerqueira T, Andrade PB, Valentão P, et al. The biotechnological potential of *Asparagopsis armata*: What is known of its chemical composition, bioactivities and current market? *Algal Research*. 2021;60:102534.
84. Pinteus S, Lemos MF, Alves C, Silva J, Pedrosa R. The marine invasive seaweeds *Asparagopsis armata* and *Sargassum muticum* as targets for greener antifouling solutions. *Science of The Total Environment*. 2021;750:141372.

Figures

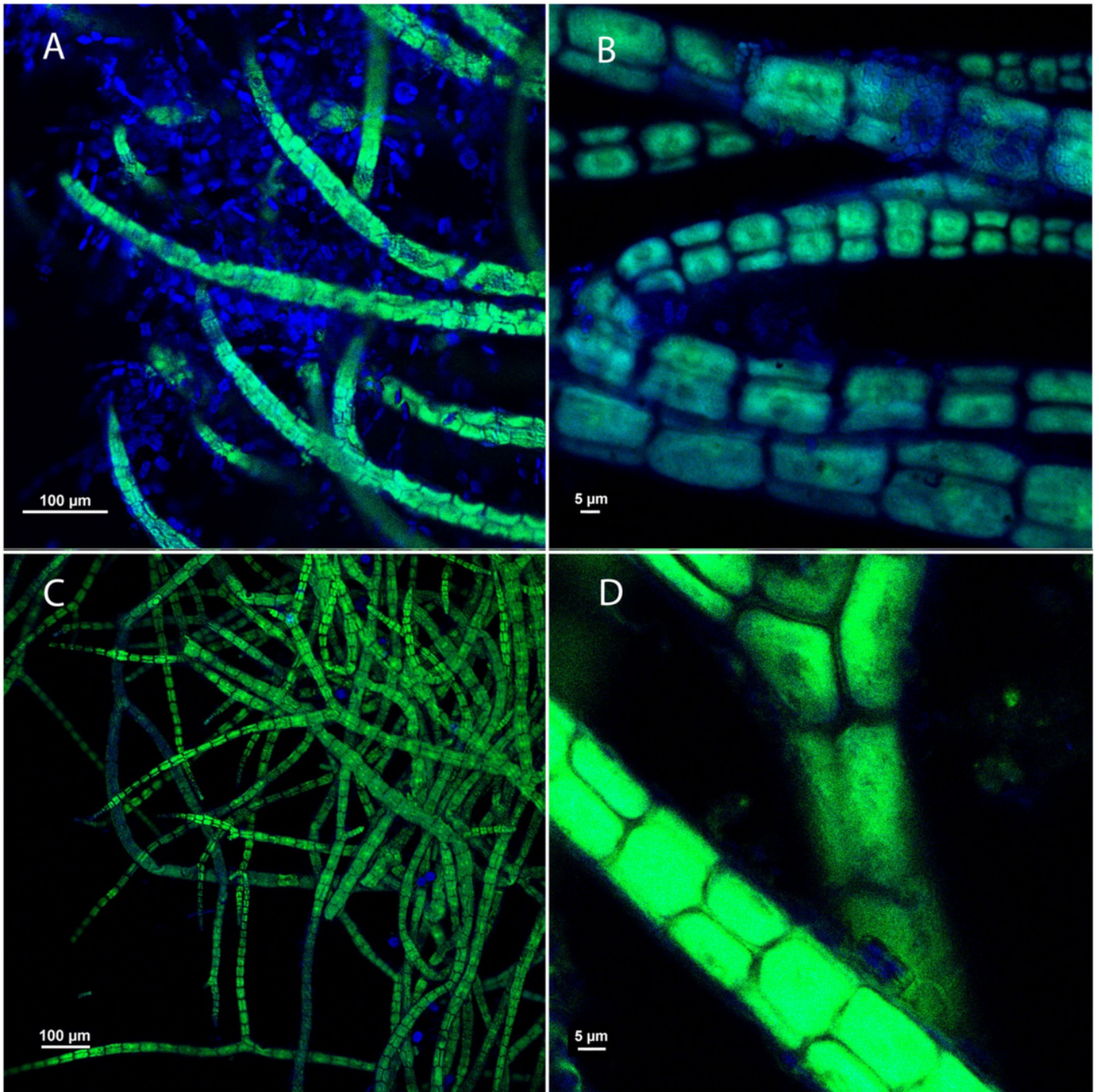


Figure 1

Laser-scanning confocal microscopy of *Asparagopsis taxiformis* (L6) and the epiphytic community. (**A**, **B**) *A. taxiformis* filaments sampled from the wild. (**C**, **D**) *A. taxiformis* filaments from culture. Seaweed filaments are shown in green and the epiphytic community is shown in blue.

