The two oceans project: Searching for novel carbohydrate sulfoconjugation enzymes

Ricardo Prins

Supervisor: Prof J. Kossmann

Faculty of Agricultural Sciences

Department Genetics

Institute for Plant Biotechnology

March 2017

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF

Declaration

By submitting this thesis/dissertation electronically, I Ricardo Prins declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Ricardo Prins

March 2017

“It always seems impossible until it's done.”

Nelson Rolihlahla Mandela

Abstract

Marine algae contain a wide range of carbohydrate-based bioactive compounds, which promote extra-nutritional health benefits. In marine algae, the most sought after bioactive compounds are sulfoconjugated polysaccharides (SPs) which have the potential to be used as functional ingredients in the nutraceuticals industry. SPs are of particular biological interest to human health and have been shown to exert anti-inflammatory, anti-tumorigenic, immunostimulatory, and more recently anti-viral activities. The specific biological properties of SPs are attributed to their monomer composition, sulfate content, and sulfoconjugation pattern present on the sugar backbone. The biochemical mechanisms through which these polysaccharides are sulfoconjugated are well characterised in animals and are catalysed by a superfamily of enzymes termed sulfotransferases (SULTs, EC 2.8.2). In vivo, SULTs are responsible for the enzymatic process termed sulfoconjugation. This entails the transfer of a sulfuryl group (SO_3) from a donor molecule, classically 3'-phosphoadenosine 5'-phosphosulfate (PAPS), to an acceptor molecule e.g. alcohol, amine, sugar or phenolic compound.

The current state-of-the-art involves the isolation and purification of sulfoconjugated compounds from marine organisms. These compounds are then tested for various bioactivities (Hamed et al., 2015; Wijesekara et al., 2011). In contrast, here we used a high-throughput RNA-seq approach to identify 18 putative SULT transcripts and subsequently isolated one carbohydrate SULT. Eight of the SULTs identified from the *P. corallorhiza* transcriptome had a Sulfotransfer_1 domain. However, four of the eight Sulfotransfer_1 SULTs also had a Sulfotransfer_3 domain. Additionally, seven SULTs had a Sulfotransfer_2 domain, with one also possessing a Gal-3-O_sulfotr domain. The SULTs identified with overlapping domains were not exclusively assigned to any one SULT family, as they had statistically significant homology (<0.05) to either the Sulfotransfer_1 and Sulfotransfer_3 or the Sulfotransfer_2 and the Gal-3-O_sulfotr domain containing SULT entries, respectively. Additionally, this approach included the diurnal sampling of a ubiquitously distributed red algal

species, *Plocamium corallorhiza,* as a generic representative of the Rhodophyta phylum. A diurnal differential gene expression analysis was used to assess potential differences in gene expression, specifically with regard to galactan biosynthesis. Ultimately, 639 nocturnally differentially expressed genes (DEGs) were identified. However, no DEGs relating to galactan biosynthesis as it relates to starch and sucrose, galactose or sulfur metabolism pathways, were present within the 639 DEGs.

Opsomming

Mariene alge bevat ‘n wye verskeidenheid van koolhidraat gebaseerde bioaktiewe verbindings wat meer as net voedingswaarde verskaf, maar ook gesondsheids-voordele inhou. Die mees gesogte bioaktiewe verbingings in mariene alge is, gesulfo-konjugeerde polisakkariede (SPs), wat potensieel as ‘n funksionele bestandeel in die gesondsheidsindustrie gebruik kan word. SPs besit anti-inflammatoriese, anti-gewas, immuun stimileerende eienskappe, en meer. Onlangs is anti-virale bioaktivitiet ook hiervoor bewys. Die bogenoemde bioaktiwiteite maak SPs van besonderse belangvir menslike gesondheid. Die spesifieke biologiese eienskappe van SPs word toegeskryf aan hul monomeer samestelling, sulfaat-inhoud en die gesulfo-konjugeerde patroon teenwoordig op die suiker-ruggraat. Die biochemiese meganismes waardeur hierdie polisakkariede gesulfo-konjugeered word, is goed nagevors in diere en word gekataliseer deur ‘n superfamilie van ensieme, naamlik sulfotransferases (SULT, EC 2.8.2).

Die huidiglike benadering, behels die isolasie en suiwering van gesulfo-konjugeerde verbindings vanaf mariene organismes. Hierdie verbindings word dan getoets vir verskeie bio-aktiwiteite (Hamed et al., 2015; Wijesekara et al., 2011). Ons benadering poog om die ensieme, naamlik SULTs, wat verantwoordelik is vir sulfo-konjugering te identifiseer en te isoleer vanaf ‘n makro rooi alg. Hier was *RNA-seq* gebruik op *Plocamium corallorhiza*, ‘n algemene Suid-Afrikaanse rooi alg. Gevolglik was 18 SULTs geïdentifiseer en een koolhidraat SULT was geïsoleer. Agt van die SULTs geïdentifiseer uit die *P. corallorhiza* transkriptoom het 'n *Sulfotransfer*_1 domein. Alhoewel, vier van die agt *Sulfotransfer*_1 SULTs het ook 'n *Sulfotransfer*_3 domein. Daarbenewens, sewe SULTs het 'n *Sulfotransfer*_2 domein, met een wat ook in besit van 'n Gal-3-O_*sulfotr* domein. Die SULTs geïdentifiseer met oorvleuelende domeine behoort nie uitsluitlik net aan een familie nie, omdat die domeine het statistiese beduidende homologoë (<0,05) om óf die *Sulfotransfer*_1 en *Sulfotransfer*_3 of die *Sulfotransfer*_2 en die Gal-3-O_*sulfotr* domein inskrywings bevat, onderskeidelik. Daarbenewens, sluit hierdie benadering ‘n middag en middernag bemonstering in vir

geenuitdrukking analises, spesifiek ten opsigte van galaktaan biosintese. Alhoewel, geen naglikse differensieele geenuitdrukking (DEGs) van belang tot galaktaan sintese, soos dit verband hou met stysel and sukrose, galaktose and swael metabolisme, geïdentifiseer was nie, is 639 DEGs wel geïdentifiseer.

Acknowledgements

In the completion of this work, I have accumulated many debts, whether personal or professional, which I am delighted to acknowledge.

My interest in high throughput screening systems originated from conversations with the academic staff of the Institute for Plant Biotechnology (IPB). I thank Prof Jens Kossmann, Dr Shaun Peters, and Dr Bianke Loedolff for immeasurable insight, guidance, and enthusiasm. Additionally, I would like to extend a special thank you to Bianke for empathy and understanding. To my fellow IPB compadres, whether staff or student, I deeply valued the support lent through the years - no laughter, obscurities, or the friendly nature of the IPB will be forgotten.

For sticking it through the hard and unbearable times my gratitude goes to Felicia Wells. She has been the pillar that has kept this house standing. Additionally, I also must acknowledge the help of James Hunter and Felicia Wells in the diurnal sampling of the algal specimens. A special thank you to Dr Tammy Robinson, from the Stellenbosch University marine research group, for helping to identify the various sampled algae. Finally, the back and forth editing of this manuscript would not have been possible without Dr Shaun Peters – whose experience and insight has added immeasurably to this work.

Table of contents

The two oceans project: Searching for novel carbohydrate sulfoconjugation enzymes 0

Declaration 1

Abstract 3

Opsomming 5

Acknowledgements 7

Table of contents 8

List of figures and tables 11

Abbreviations 16

Chapter 1: Literature review 18

1. Marine bioactive compounds 18

1.1. Bioactive marine polysaccharides. 20

1.1.1. Bioactive marine proteins and peptides 23

1.1.2. Marine bioactive fatty acids and sterols 24

1.2. Sulfoconjugated polysaccharides from macro red algae: Their composition, biosynthesis, and most significant sources. 28

1.3. Sulfotransferases 33

1.4. Project aims and objectives 41

Chapter 2: Materials and methods 43

2.1. Diurnal sampling and identification of macro red algae 43

2.2. Determining the presence of acidic and/or sulfoconjugated polysaccharides in *P. corallorhiza* extracts. 43

2.3. RNA extractions, integrity analysis and high throughput Next-Generation Sequencing (NGS) 44

2.3.1 *P. corallorhiza* RNA extractions and quality assessment 44

2.3.2. RNA-sequencing of a midday and midnight *P. corallorhiza* sample 45

2.4. *P. corallorhiza* RNA-seq data pre-processing, transcriptome assembly and diurnal differential gene expression analysis. 45

2.4.1. RNA-seq pre-processing and transcriptome assembly 45

2.4.2. Diurnal differential gene expression analysis and visualisation 45

2.5. In silico transcriptome annotation using sequence homology, protein domain information and Gene Ontology (GO) term allocation. 46

2.6. Carbohydrate SULT transcript identification and cloning. 47

Chapter 3: Results 49

3.1. *P. corallorhiza* as a generic representative of the Rhodophyta phylum, for SULT identification 49

3.2. RNA quantification and integrity analysis 51

3.3. RNA pre-processing, transcriptome assembly and annotation statistics 52

3.4. Analysis of the draft *P. corallorhiza* transcriptome: metabolic pathways associated with algal SPs biosynthesis. 55

3.5. Mining of annotated contigs for sulfotransferase transcript identification 62

3.6. ORF cloning of a *P. corallorhiza* glycolipid SULT (GLST1) and a putative carbohydrate SULT (PCHST1), grouping to the sulfotransfer_1 and _2 families, respectively 64

3.7. Diurnal differential gene expression analysis: Single gene and gene network analysis. 65

Chapter 4: Discussion 72

4.1. *Plocamium corallorhiza*, the generic Rhodophyta species used for SULTs identification, accumulates sulfoconjugated polysaccharides. 72

4.2. Assembling the transcriptome of *P. corallorhiza.* 73

4.3. In silico transcriptome annotation using sequence homology, protein domain

information and Gene Ontology (GO) term assignment. 75

4.4. Putative sulfotransferase transcript identification and subsequent cloning and validation. 77

4.5. Potential differences in diurnal gene expression estimates: Analysis of differentially expressed genes (DEGs) at gene and gene network level. 80

Chapter 5: Conclusion 83

Literature cited 84

Appendix A 102

List of figures and tables

Chapter 1: Literature review

Figure 1. The phylogenetic relationship among marine organisms used for their natural products, as domesticated foods and sources of patented DNA sequences, obtained from Arrieta et al. (2010). 19

Figure 2. Sulfoconjugated polysaccharides found in different marine phyla, their principal monomer constituents, and proposed phylogenetic relationships. Highlighted in dark grey is the structure 3-β-D-galactopyranose, which is highly conserved among marine organisms that accumulate SPs. Further, the different types of sulfoconjugation i.e. 2-, 4- and 6-substitution, respectively is highlighted by light grey ellipses (Pomin and Mourão, 2008). 30