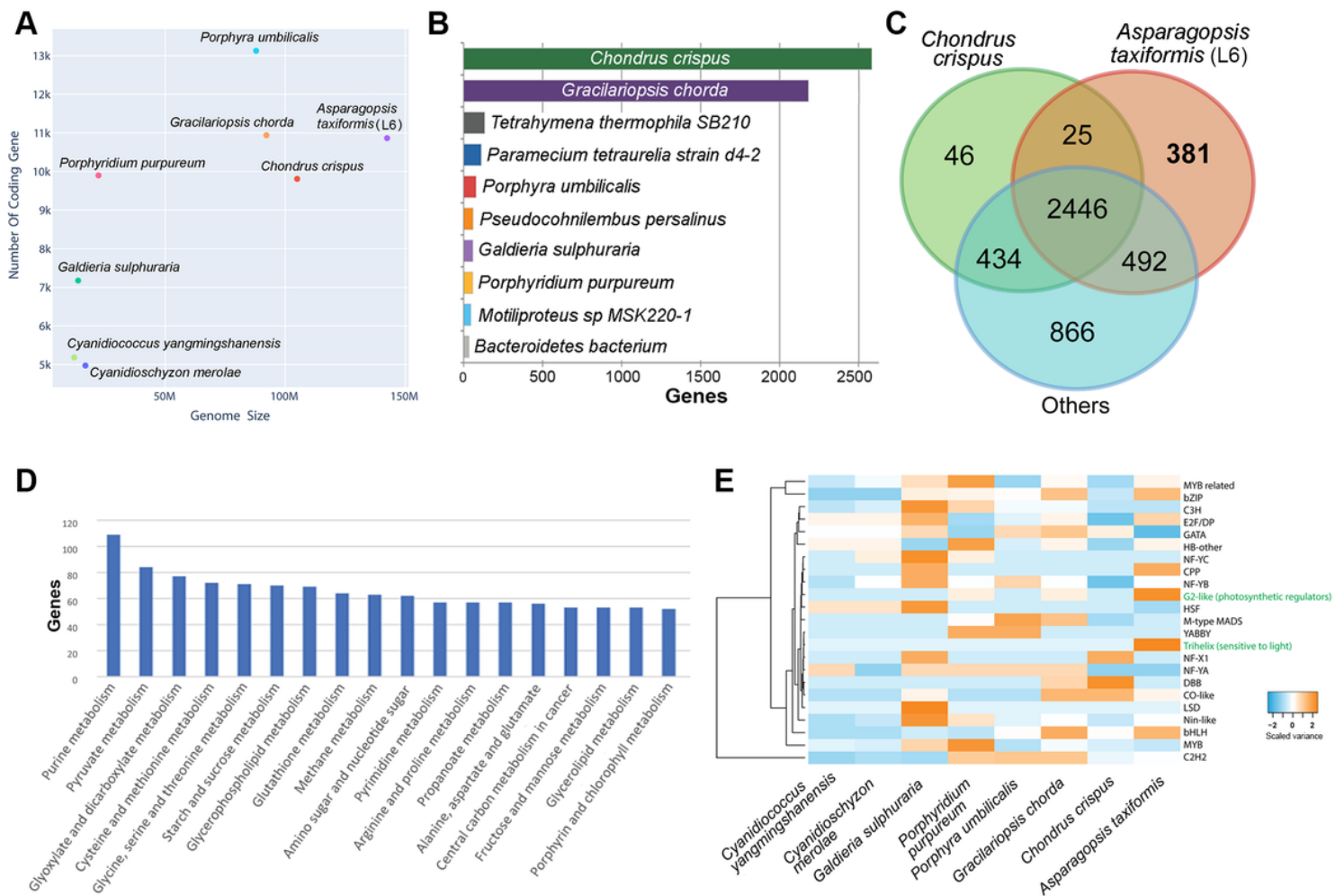


Figure 2

The *Asparagopsis taxiformis* (Lineage 6) draft genome and comparative functional overview. **(A)** Comparison of genome size and the number of protein-coding genes among 8 red algal species with completed genome and proteome data. **(B)** The species distribution based on best matched BLAST results. **(C)** Shared Pfam entries among *A. taxiformis*, *Chondrus crispus* and the remaining 6 red algae genomes. **(D)** Number of genes associated with metabolism KEGG pathways (50 or more genes). **(E)** Interspecies comparison of transcription factors, where green font represents relatively large expansion in *A. taxiformis*.

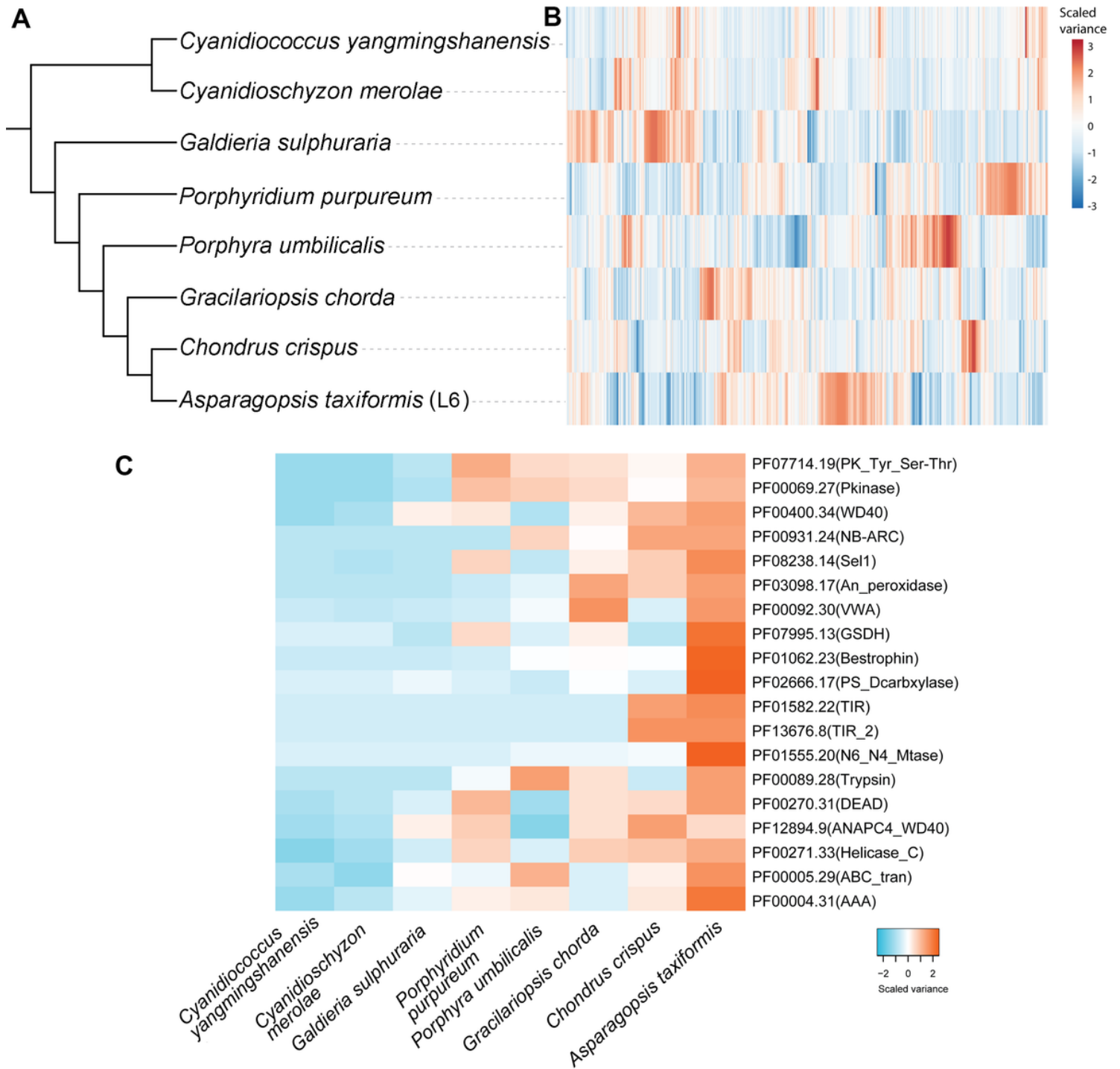


Figure 3

Comparative Pfam genomics of *Asparagopsis taxiformis*(L6) and other red seaweed genomes. **(A)** Phylogenetic tree constructed based on all the orthologous sequences inferred from OrthoFinder. **(B)** Heatmap showing the clustering of Pfam families based on the gene abundance in each species. Each row is a Pfam domain. **(C)** The top 20 Pfam domains that underwent gene duplications in *A. taxiformis* following branching from *Chondrus crispus*. The clustering in the heatmap was based on the number of genes for each Pfam entry (row) in the corresponding species (column).