Figure 3. The proposed biochemical pathway for agar biosynthesis in Rhodophyceae. Floridean starch that was accumulated through photosynthesis provides the precursor sugars for algal SPs biosynthesis. Floridean starch is degraded into linear alpha 1, 4-glucans through the actions of isoamylase, glucan water dikinase and amylopullulanase - this breakdown is reversible. The formed linear glucans are subsequently converted into glucose and glucose-1-phosphate. Through various enzymatic conversions and modifications such as isomerization, epimerization, and nucleotide sugar formation, GDP-L-galactose and UDP-D-galactose the principal algal SPs constituents are formed. GDP and UDP-galactosyltransferase polymerises GDP-L-galactose and UDP-D-galactose monomers to form the sugar backbone of agar. Either prior or post agar backbone synthesis, chemical group modifications such as sulfoconjugation is added onto the agar through the actions of sulfotransferases. Present sulfate groups are removed from the agar backbone through sulfatase, forming very characteristic sulfoconjugation patterns. Enzymes identified in this study are shaded grey. Figure adapted from Barbier et al. (2005) and Hector (2013). 32

Figure 4. The general sulfoconjugation reaction as enzymatically catalysed by sulfotransferases. The sulfuryl group (orange) is transferred from the universal sulfate donor, classically 3'-phosphoadenosine 5'-phosphosulfate (PAPS), to an appropriate acceptor molecule, such as compounds containing either a hydroxyl or an amine group (blue). The by-product of the sulfoconjugation reaction is 3'-phosphoadenosine-5'-phosphate (PAP). Diagram obtained from Prather et al.(2012). 34

Figure 5. Conserved protein motifs commonly found in sulfotransferase enzymes. Based on studies of *A. thaliana* SULTs and structural research conducted on *Mus musculus*

SULTs, conserved protein motifs are illustrated. PAPS binding (regions I, II, and IV), a histidine residue responsible for proton acceptance and region III, of unknown function. The characteristically conserved motifs are indicated by boxes, with the position and size proportionally represented to average *A. thaliana* SULTs. Further, the SULT catalytic domain is indicated in grey along the gene length. Figure adapted from Hirschmann et al. (2014). ..35

Figure 6. The enzymatic transfer of a sulfuryl group from the donor molecule, PAPS, to an acceptor molecule, ROH or RNH2, by either a cytosolic or membrane-associated SULT. The SULT substrate is represented by an R and depending on cellular localisation can denote a carbohydrate, peptide, or xenobiotic compound containing a hydroxyl or amine group, respectively. The sulfate donor i.e. PAPS is biosynthesized in the cytosol sequentially by ATP sulfurylase and APS kinase. Subsequently, PAPS is transported into the Golgi lumen by the transport protein, PAPS translocase. The figure is adapted from http://what-when-how.com/molecular-biology/sulfation-molecular-biology.....37

Table 1. Summary of marine bioactive compounds, their sources, and bioactivities. .26

Table 2. Sulfotransferase genes currently identified and biochemically characterised. Organisms included are terrestrial animals and plants, algae and prokaryotes.....39

Chapter 3: Results

Figure 7. The diurnal sampling of macro algae from shallow intertidal zones on the west coast of South Africa. Top row (left to right), *P. corallorhiza*, *hypnea* sp., *Codium* sp.; Bottom row (left to right), *Caulerpa filiformis*, *Sargassum* sp., unidentified algal specimen...50

Figure 8. Agarose gel electrophoresis post-stained with toluidine blue O, demonstrate the presence of acidic and/ or sulfoconjugated polysaccharides in the alga *P. corallorhiza* & model system *A. thaliana*. 1. Heparan sulfate, 2. Sulfoconjugated arabinogalactan, 3. Non-sulfoconjugated galactan, 4. *P. corallorhiza* whole algal extract, 5. *A. thaliana* leaf extract, 6. *P. corallorhiza* SPs, 7. *P. corallorhiza* desulfonated SPs. 50

Figure 9. *P. corallorhiza* RNA integrity analysis of both a day and night time sample. Formaldehyde agarose gel electrophoresis and Agilent bioanalyser 2100 analysis showing 18S and 28S rRNA fragments with accompanying RNA integrity number (RIN), sample concentration, and 28S/18S rRNA ratios. Fluorescence (FU) and nucleotides (nt). *(B.02.08, Anomaly Threshold(s) manually adapted). 51

Figure 10. Analyses of contiguous assemblies. The distribution of 119 675 de novo assembled contigs covering 65.46 Mb, using a k-mer size of 24. 52
Figure 11. Annotations of contiguous assemblies through Gene Ontology (GO) term transfers. *P. corallorhiza* contig subset ≥ 900 was annotated through the four-step analysis in the BLAST2GO pipeline including BLASTx, InterProScan, mapping, and annotation. Annotations were refined by annex and GOslim execution. 54

Figure 12. The KEGG starch and sucrose metabolic pathway with Enzyme Code (EC) assigned to contigs identified from thallus RNA. Coloured blocks represent enzymes encoded in specific contigs, with the EC to which the contig mapped. 57

Figure 13. The KEGG galactose metabolic pathway with Enzyme Code (EC) assigned to contigs identified from thallus RNA. Coloured blocks represent enzymes encoded in specific contigs, with the EC to which the contig mapped. 59

Figure 14. The KEGG sulfur metabolic pathway with Enzyme Code (EC) assigned to contigs identified from thallus RNA. Coloured blocks represent enzymes encoded in specific contigs, with the EC to which the contig mapped. 61

Figure 15. PCR amplification and isolation of *P. corallorhiza* SULT transcripts, GLST1 (1029 bp) and PCHST1 (1146 bp). PCR screening of colonies containing GLST1::pJet1.2 or PCHST1::pJet1.2 plasmid construct. 64

Figure 16. Directed acyclic graph of the Fisher's exact test for enriched Gene Ontology (GO) terms present on the nocturnally downregulated DEGs. Assessment of the GO terms for molecular function in the diurnal differentially expressed genes compared to all GO terms for all the transcripts. The Fishers Exact Test was used with a FDR corrected p-value cut-off ≤ 0.01. 68

Figure 17. Directed acyclic graph of the Fisher's exact test for enriched Gene Ontology (GO) terms present on the nocturnally upregulated DEGs. Assessment of the GO terms for molecular function in the diurnal differentially expressed genes compared to all GO terms for all the transcripts. The Fishers Exact Test was used with a FDR corrected p-value cut-off ≤ 0.01. 69

Figure 18. Directed acyclic graph of the Fisher's exact test for enriched Gene Ontology (GO) terms present on the nocturnally upregulated DEGs. Assessment of the GO terms for cellular components in the diurnal differentially expressed genes compared to

all GO terms for all the transcripts. The Fishers Exact Test was used with a FDR corrected p-value cut-off ≤ 0.01. 70

Figure 19. Directed acyclic graph of the Fisher's exact test for enriched Gene Ontology (GO) terms present on the nocturnally upregulated DEGs. Assessment of the GO terms for biological processes in the diurnal differentially expressed genes compared to all GO terms for all the transcripts. The Fishers Exact Test was used with a FDR corrected p-value cut-off ≤ 0.01 71

Table 3. Summary data for Illumina RNA-sequencing, pre-processing, de novo transcriptome assembly, and evaluation. 52

Table 4. Selected metabolic pathways associated with the *P. corallorhiza* transcriptome identified through KEGG pathway analysis. 55

Table 5. Putative Sulfotransferase transcripts identified. A subset of contigs identified as SULTs through BLAST2GO annotation pipeline of the de novo assembled *P. corallorhiza* transcriptome. 62

Table 6. Transcripts with the highest log2 fold change in the RNA-seq differential expression analysis for *P. corallorhiza* diurnal samples. Diurnal gene expression analyses were performed with Wald statistical analysis (log2 Fold change ≥ 5 or ≤ -5, with the Bonferroni and False Discovery Rates (FDR) corrected p-values ≤ 0.01). Transcripts identified using above criteria were considered as favourable candidates for potentially statistically significant DEGs. 65

Appendix A

Figure 20. BLASTx hits with regard to species distribution for assembled contigs ≥ 900. (A) Total species distribution derived from BLASTx alignments **(B)** Top-hit species distribution for all BLASTx hits used for annotation inferences. 103

Figure 21. GC-content distribution of the *P. corallorhiza* transcriptome. The X-axis represents the relative GC-content in percentages of a contig. The Y-axis represents the number of contigs containing particular GC- percentages normalised to the total number of contigs. 103

Figure 22. *Coverage over GC content analysis of the P. corallorhiza draft transcriptome..* 104

Figure 23. GO term annotation distribution from BLAST2GO pipeline for *P. corallorhiza* transcriptome. GO terms relating to biological processes, molecular functions and cellular components were assigned to the *P. corallorhiza* transcriptome using the BLAST2GO pipeline, including BLASTx and InterProScan analyses. 105

Figure 24. Volcano plot visually presenting the DEGs for the night compared to day RNA-seq sample. DEGs (red dots) were identified through Wald statistical analysis (log2 Fold change ≥ 5 or ≤ -5, with the Bonferroni and False Discovery Rates (FDR), corrected p-values ≤ 0.01). **(A)** Upregulated transcripts (red dots) in the Night compared to the Day temporal sample. **(B)** Downregulated transcripts in the Night when compared to the day temporal sample 107

Abbreviations

AMP	Adenosine monophosphate
APS	Adenylyl sulfate
ARC-BTP	Agricultural research council's biotechnology platform
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
CAF	Central analytical facility
cDNA	Complementary deoxyribonucleic acid
CVD	Cardiovascular disease
DE	Differentially expressed
DEG	Differentially expressed gene
DHA	Docosahexaenoic acid
DNA	Deoxyribonucleic acid
EBI	European bioinformatics institute
EC	Enzyme codes
EPA	Eicosapentaenoic acid
EST	Expressed sequence tag
FDR	False discovery rates
FU	Fluorescence
g	Earth's gravitational force
Gb	Gigabase
GC	Guanine-cytosine
GDP	Guanosine Diphosphate
GIT	Gastrointestinal tract
GLST1	Glycolipid sulfotransferase 1
GO	Gene ontology
GSEA	Gene set enrichment analysis
GTP	Guanosine triphosphate
HIV-1	Human immunodeficiency virus type 1
HMM	Hidden Markov models
HSV-1	Herpes simplex virus type 1
IPB	Institute for Plant Biotechnology
KEGG	Kyoto Encyclopedia of genes and genomes
L	Litre
LB	Luria-Bertani
min	Minute

mL	Milli-litre
NADH	Nicotinamide adenine dinucleotide
NCBI	National center for biotechnology information
ng	Nanogram
NGS	Next-generation sequencing
nr	Non-redundant protein database
NRF	National Research Foundation
nt	Nucleotides
ORF	Open reading frame
PAP	3’-phosphoadenosine-5’-phosphate
PAPS	3'-Phosphoadenosine-5'-phosphosulfate
PCHST1	Putative carbohydrate sulfotransferase 1
PCR	Polymerase chain reaction
Poly-Gluc	Poly-glutamic acid
PPi	Pyrophosphate
RIN	RNA integrity number
RNA	Ribonucleic acid
SA	South Africa
SP	Sulfoconjugated polysaccharide
SULT	Sulfotransferase
TBE	Tris/borate/EDTA
µg	Micro-gram
µL	Micro-Litre
TCA	Tricarboxylic acid
UDP	Uridine diphosphate
US$	United States dollar
USA	United States of America
UTP	Uridine triphosphate
UV	Ultraviolet