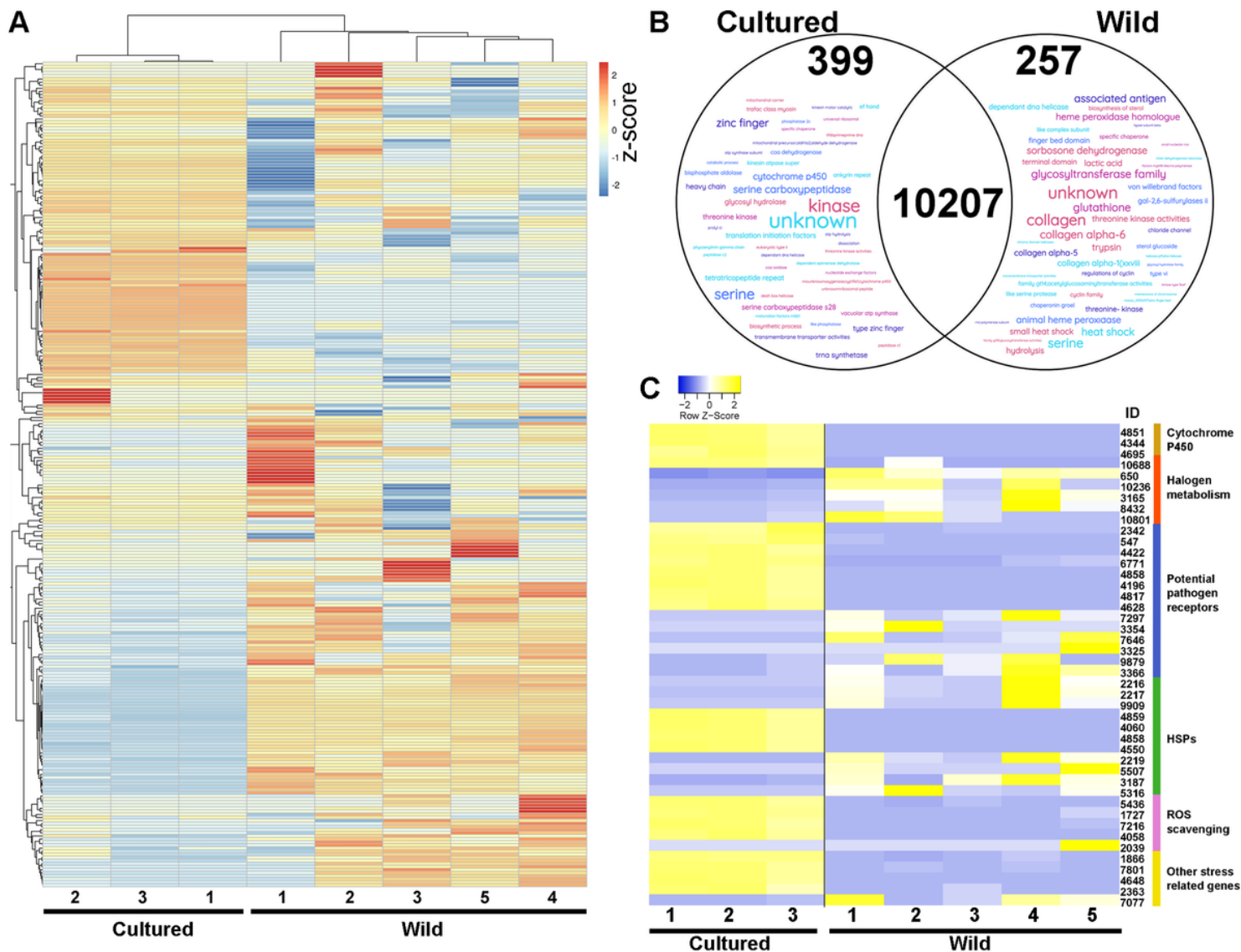


Figure 4

Summary of differential gene expression between *A. taxiformis* (L6) cultured and wild sporophytes and annotation. **(A)** Heatmap with hierarchical clustering showing relative gene expression between cultured and wild sporophytes. **(B)** Venn shows those genes with greater than 4-fold higher expression in cultured or wild sporophytes, and highlights (Word Cloud) their proposed function based on relative abundance of annotation. **(C)** Heatmap showing relative gene expression of genes significantly different between cultured and wild sporophytes that are associated with stress and resistance.

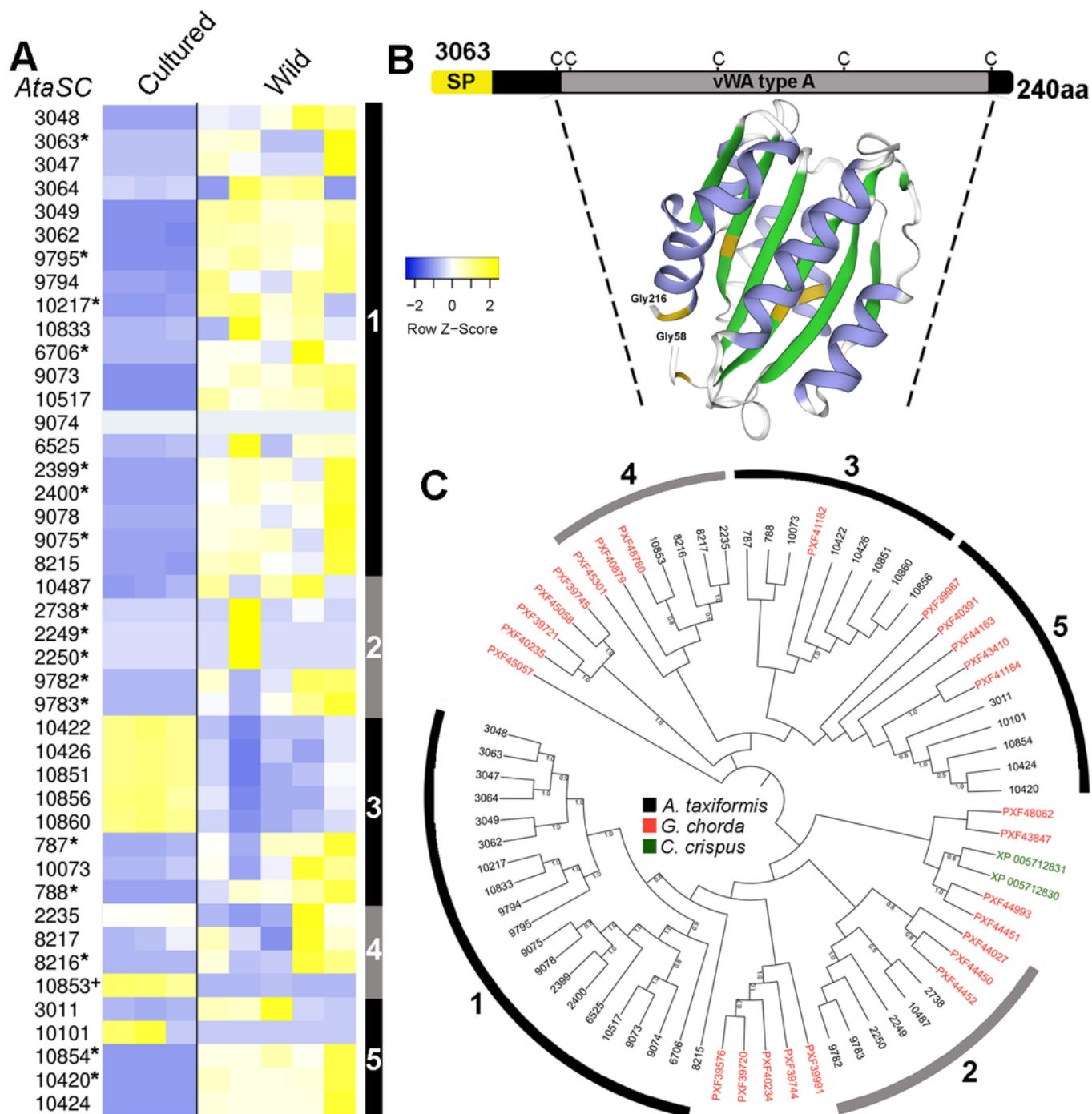


Figure 6

Identification and comparative analysis of *A. taxiformis* (L6) rhodophyte collagen-alpha-like proteins (RCAPs). **(A)** List of RCAPs and their relative gene expression (TPM log₂) in cultured and wild sporophyte. * and +, indicates significantly higher ($P < 0.05$) in wild and cultured samples, respectively, and with >4 FDR fold-change. **(B)** Schematic showing organization of *AtaL6* 3063 and structure model. SP, signal peptide; vWA type A, vWA type A domain. **(C)** Phylogenetic tree of RCAPs identified in *A. taxiformis* (L6), *G. chorda*

and *C. crispus*. Five clusters that include *A. taxiformis* (L6) are shown and correspond to that shown in (A).