Chapter 1: Literature review

1. Marine bioactive compounds.

Approximately 70% of the earth consists of oceans, the majority of which are unexplored. Within these environments, marine organisms encounter diverse environmental conditions, often leading to interesting and unique metabolic adaptations (including metabolic processes and chemical compounds) which have no terrestrial equivalent (Hamed et al., 2015). These organismal adaptations coupled with high biodiversity makes the marine environments an attractive proposition for the discovery and development of novel bioactive compounds. Presently, the use of terrestrial plants as sources of medicinal compounds supports approximately 80% of the global population. However, the increasing rate of occurrence of human diseases such as cancer (Talero et al., 2015), infectious deceases (Boucher et al., 2009; Spellberg et al., 2004), diabetes (Lordan et al., 2011) and cardiovascular disease (Lordan et al., 2011; Rangel-Huerta et al., 2015; Talero et al., 2015) is fuelling the need for the discovery and development of new and improved drugs from other natural sources. Collectively these factors are guiding contemporary research to identify and characterise naturally occurring bioactive compounds, from the marine environment (Itokawa et al., 2008; Plaza et al., 2008; Schaeffer and Krylov, 2000).

In this regard, the use of marine species by humans (direct consumption, aquaculture, and biotechnological applications) is growing at an unprecedented rate. These is evident in the number of marine natural products discovered and even more so with the number of patents associated with marine genetic resources (Figure 1). Using natural product databases and the NCBI's patent database, Arrieta et al. (2010) estimate over 18 000 marine natural products as of 2010, growing annually at 4% and marine gene patents at 4 900, growing annually at 12%. Remarkably, both the rates for marine natural product discovery and marine gene patents are growing faster than marine species discovery i.e. almost 1% annually (Arrieta et al., 2010). The number of

known marine organisms from which bioactive compounds have been isolated is numerous.

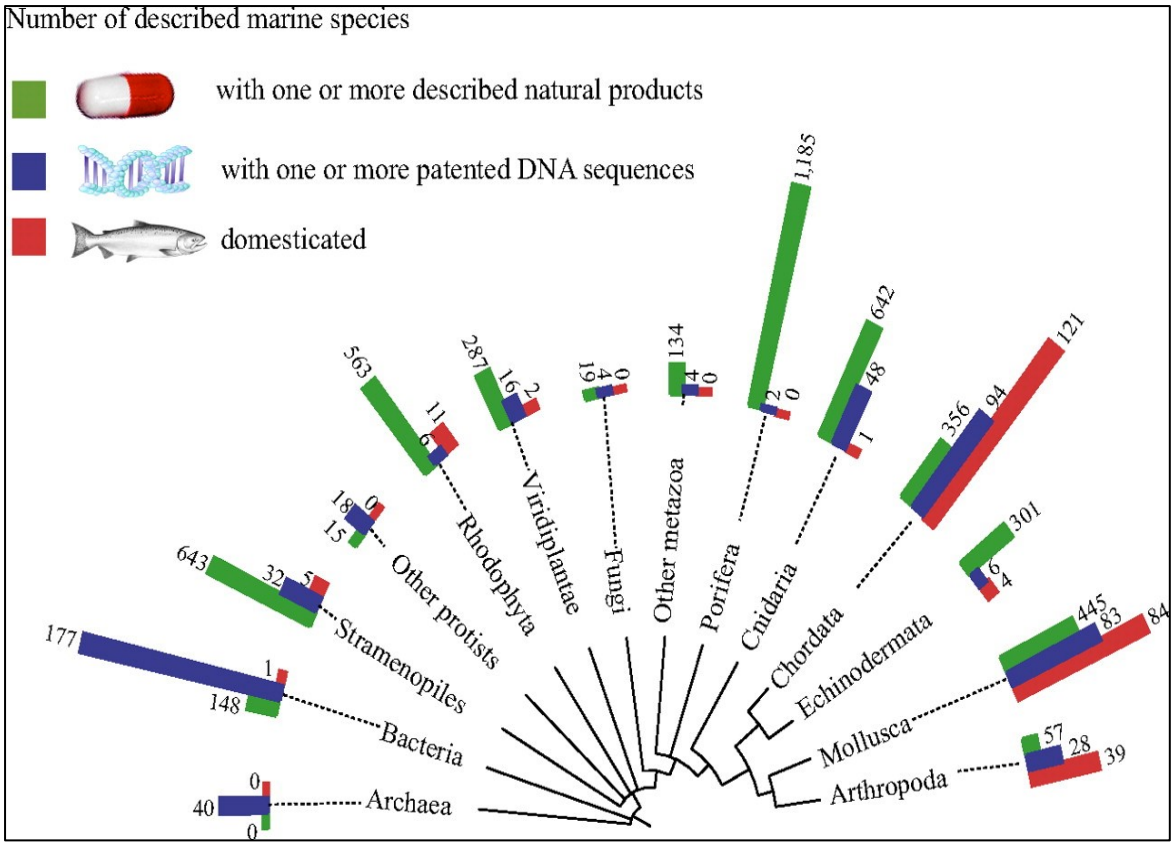


Figure 1. The phylogenetic relationship among marine organisms used for their natural products, as domesticated foods and sources of patented DNA sequences, obtained from Arrieta et al. (2010).

Compounds displaying anti-viral, antioxidant, anti-tumorigenic, anti-dementia, antimicrobial, immunostimulatory, immunomodulatory and anti-inflammatory bioactivities have been isolated from crustaceans, fish, micro- and macro-algae and lower organisms such as cyanobacteria (Arrieta et al., 2010; Hamed et al., 2015; Harnedy and Fitzgerald, 2011; Wijesekara et al., 2011). The chemical makeup of these bioactive compounds are diverse and this review will briefly discuss marine bioactive compounds such as polysaccharides, oligosaccharides, proteins, peptides, fatty acids and sterols.

1.1. Bioactive marine polysaccharides.

Bioactive polysaccharides are found in both terrestrial and marine organisms (Dantas-Santos et al., 2012; Lehmann and Robin, 2007). Heparin, a proteoglycan and a frequently used anticoagulant drug is sourced from terrestrial animal tissue while macro algal polysaccharides have been used for decades in the food and pharmacology industries. It should be noted that polysaccharides derived from the marine environment often have chemical modifications such as sulfoconjugation, acetylation, methylation, and pyruvilation (Pomin and Mourão, 2008). Among the most frequently isolated marine bioactive polysaccharides, it is specifically sulfoconjugated polysaccharides (SPs) that display multiple and potent bioactivities (Wijesekara et al., 2011). Marine macro algae represent the most significant source of non-animal SPs, largely because of their abundance, ease of extraction and physicochemical properties, making them applicable in the food industry as an emulsifying agent and functional ingredient. In addition, their bioactive properties make them applicable in the pharmacology industry, with application in human and animal health. The negative health implications associated with animal-sourced SPs, such as the USA heparin contamination case of 2008, which led to the death of 16 people and hundreds of severe adverse drug reactions, are encouraging the search for natural and artificially synthesised alternative bioactive SPs (Bogdanich, 2008; Silbert and Sugumaran, 2002).

The global market value for algal SPs i.e. agars, alginates and carrageenans, is estimated at US$ 1 billion. The three major divisions of algae, that is Chlorophyta (green algae), Phaeophyta (brown algae) and Rhodophyta (red algae), largely accumulates different types of SPs, all with different global market demands (Pomin and Mourão, 2008). Green algae accumulate the SPs ulvans, along with other carbohydrates such as starch, xylans and mannans. Other SPs such as alginates, fucans and laminarans are accumulated by brown algae. Red algae accumulate sulfoconjugated galactans such as agarans and carrageenans, alongside floridoside and Floridean starch. Cellulose is commonly found in all three major algal divisions

(Hamed et al., 2015; Viola et al., 2001). SPs extracted from the green alga, *Ulva conglobate* had anticoagulant bioactivity (Mao et al., 2006). Other SPs (fucoidans) isolated from the brown macro alga, *Laminaria japonica,* displayed antioxidant and anticoagulant effects (Wang et al., 2010; Wijesekara et al., 2011). In *U. conglobate* SPs, the major constituent sugars were rhamnose and glucose, while in *L. japonica* fucoidans, fucose was the major constituent sugar (Kitazawa et al., 2013; Wijesekara et al., 2011). Bioactive SPs from the red alga, *Cryptonemia crenulata*, with galactose as the major sugar constituent and *Nemalion helminthoides*, where mannose was the primary monomer, both displayed anti-viral bioactivity against the Herpes simplex virus type 1 (HSV-1) (Wijesekara et al., 2011). Other bioactive marine polysaccharides include the SPs, calcium spirulan, extracted from the cyanobacterium, *Arthrospira platensis,* which showed anti-viral bioactivity towards HSV-1 and the Human Immunodeficiency Virus type 1 (HIV-1) (Hamed et al., 2015; Hayashi et al., 1996). Marine shellfish and crustaceans such as shrimps, crabs, lobster, and krill are also known to contain bioactive polysaccharides. Chitin is abundantly found in nature, commonly in fungi and insects and in the shells of crustaceans and shellfish. Chitosan and chitooligosaccharides (derivatives of chitin) are extensively used in the food industry as an edible protective film on food, a food additive and in pharmacology as an antimicrobial and anti-obesity compound (Hayes and Tiwari, 2015).

Further, the prebiotic effects of marine bioactive polysaccharides make them a very valued product, with regard to functional foods (Hayes and Tiwari, 2015). The term “functional foods” was first introduced in the 1980s in Japan, because of a greater understanding of the links between food consumption and human health. Functional foods can be defined as any food product fortified to have extra nutritional value beyond sustenance, leading to beneficial physiological effects. Dietary fibre is possibly the prototypic bioactive carbohydrates and prebiotic, owing to its long-standing use and the extensive research done on it (Brownlee et al., 2005; Dawczynski et al., 2007; van der Kamp et al., 2004). Here, edible algae, also termed seaweeds, are commonly used as a functional food alongside chitosan and chitooligosaccharides (Hayes and

Tiwari, 2015). Structurally the most important feature of prebiotic polysaccharides is resistance to digestion in the upper gastrointestinal tract (GIT). Followed by relocation to the large intestine and selective breakdown by the beneficial GIT microbiota. In the large intestine, these prebiotics are transformed to bioactive compounds or alternatively they act to promote the growth of selective beneficial microbiota, like *Lactobacillus* spp. and *Bifidobacterium* spp. at the expense of detrimental microorganisms (Schieber, 2012; Ernst and Magnani, 2009).