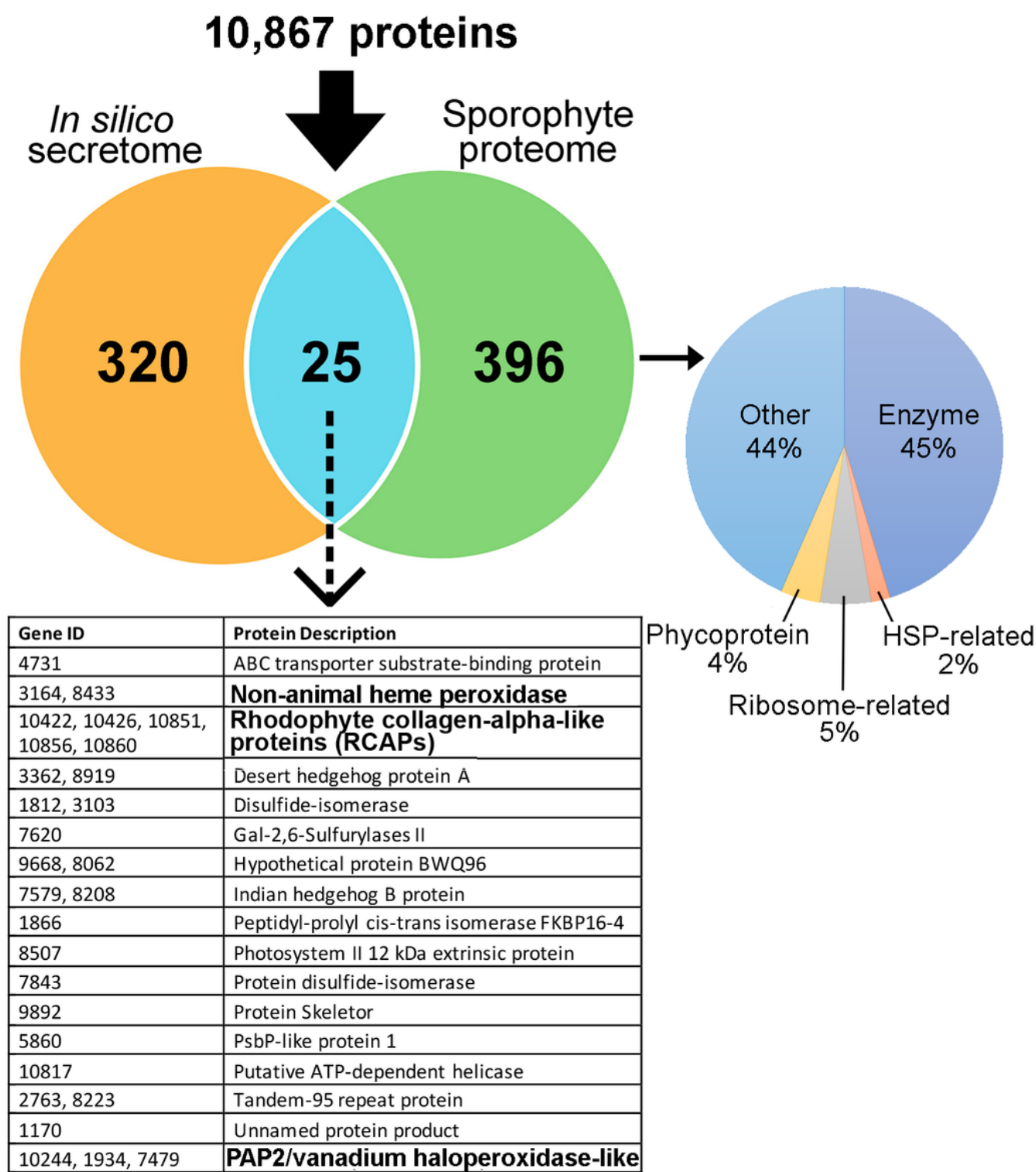


Figure 7

Venn showing *in silico* secretome of *A. taxiformis* (L6) (based on SignalP and TMHMM predictions) and sporophyte proteins identified, with identified sporophyte proteome (table below Venn).

Supplementary Files

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

- [FileS1.xlsx](#)
- [FileS2.xlsx](#)
- [FileS3.xlsx](#)
- [Supfiguresfiles1through8.docx](#)