Again, outside of the functional food arena, the medical applications of bioactive polysaccharides include, but are not limited to anti-inflammatory, anti-tumorigenic, immunostimulatory, anti-ulcer and more recently anti-viral effects (Hamed et al., 2015; Wagner, H and Kraus, 2000; Wijesekara et al., 2011). Biologically active polysaccharides isolated from different marine organisms display varying biological activities. Both overlapping and distinct bioactivities have been observed for marine bioactive polysaccharides with diverse structural features. This, coupled with experimental data, suggest that not only is the polysaccharide backbone structure important i.e. major constituent sugar and types of glycosidic bonds, but also the other chemical modifications present on the sugar backbone. According to Jiao et al. (2011), the polymer backbone structure, molecular weight, constituent monomers and patterns of chemical modifications such as sulfoconjugation, acetylation and benzoylation of marine bioactive polysaccharides determines their bioactivity and specificity. In this regard, three chemical derivatives were produced from *L. japonica* fucoidans i.e. oversulfoconjugated, acetylated and benzoylated fucoidans (Wang et al., 2009). The three derivatives were tested in vitro for their antioxidant activities. All three products had strong antioxidant activities; however, the benzoylated product showed stronger scavenging activities towards superoxide and hydroxyl radicals. The acetylated derivative produced the strongest scavenging activity towards hydroxyl and 1,1-diphenyl-2-picrylhydrazy radicals. In general, the acetylated and benzoylated fucoidans had more potent antioxidant effects than the oversulfoconjugated fucoidans in certain tests, leading the authors to conclude that the differences in antioxidant

effects observed are linked to the substitution groups on the fucoidan backbone (Wang et al., 2009). Additionally, Ghosh et al. (2009) demonstrated how different SPs inhibits the HSV-1 strain, based on different structure-activity relationships. When a 3-O-sulfoconjugated octasaccharide, derived from heparin, was tested for its anti-viral activities, it was more effective at blocking HSV-1 infection than its counterpart, a 3-OH octasaccharide (Ghosh et al., 2009). The study suggested that a specific sulfoconjugation pattern is needed to effectively block HSV-1 infection. It is evident that modifying substitution groups often leads to altered bioactivities. At our current rate of discovery, it is estimated that another 250 to 1000 years will be needed for a complete register of all marine species (Arrieta et al., 2010). Further, noting the structural diversity and backbone modifications possible for marine bioactive SPs from different species, the discovery of novel marine SPs for the treatment of current and emerging human deceases is very likely in years to come.

1.1.1. Bioactive marine proteins and peptides.

Proteins and peptides derived from marine organisms display various bioactivities (Hamed et al., 2015). Proteins are linear polymers made up of amino acids encoded within the genetic code of an organism, along with other amino acids linked by peptide bonds. (Hamed et al., 2015; Tobergte and Curtis, 2013). Proteins and peptides are required nutritionally by humans, however, they also play pivotal physiological roles e.g. acting as hormones, transporters and structural components (Hamed et al., 2015). Rich sources of proteins and peptides from the marine environment include fish, shellfish, seaweeds and microalgae, collectively termed seafood (Hamed et al., 2015). Seafood is an outstanding source of proteins and peptides, containing generally all the essential amino acids in relatively the correct proportions for human consumption. To this end, both micro and macro algae represent good candidates for inclusion into functional foods, because of their protein content and amino acid profile. However, extreme variations in the protein content and amino acid profile for algae can be observed between phyla and species (Harnedy and Fitzgerald, 2011). Commonly, red algae have a higher protein content than brown and green algae and even

conventional high-protein foods such as soybean, eggs, fish and cereals (Harnedy and Fitzgerald, 2011). In addition to the nutritional value of marine-derived proteins, some also displayed bioactivities such as antihypertensive, antioxidant, anticoagulant, and anti-tumorigenic activities. Phycobilliprotein components, such as allophycocyanin from the cyanobacterium, *Anacystis nidulans*, and phyco-erythrobilin from the red algae genus *Porphyra* showed potent antioxidant bioactivities (Ge et al., 2006; Yabuta et al., 2010). Aneiros & Garateix (2004), on the pharmacological applications of bioactive marine peptides, showed that peptides obtained from diverse marine species had numerous applications in the pharmacology industries. Depending on the amino acid composition and sequence, bioactive peptides generally consist of 3 to 20 amino acid residues. These peptides do not display any bioactivity unless released from the source protein, through either digestion or hydrolysis. (Hayes, 2012). Protein hydrolysates from cuttlefish processing wastewaters, derived through enzymatic hydrolysis, showed antihypertensive and antioxidant activities (Amado et al., 2013) Similarly, Parasin I and Pelteobagrin, peptides isolated from catfish (*Parasilurus asotus and Pelteobagrus fulvidraco*), displayed antihypertensive and antimicrobial activities (Hamed et al., 2015; Park et al., 1998; Su, 2011).

1.1.2. Marine bioactive fatty acids and sterols.

The lipid profile of seafood is generally characterised by a low saturated fatty acid content, considered as beneficial for human health (Hamed et al., 2015). Therefore, seafood is commonly regarded as a healthy alternative, because of the assumption that increased saturated fatty acid intake may lead to the development of cardiovascular disease (CVD). However, modern science is challenging this claim and hypothesise that increasing sugar consumption might be responsible for the rising CVD rate (Hamed et al., 2015; Hooper et al., 2015; Leslie, 2016; Lustig et al., 2012). Nevertheless, polyunsaturated fatty acids (PUFAs), which are essential for human health, is the main lipid constituent of seafood. Humans are unable to synthesise longer chain PUFAs i.e. PUFAs comprised of 18 or more carbons, and thus nutritionally require them. Different longer chain PUFAs that are commonly found in

seafood include oleic acid, hexadecatrienoic, omega-3 fatty acids such as eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA), in addition to omega-6 fatty acids such as arachidonic acid and linoleic acid (Hamed et al., 2015).

Health benefits attributed to bioactive PUFAs include risk reduction for CVD, arthritis, diabetes and obesity development. The regulation of blood clotting, blood pressure and functional brain and nervous system development (Hamed et al., 2015). Further, EPA and DHA have inhibitory effects on key enzymes in the metabolic process of lipid synthesis and they enhance lipid breakdown. As obesity becomes a global epidemic, Li et al. (2008) discuss how the obesity alleviating and preventative effects of EPA and DHA can be used in this ongoing struggle. Sterols are a class of membrane lipids produced by both eukaryotes and certain bacteria and are commonly found in seafood, such as crab, algae, sea urchins and fish (Hamed et al., 2015). Bioactivities displayed by marine sterols and their derivatives include antioxidant, anti-tumorigenic and antifungal effects (Carmely et al., 1989; Hamed et al., 2015; Lee et al., 2004). Well researched marine bioactive sterols include fucosterol and eryloside A, respectively extracted from the marine sponges, *Erylus lendenfeldi* and *Erylus formosus* (Lee et al., 2004; Stead et al., 2000). A copious number of important bioactive compounds have been isolated from marine sources using bioactivity based assay techniques (Table 1).

Table 1. Summary of marine bioactive compounds, their sources, and bioactivities.

COMPOUND	SOURCE	BIOACTIVITY	LITERATURE
POLYSACCHARIDES			
CHITIN AND CHITOSAN	Crustacean shells	Antimicrobial Anti-obesity Anti-tumorigenic	(Hamed et al., 2015; Hayes and Tiwari, 2015; Xia et al., 2010)
SPS	*Porphyra haitanensis* (red alga)	Antioxidant Anticoagulant	(Zhang et al., 2010)
SPS	*Gymnogongrus griffithsiae* (red alga)	Anti-viral (HSV-1)	(Lee et al., 2010; Wijesekara et al., 2011)
SPS	*Ulva conglobata* (green alga)	Anticoagulant	(Mao et al., 2006)
SPS	*Laminaria japonica* (brown alga)	Antioxidant Anticoagulant	(Wang et al., 2010; Wijesekara et al., 2011)
SPS	*Spirulina platensis* (cyanobacterium)	Anti-viral (HSV-1 and HIV-1)	(Hayashi et al., 1996)
OLIGOSACCHARIDES			
CHITOOLIGO-SACCHARIDE	Crustacean shells	Antimicrobial Anti-inflammatory Hypo-cholesterolemic	(Xia et al., 2010)
PROTEINS			
ALLO-PHYCOCYANIN	*Anacystis nidulans* (cyanobacterium)	Antioxidant	(Ge et al., 2006)
PHYCO-ERYTHROBILIN	*Porphyra* spp. (red alga)	Antioxidant	(Yabuta et al., 2010)
PEPTIDES			

TABLE 1 CONTINUED

ONNAMIDE	*Theonella swinhoei* (sponge) prokaryotic symbiont	Anti-tumorigenic	(Piel et al., 2004)
PARASIN I	*Parasilurus asotus* (catfish)	antimicrobial	(Park et al., 1998)
PELTEOBAGRIN	*Pelteobagrus fulvidraco* (catfish)	antimicrobial	(Su, 2011)
PROTEIN HYDROLYSATE	*Illex argentines* (cuttlefish)	Anti-hypertensive Antioxidant	(Amado et al., 2013)
POLYUNSATURATED FATTY ACIDS			
OMEGA-3-FATTY ACID (EPA)	salmon, pilchards and sardines	Anti-obesity Anti-inflammatory Risk reduction of CVD	(Kris-Etherton et al., 2002; Li et al., 2008; Wall et al., 2010)
OMEGA-3-FATTY ACID (DHA)	salmon, pilchards and sardines	Anti-obesity Anti-inflammatory Risk reduction of CVD	(Li et al., 2008; Wall et al., 2010)(Kris-Etherton et al., 2002)
STEROLS			
FUCOSTEROL	*Pelvetia siliquosa* (brown alga)	Anti-diabetic Antioxidant	(Lee et al., 2003, 2004)
ERYLOSIDE A	*Erylus lendenfeldi (sponge)*	Anti-tumorigenic Antifungal	(Carmely et al., 1989)
ERYLOSIDE F	*Erylus formosus* (sponge)	Thrombin receptor antagonist Antiplatelet	(Stead et al., 2000)

Other marine bioactive compounds include phenolic compounds (terpenes), photosynthetic pigments and vitamins (Hamed et al., 2015). Algae also possess small quantities of amino acids derived bioactive molecules such as taurine and kainoids. Taurine, a sulfur-containing amino acid derived from cysteine and methionine, is conditionally essential for biological processes such as bile-acid conjugation, neurological development, immune function, and retinal development. Bioactivities attributed to taurine include antioxidant, improved detoxification, membrane stabilising and anticonvulsant effects (Gaull, 1989). Further, infant milk formulas have been fortified with taurine since 1980 in order to replicate the concentrations found in lactating mothers (Harnedy and Fitzgerald, 2011). Kainoids are structurally related to the amino acids aspartate and glutamate, correspondingly fulfilling similar functions. These kainoids are of interest, because of their neuroexcitatory, insecticidal and anthelmintic bioactivities (Harnedy and Fitzgerald, 2011)

1.2. Sulfoconjugated polysaccharides from macro red algae: Their composition, biosynthesis, and most significant sources.

Marine eukaryotes account for 97% of the 18 000 marine natural products currently known (Arrieta et al., 2010). A subset of approximately 9% is marine natural products derived from red algae, while products derived from tunicates, cnidarians and sponges make up the remaining percentages. Algal polysaccharides i.e. agar, alginate and carrageenans represent high-value commodities used in the food, pharmaceutical and biotechnological industries. With the exception of alginate, sulfoconjugated galactans (agarans and carrageenans) are mainly extracted from macro red algae such as the carrageenan-rich genera *Chondrus*, *Gigartina*, *Kappaphycus*, *Eucheuma* and *Hypnea* and the agar rich genera such as *Gelidium*, *Gelidiella* and *Gracilaria* (Rhein-Knudsen et al., 2015). Sulfoconjugated galactans are SPs that form part of the cell wall and extracellular matrix of marine algae and numerous marine invertebrates (Pomin and Mourão, 2008). It is postulated that SP biosynthesis in marine organisms is a physiological adaptation to the marine environment (Ho, 2015), as a positive

correlation between plant SP content and salinity has been observed (Aquino et al., 2011).

Red and green algae accumulate sulfoconjugated galactans and similarly also do marine ascidians and sea urichins. In contrast, brown algae accumulate sulfoconjugated fucoidans and alginates (Figure 2). Carrageenans and agarans are the main sulfoconjugated galactans from marine red macro algae. Both carrageenans and agarans are anionic polysaccharides comprised of sulfoconjugated linear D-galactopyranose units linked alternatively by α-1,3 and β-1,4 glycosidic bonds. The α-1,3 linked galactopyranose unit of agarans is however in the L-conformation, in contrast with macro algal carrageenans where both α-1,3 and β-1,4 galactopyranose units are in the D-conformation (Rhein-Knudsen et al., 2015). There are three types of commercially used carrageenans i.e. κ, ι and λ. In general, ι- and λ-carrageenans have two to three sulfate residues per dimer, in contrast to the one sulfate for κ-carrageenans. Further, κ-carrageenans have a sulfate content of 25%-30%, ι-carrageenans 28%-30%, and λ-carrageenans 32%-39% - however, large fluctuations are possible. Typically, agarans contain remnants of their precursor, porphyrin (D-galactose and L-galactopyranose 6-sulfate) adding to the structural diversity of these SPs. Both carrageenans and agarans contain 3,6 anhydro-α-galactopyranose residues (Rhein-Knudsen et al., 2015). A common chemical substitution found on the agaran backbone is methylation, resulting in methylated galactose units such as 6-O-methyl-D-galactose and 4-O-methyl-L-galactose. Both carrageenans and agarans are highly heterogeneous polysaccharides, with structural properties being influenced by both environmental and intrinsic (inherent) factors such as seasonal changes, salinity, extraction methods, algal species and life cycle (Daugherty and Bird, 1988; Rhein-Knudsen et al., 2015).

The primary sulfoconjugated galactan found in green algae is composed of 3-β-D-galactopyranose 4-sulfate units. Further, fucoidans found in brown algae are characterised by α-L-fucose monomers bound alternatively by α-1,3 and β-1,4 glycosidic bonds with varying sulfoconjugation patterns present on the sugar

backbone. Similarly, SPs can also be found in other marine organisms such as marine angiosperms, sea urchins, and ascidians. Finally, other SPs such as the glycosaminoglycan, keratan sulfate, can also be found in trace amounts in fishes (Figure 2). The algal industry is valued at US$ 6 billion, with seaweeds used directly or indirectly in human consumption accounting for approximately US$ 5 billion and the algal polysaccharides industry accounting for the remaining US$ 1 billion (Kraan, 2012). The South African algal industry primarily focused on value added algal polysaccharides for their application in the food, pharmaceutical and biotechnological applications.

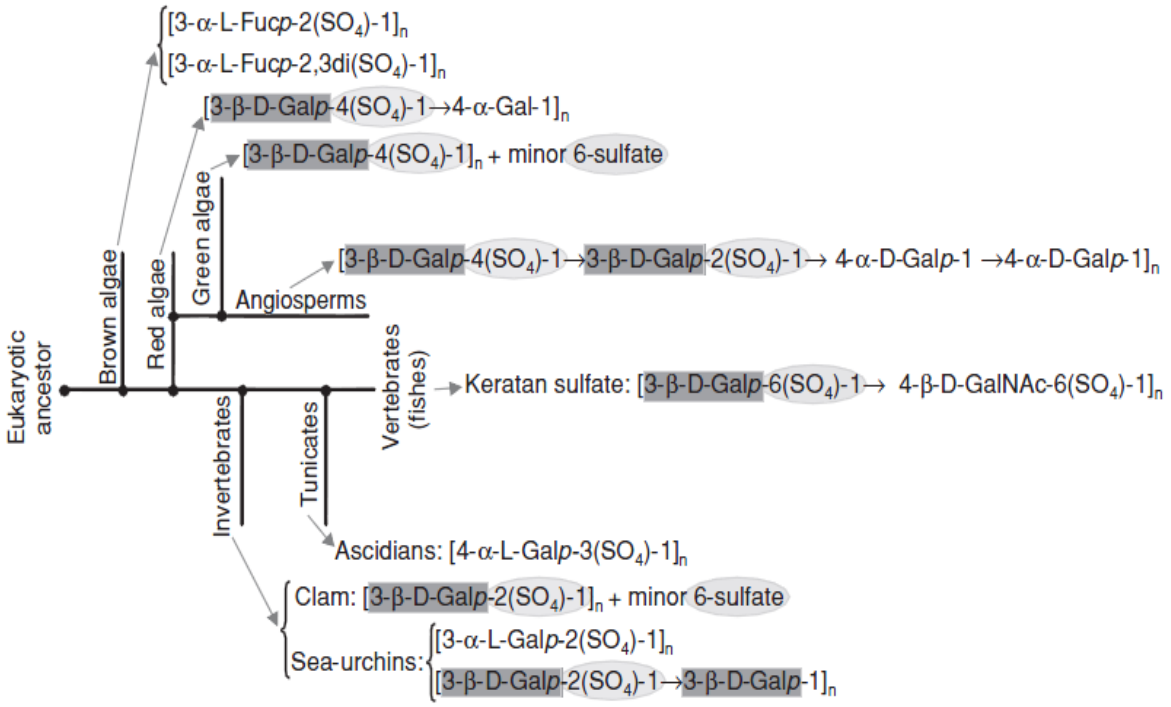


Figure 2. Sulfoconjugated polysaccharides found in different marine phyla, their principal monomer constituents, and proposed phylogenetic relationships. Highlighted in dark grey is the structure 3-β-D-galactopyranose, which is highly conserved among marine organisms that accumulate SPs. Further, the different types of sulfoconjugation i.e. 2-, 4- and 6-substitution, respectively is highlighted by light grey ellipses (Pomin and Mourão, 2008).

One of the principal players in South African is Taurus Chemicals (Cape Kelp) (Pty) Ltd who is involved in the cultivation and harvesting of brown and red algae for the global algal polysaccharide and seaweed trade. SPs are commonly extracted from natural populations of *Gracilaria verrucosa*, *Ecklonia maxima*, *G. pristoides,* and

Gelidium amanzii. (Dzein Studio, 2010). Other areas where seaweeds are used locally include raw material for medical research and abalone feed. One algal species used locally both for abalone feed and as a source of bioactive terpenes is *Plocamium corallorhiza*.

Despite the scale of the global algal polysaccharide industry, in particular, the sulfoconjugated galactan industry, the underlying biosynthetic pathways remain a mystery. However, this is a highly active area of research with several proposed biosynthetic pathways to date (Figure 3). Rhodophyta accumulates Floridean starch during irradiance through photosynthesis. Dark periods are characterised by the degradation of the accumulated starch into linear glucans and glucose. Manley & Burns (1991) and Barbier et al. (2005a) proposed that the precursor sugars for red algal SPs were derived from these breakdown products. To this end, the enzymes isoamylase, glucan water dikinase (R1) and amylopullulanase are responsible for linear α-1,4-glucan formation. Subsequently, glucose and glucose-1-phosphate are derived from these linear glucans through the actions of glucan 1,4-glucosidase and glycogen phosphorylase, α-amylase and β-amylase, respectively. These sugars, through the enzymatic actions of various other enzymes, become the precursor sugars GDP-mannose and UDP-D-glucose, used for algal SP biosynthesis. The nucleotide sugars are transported into the Golgi apparatus, after epimerization by GDP-mannose 3,5-epimerase and UDP-galactose 4-epimerase, as GDP-L-galactose and UDP-D-galactose. In the Golgi apparatus, GDP-L-galactose and UDP-D-galactose are used by UDP and GDP galactosyltransferases as the building blocks for different sulfoconjugated algal polysaccharides, most notably agarans and carrageenans. There exist discrepancies as to when chemical group substitutions such as methyl, pyruvate and sulfate conjugation occur i.e. before polysaccharide backbone polymerization or afterwards. A consensus, however, does exist that algal polysaccharides carry various chemical group substitutions, undergoing continuous alterations depending on inherent or environmental changes (Daugherty and Bird, 1988; Marinho-Soriano, 2001; Pomin and Mourão, 2008). One of the most prominent

chemical modifications is sulfoconjugation thought to be executed by sulfotransferase (SULTs) and removed by sulfatases (Collén et al., 2013).

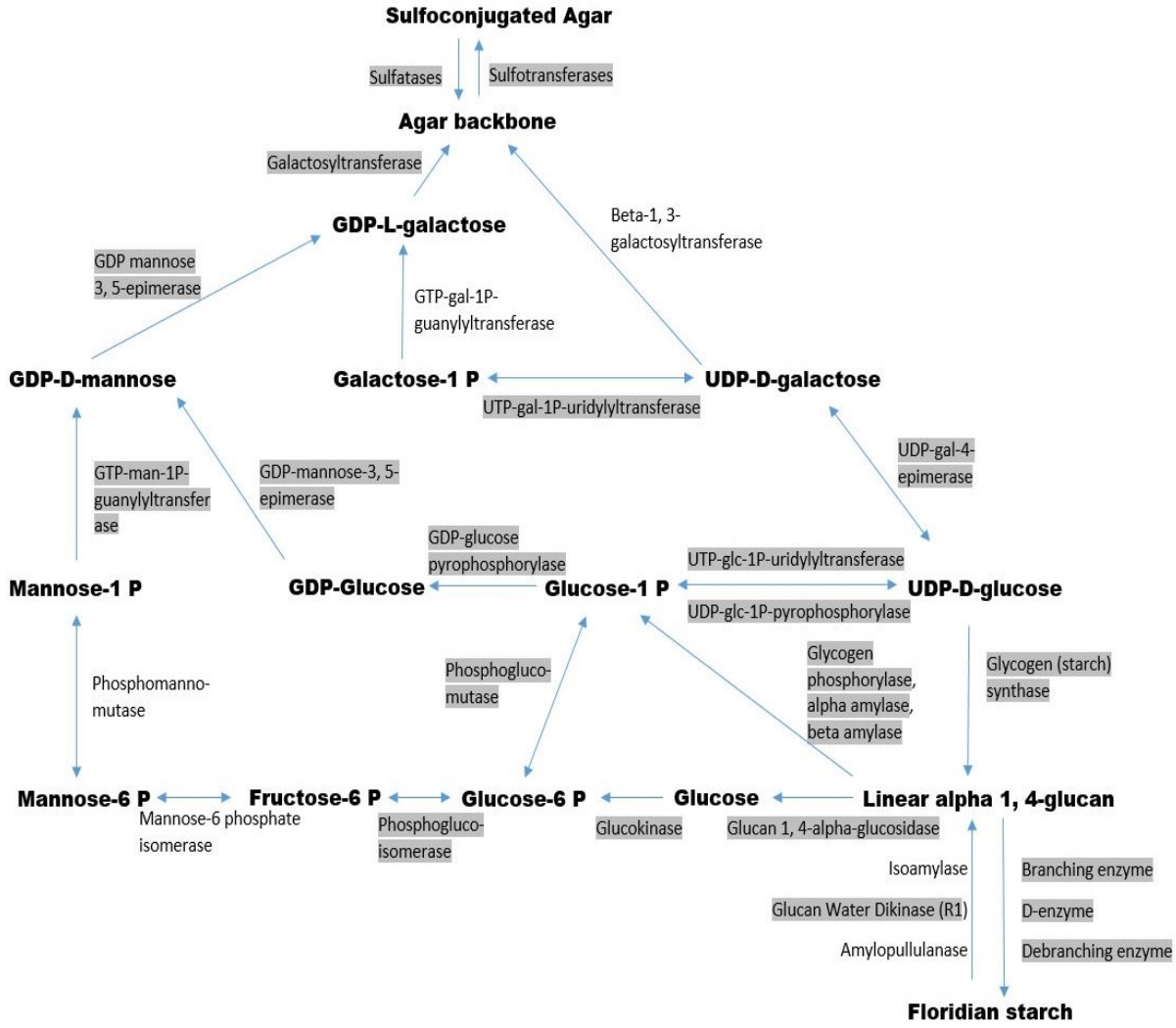


Figure 3. The proposed biochemical pathway for agar biosynthesis in Rhodophyceae. Floridean starch that was accumulated through photosynthesis provides the precursor sugars for algal SP biosynthesis. Floridean starch is degraded into linear alpha 1,4-glucans through the actions of isoamylase, glucan water dikinase and amylopullulanase - this breakdown is reversible. The formed linear glucans are subsequently converted into glucose and glucose-1-phosphate. Through various enzymatic conversions and modifications such as isomerization, epimerization, and nucleotide sugar formation, GDP-L-galactose and UDP-D-galactose, principal algal SP constituents are formed. GDP and UDP-galactosyltransferase polymerise GDP-L-galactose and UDP-D-galactose monomers to form the sugar backbone of agar. Either prior or post agar backbone synthesis, chemical group modifications such as sulfoconjugation is added onto the agar backbone through the actions of sulfotransferases. Present sulfate groups are removed from the agar backbone through sulfatase, forming very characteristic sulfoconjugation patterns. Enzymes identified in this study are shaded grey. Figure adapted from Barbier et al. (2005) and Hector (2013).

1.3. Sulfotransferases

The chemistry of algal SPs is well studied, in contrast, we currently have a limited understanding of the underlying genetic mechanisms responsible for the biosynthesis of these algal SPs (Pomin & Mourão 2008). Current biosynthetic models for algal SP biosynthesis are based on a combination of red algal and terrestrial plant biochemistry (Barbier et al., 2005; Chang et al., 2014; Hector, 2013; Manley and Burns, 1991). Additionally, as stated in section 1.2, macro algal SPs are highly heterogenic molecules, with the heterogeneity of Rhodophyta SPs being largely dependent on their sulfate content and sulfoconjugation pattern (Pomin and Mourão, 2008). To date, the isolation and biochemical characterization of algal SULTs remain elusive. However, a micro algal expressed sequence tag (EST) library, displaying SULT activity, has been described, along with an accompanying patent for the synthesis of SPs (Arad et al., 2010; Arad and Levy-Ontman, 2010). Based on structural studies conducted on animal SPs e.g. chondroitin and dermatan sulfate, in combination with biochemical investigations on SULTs from terrestrial organisms e.g. humans, mice and plants, substrate sulfoconjugation can occur at four different carbon locations i.e. C-2, C-3, C-4 and C-6, (Hector, 2013; Silbert and Sugumaran, 2002).

Above, the paucity of research done on algal SULTs is contrasted by research on SULTs from terrestrial organisms, especially animals (Hernàndez-Sebastiá et al., 2008; Sakakibara et al., 2002). SULTs are the enzyme superfamily responsible for enzymatic sulfation and sulfonation, which are collectively termed sulfoconjugation. Figure 4 illustrates the sulfoconjugation reaction catalysed by SULTs, this is the enzymatic transfer of a sulfuryl group (SO_3) from a donor molecule, classically 3'-phosphoadenosine 5'-phosphosulfate (PAPS), to an acceptor molecule i.e. compounds containing either hydroxyl or amine groups (Chapman et al., 2004; Hernàndez-Sebastiá et al., 2008; Paul et al., 2012).

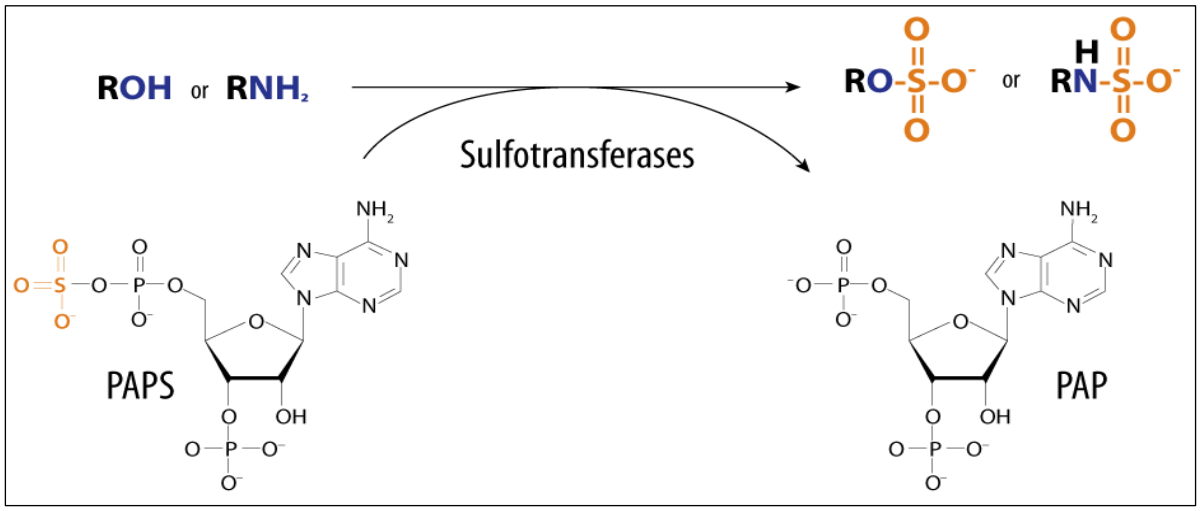


Figure 4. The general sulfoconjugation reaction as enzymatically catalysed by sulfotransferases. The sulfuryl group (orange) is transferred from the universal sulfate donor, classically 3'-phosphoadenosine 5'-phosphosulfate (PAPS), to an appropriate acceptor molecule, such as compounds containing either a hydroxyl or an amine group (blue). The by-product of the sulfoconjugation reaction is 3'-phosphoadenosine-5'-phosphate (PAP). Diagram obtained from Prather et al. (2012).

Suitable SULT substrates are diverse, such as iodothyronines, steroid hormones, catecholamine, cholesterols, phenolic compounds, salicylic acid, tyrosylproteins, carcinogens, chondroitin, heparan sulfate, neurosteroids, flavonone, brassinosteroids and benzyl, methionine and phenylalanine derived desulfo-glucosinolate (dsGl) (Hernàndez-Sebastiá et al., 2008; Hirschmann et al., 2014). In this regard, substrate specificity has traditionally been used to facilitate SULT classification, however, a classification system for SULTs according to nucleotide sequence i.e. gene or cDNA similarity, has been established recently (Glatt, 2000). Based on Hidden Markov Models (HMM), putative SULTs characteristically possess at least one of seven Pfam motifs, assigning them into specific SULT families. Subsequently, the SULT superfamily can be divided into distinct families, depending on the protein catalytic domain present, i.e. sulfotransfer_1 (PF00685), sulfotransfer_2 (PF03567), sulfotransfer_3 (PF13469), sulfotransf (PF09037), gal-3-O_sulfotr (PF06990), Arylsulfotrans_1 (PF05935) and Arylsulfotrans_2 (PF14269) families, respectively (Hirschmann et al., 2014).

SULT commonly also have four highly conserved regions i.e. regions I to IV that were identified through sequence alignments from eleven cytosolic SULTs from plants, animals and bacteria (Marsolais and Varin, 1995). Localised at the N-terminus, region I interact with the 5'-phosphate of PAPS and forms part of the PAPS binding domain. Still forming part of the PAPS binding domain, which is essential for PAPS co-factor binding, is region II which starts with a very characteristic histidine residue (responsible for proton reception during sulfuryl transfer). The C-terminal section of region II forms a 3' P-motif, responsible for binding the 3'-phosphate of PAP. Further, regions III and IV contain a highly conserved poly-glutamic acid (Poly-Gluc) motif of which the function is currently still unknown. Lastly, localised at the C-terminal, region IV is a P-loop GxxGxxK motif. Regions I and IV are highly conserved and can be found in at least 20 *A. thaliana* SULTs (Hirschmann et al. 2014; Hernàndez-Sebastiá et al. 2008, Figure 5).

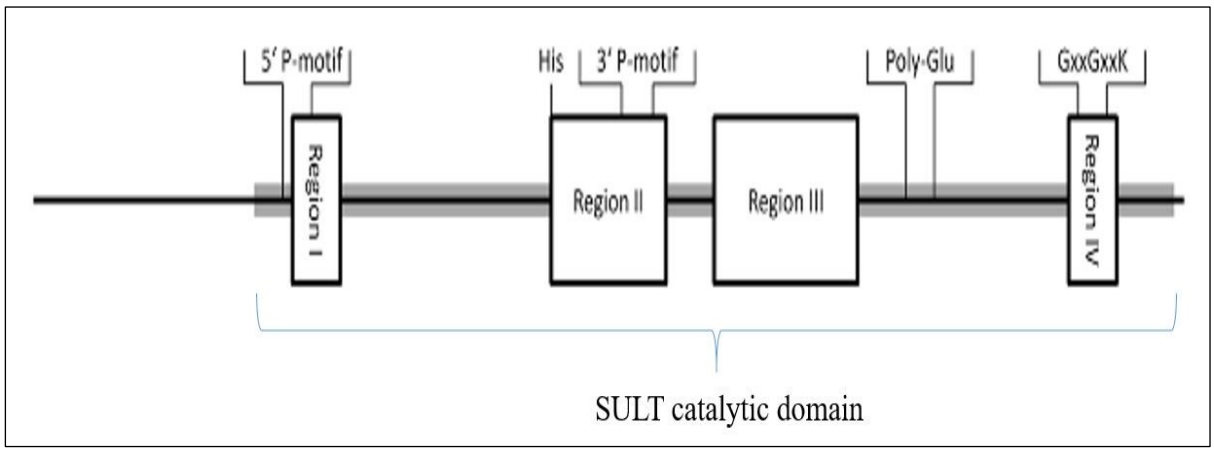


Figure 5. Conserved protein motifs commonly found in sulfotransferase enzymes. Based on studies of *A. thaliana* SULTs and structural research conducted on *Mus musculus* SULTs, conserved protein motifs are illustrated. PAPS binding (regions I, II, and IV), a histidine residue responsible for proton acceptance and region III, of unknown function. The characteristically conserved motifs are indicated by boxes, with the position and size proportionally represented to average *A. thaliana* SULTs. Further, the SULT catalytic domain is indicated in grey along the gene length. Figure adapted from Hirschmann et al. (2014).

SULTs are further categorised into two main classes, that is, either cytosolic or membrane-associated (Hernàndez-Sebastiá et al., 2008). Our early understanding of these enzymes come from studies conducted on cytosolic SULTs, which play a role in

the metabolism of small endogenous and exogenous molecules like xenobiotic compounds, bioamines and hormones. Traditionally sulfoconjugation was thought to be principally responsible for drug detoxification, but it has also been implicated in the modulation of hormones, neurotransmitters and the immune system. (Chapman et al., 2004). The biotransformation of endogenous and exogenous compounds are classified into phase one and phase two reactions. Conjugation reactions which include sulfoconjugation and glucuronidation, form part of phase 2 compound biotransformation (Jančová and Šiller, 2012). The importance of cytosolic SULT-mediated detoxification can be inferred from the absence of UDP–glucuronosyltransferases (UGTs) transcripts in fetal liver, until week 20 of pregnancy. The UGT superfamily is vital for the detoxification of a diverse spectrum of natural and synthetic compounds in vertebrates. The absence of UGT transcripts in fetal liver until week 20 of pregnancy, may suggest a more substantial role for SULT-mediated detoxification as UGTs and SULTs have overlapping substrate specificities. (Jančová and Šiller, 2012). The conjugation of endogenous and xenobiotic molecules increases the water solubility of these compounds, while also decreasing their half-life and passive diffusion over cell membranes. These physicochemical changes are advantages for enhanced excretion and reduced reabsorption via the biliary and urinary system (Glatt, 2000). In contrast to cytosolic SULTs, membrane-associated SULTs are confined to the either the Golgi apparatus or plasma membrane and responsible for the sulfoconjugation of larger compounds including polysaccharides, peptides, proteins, glycosaminoglycans and lipids (Jančová and Šiller, 2012; Paul et al., 2012; Varin et al., 1997). Figure 6 illustrates the general sulfoconjugation mechanism of both classes of SULTs, in addition to their subcellular localisation.

The majority of biochemical and structural research have been done on animal SULTs. This is exemplified in that the first structural studies conducted on SULTs, were on the mouse oestrogen SULT. Additionally, for all 13 human SULTs identified, there exist biochemical data such as enzyme kinetics and substrate specificities. Again, the *A.*

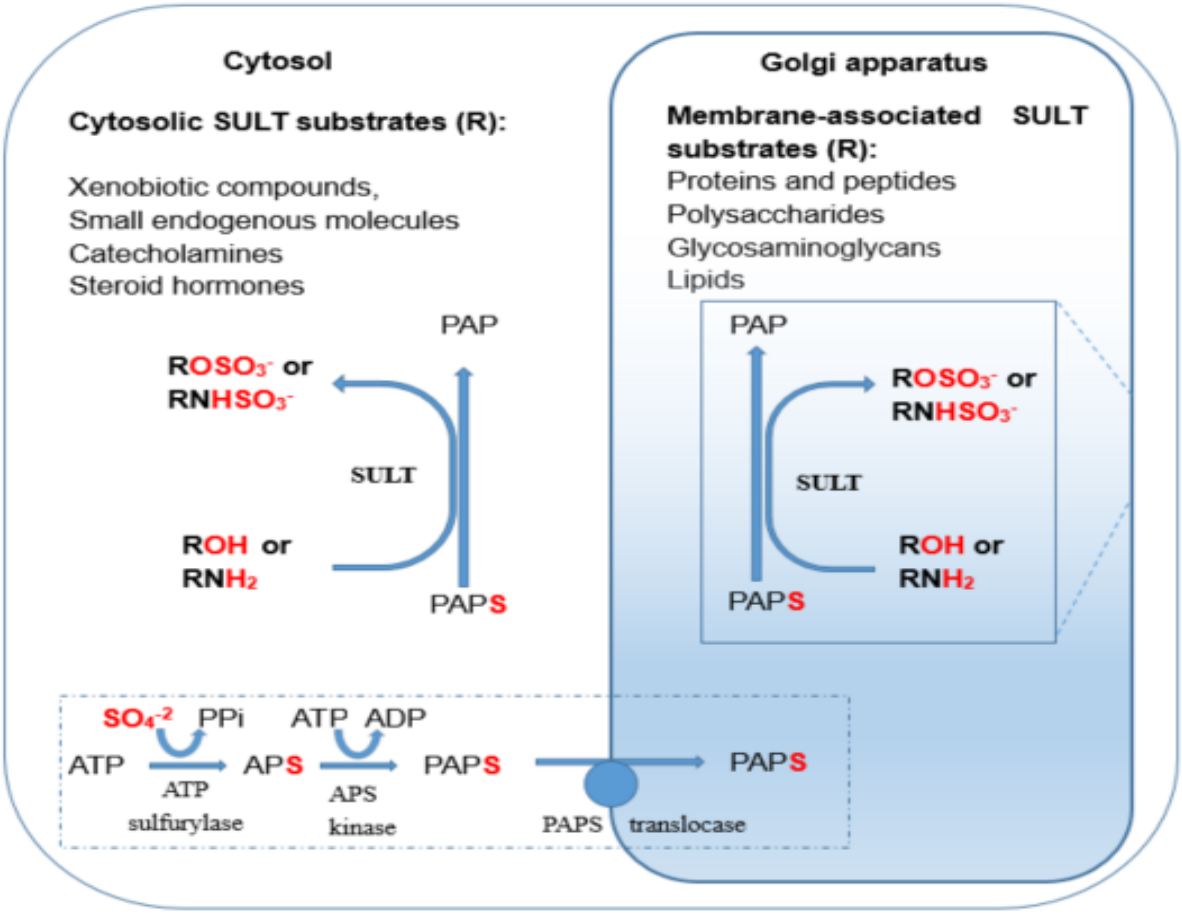


Figure 6. The enzymatic transfer of a sulfuryl group from the donor molecule, PAPS, to an acceptor molecule, ROH or RNH2, by a either cytosolic or membrane-associated SULT. The SULT substrate is represented by an R and depending on cellular localisation can denote a carbohydrate, peptide, or xenobiotic compound containing a hydroxyl or amine group, respectively. The sulfate donor i.e. PAPS is biosynthesized in the cytosol sequentially by ATP sulfurylase and APS kinase. Subsequently, PAPS is transported into the Golgi lumen by the transport protein, PAPS translocase. The figure is adapted from http://what-when-how.com/molecular-biology/sulfation-molecular-biology.

thaliana SULT superfamily signifies the best-characterised plant SULT superfamily to date and is represented by 21 members (Hirschmann et al., 2014). Extensive research has been done on desulfo-glucosinolate (dsGI) SULTs in *A. thaliana*, such that dsGI-SULTs with differing substrate affinities for indolic and aliphatic dsGI have been found. Similarly, differences in the substrate affinities of dsGI-SULTs have also been observed across different Arabidopsis ecotypes (Hirschmann et al., 2014). All 21 *A. thaliana* SULT genes were identified through sequence homology approaches via

BLAST (Altschul et al., 1990) and subsequently, ten of the 21 SULTs have been biochemically characterised (Table 2). SULTs have a broad range of substrates and therefore many functions. In this regard, from studies on terrestrial organisms the diversity in biological functions of sulfoconjugation can be understood by the copious numbers of different sulfate conjugates present in the organisms, SULT gene genetic polymorphisms, their broad substrate diversity, and enantioselectivity and tissue distribution (Hernàndez-Sebastiá et al., 2008; Hirschmann et al., 2014).

Currently, conserved protein motifs and protein architecture analysis provide clues as to the function of putative SULTs. Phylogenetic studies, however, have been of limited value in predicting biochemical properties of unknown SULTs (Hirschmann et al., 2014). Partially, this is due to observations that minor changes in SULT gene sequences lead to diverse differences in substrate affinities and enzyme kinetics. Conversely, experimental enzyme substrate specificities and phylogenetic analysis do not correspond i.e. SULTs with more than 85% sequence homology do not display similar enzymatic characteristics. For example, the in silico assignment of *A. thaliana* SULT13 as a brassinosteroid SULT, did not correspond with experimental data that it uses flavonoids as preferred substrates. Illustrated in Table 2, despite a large number of putative algal SULTs identified in databases, no algal SULT has been isolated and characterised to date. Further, with the limited ability to reconcile SULT phylogenetic analyses with SULT biochemical predictions, algal SULT kinetics, and substrates specificities remain a mystery.

Table 2. Sulfotransferase genes currently identified and biochemically characterised. Organisms included are terrestrial animals and plants, algae and prokaryotes.

ORGANISM	IDENTIFIED SULTS	BIOCHEMICALLY CHARACTERISED SULTS	SUBSTRATE SPECIFICITIES	LITERATURE
HOMO SAPIES	13	13	Iodothyronines, oestrogen, catecholamine, steroids, cholesterol, oxysterols, phenolic compounds, drugs, carcinogens,	(Gamage et al., 2006)
DANIO RERIO	13	10	Chondroitin, heparan sulfate, partially desulfated dermatan, dehydroepiandrosterone, neurosteroids, phenolic compounds	(Cadwallader and Yost, 2006; Mizumoto et al., 2009; Sugahara et al., 2003; UniProt Consortium, 2015)
ARABIDOPSIS THALIANA	21	10	Flavonol, flavanone, brassinosteroids, salicylic acid, tyrosylproteins, benzyl, methionine and phenylalanine derived desulfo-glucosinolate (dsGI)	(Hirschmann et al., 2014)
ORYZA SATIVA	35	0	unknown	(Cao et al., 2016; Chen et al., 2012)
CHONDRUS CRISPUS	12	0	unknown	(Collén et al., 2013; Ho, 2015)
PORPHYRIDIUM CRUETUM	13	0	unknown	(Ho, 2015)
ECTOCARPUS SILICULOSUS	14	0	unknown	(Ho, 2015)

TABLE 2 CONTINUED

SINORHIZOBIUM MELILOTI	1	1	N-acetylglucosamine oligosaccharides	(Ehrhardt et al., 1995; UniProt Consortium, 2015)
MYCOBACTERIUM TUBERCULOSIS	4	2	Trehalose, menaquinone	(Holsclaw et al., 2008; Pi et al., 2005)

1.4. Project aims and objectives.

Sulfoconjugated polysaccharides (SPs) from macro red algae are known to display various bioactivities such as antioxidant, antibacterial, anticoagulant, anti-tumorigenic and anti-viral properties (Wijesekara et al., 2011). Further, structural studies have determined the bioactivities observed for algal SPs are largely depended on the sulfate content and sulfoconjugation pattern present on the sugar backbone (Damonte et al., 2004; Witvrouw and De Clercq, 1997). However, to date, no macro red algal SULTs, i.e. the enzymes responsible for attaching sulfate to the sugar backbone, have been isolated and biochemically characterised. This absence of scientific enquiry into macro red algal SULTs, provides the opportunity to isolate these enzymes for downstream synthesis of novel bioactive compounds, by providing different sugar substrates to these SULTs, especially those that are not naturally found in algae (Collén et al., 2013; Ho, 2015).

All biochemically characterised SULTs are sulfoconjugating. Further, SULTs from terrestrial animals and plants are well characterised and display diverse substrate specificities. We hypothesise that algal SULTs are no exception with regard to their substrate promiscuity and may possibly sulfoconjugate a large repertoire of compounds, similar to the broad substrate specificities of currently characterised SULTs (Collén et al., 2013; Hernàndez-Sebastiá et al., 2008). This leaves a copious number of possible compounds that could be appropriate algal SULT substrates, some potentially displaying novel and more potent bioactivities than currently available algal SPs (Wijesekara et al., 2011).

In this regards, the primary aim of this study is to identify and isolate macro red algal carbohydrate SULTs. The secondary project aim is to assess the diurnal differential changes in gene expression, specifically regarding sulfoconjugated galactan biosynthesis in a macro red algal species. In order to achieve the primary and secondary research aims the following objectives were completed: 1. Diurnal macro red algae sampling and subsequent identification. 2. RNA-sequencing and transcriptome assembly. 3. Transcriptome annotation. 4. Transcriptomic data analysis, SULT gene identification, followed by SULT ORF isolation and finally; 5. Diurnal differential gene expression analysis regarding nocturnally up- or

downregulated transcripts, related to galactan biosynthesis. Ultimately this project forms part of a bigger ambition to synthesise improved SPs with application in the nutraceuticals industry, through heterologously expressed SULTs - that are then incubated (in vitro) with various carbohydrates (both simple and complex) to generate novel bioactive SPs.

Chapter 2: Materials and methods

2.1. Diurnal sampling and identification of macro red algae

A diurnal sampling of macro red algae was carried out at Strand, Western Cape, South Africa (-34.123963, 18.831267). Strand was chosen as an appropriate sampling area because of accessibility to shallow intertidal zones and a variety of different algae specimens. Sampling hours were between 12:00pm to 02:00pm & 12:00am to 02:00am, corresponding to low tide (2nd and 3rd May 2015). Based on the phenotypic red pigmentation typical of macro red algae, samples were collected, placed in 50 mL conical tubes, flash frozen in liquid nitrogen and stored at -80° C. Dr Tammy Robinson (Department of Botany and Zoology, Marine biology unit, Stellenbosch University) performed the phenotypical identification of the sampled algae. We proceeded with the macro red alga *Plocamium corallorhiza* due to its ubiquitous nature and literature espousing the presence of SPs in algae from the same genera (Branch et al., 2013; Falshaw et al., 1999; Wynne, 2002).

2.2. Determining the presence of acidic and/or sulfoconjugated polysaccharides in *P. corallorhiza* extracts.

In an attempt to confirm the presence of acidic and/or sulfoconjugated polysaccharides in *P. corallorhiza* , toluidine blue O stains were performed on (1) Heparan sulfate and (2) sulfoconjugated arabinogalactan, as positive controls, (3) leaf extracts of *A. thaliana,* as a negative control, (4) *P. corallorhiza* agaran extracts (desulfoconjugated and untreated) and (5) whole alga extracts. Alga samples were rinsed to remove any debris. All samples were oven dried (90°C for 30min) prior to extractions. The samples were separated into two groups (each group consisted of 3 technical replicates), one subjected to desulfoconjugation by means of alkali pre-treatment and the other remained untreated. The agaran desulfoconjugation, i.e. removing agaran sulfate group, was done according to Freile-Pelegrin and Murano (2005) using 5% NaOH alkali treatment and subsequently washed under the tap for 30 minutes at room temperature.

Next, native agaran were extracted as previously described by Marinho-Soriano et al., (1999), with the following modifications: briefly, each cleaned and dried sample was added to 500 mL of MilliQ water (pH adjusted to 6.5 with glacial acetic acid) in an Erlenmeyer flask and extractions were performed in an autoclave (1h, 110^{o}C). Samples were subsequently centrifuged (8000g, 10min). Gravity assisted separation was used as an alternative to cotton tissue and pressure glass fibre filtration, as a means of removing large particles. Lastly, the extracts were freeze-dried overnight (VirTis® BenchTop™ 2K Freeze Dryer).

Comparing the presence of acidic and/or sulfoconjugated polysaccharides in *P. corallorhiza* and *A. thaliana*, rudimentary tissue extractions were performed. Algal thallus tissue and *A. thaliana* leaf tissue were flash frozen in liquid nitrogen and ground in a chilled mortar and pestle (4^{0}C). Subsequently, 450µL of cell lysis buffer RLC (QIAGEN) was added to 100mg of ground tissue and the lysate transferred to two shredder spin columns (QIAGEN) and centrifuged (13 000g, 2 min). The flow-through of the supernatants was transferred to new 1.5mL Eppendorf tubes, for each sample. The thallus and leaf tissue extracts were analysed by agarose gel electrophoresis (0.5x Tris/Borate/EDTA [TBE] buffer), followed by toluidine blue O staining, as previously described (Volpi and Maccari, 2002).

2.3. RNA extractions, integrity analysis and high throughput Next-Generation Sequencing (NGS).

2.3.1 *P. corallorhiza* RNA extractions and quality assessment

Thallus RNA extractions for the midday and midnight samples, were prepared using the RNeasy (QIAGEN) plant mini kit following the manufacturer's instructions. Lysis buffer RLC was used, in addition to the optional on-column DNase digestion procedure. RNA integrity was assessed, through formaldehyde agarose gel electrophoresis according to QIAGEN (2012) as modified from (Sambrook and Russell, 2001). The RNA integrity was determined using an Agilent bioanalyser 2100 (Agilent) at Stellenbosch University's Central Analytical facility (CAF).

2.3.2. RNA-sequencing of a midday and midnight *P. corallorhiza* sample.

Next-generation sequencing (NGS) library preparations and sequencing were performed at the Agricultural Research Council's Biotechnology Platform (ARC-BTP, Pretoria, South Africa). cDNA libraries were prepared using the Truseq® stranded mRNA sample preparation kit according to the manufacturer's instructions. Subsequently, the samples were multiplexed and sequenced using the Illumina Hiseq 2500 with paired-end sequencing (Illumina Inc., 2013). The respective midday and midnight sample was sequenced 3 times over and generated 22.56 Gigabase (Gb) of raw sequencing data in total.

2.4. *P. corallorhiza* RNA-seq data pre-processing, transcriptome assembly and diurnal differential gene expression analysis.

2.4.1. RNA-seq pre-processing and transcriptome assembly

All bioinformatic analyses were conducted using the CLC bio genomics workbench (version 9.0) (QIAGEN Aarhus A/S, 2015) supplemented with either the stand alone or plugin client of BLAST2GO (version 3.2.7) (Biobam, 2015). The RNA-seq quality assessment and data pre-processing were performed with the following parameters: Phred scale quality score limit (0.05), trimming of ambiguous nucleotides with maximum number of ambiguities (2), Illumina adapter trimming (AR006 and AR007) (Illumina, 2013) and stringently seven nucleotides were trimmed from each 5' and 3' read terminus.

De novo assembly using the combined processed diurnal samples was executed with different k-mer values (Chikhi and Medvedev, 2014) and default workbench parameters. Briefly, parameters were, k-mer size (automatic size of 24 and manual sizes of 30, 35, 40), map reads back to contigs (slow) with mismatch cost (2), insertion coast (3), deletion cost (3), length fraction (0.5) and similarity fraction (0.8). Contigs were updated during read back mapping and contig scaffolding was performed.

2.4.2. Diurnal differential gene expression analysis and visualisation

The assembled contigs obtained from the automatic word size of 24 were used as references for the expression analysis. Contig length filtering (≥ 900 bp), resulted in 11 816 contigs (9.873%) remaining from the 119 675 contigs produced originally,

these were subsequently used for diurnal differential gene expression analyses of the processed day and night temporal samples, respectively. In order to identify transcripts differentially expressed (DE) at night when compared to the day, the Wald statistical analysis test was used (log_2 Fold change ≥ 5 or ≤ -5, with the Bonferroni and False Discovery Rates (FDR) corrected p-values ≤ 0.01).

The day RNA-seq reads of 47 782 626 (average length 107) and the night RNA-seq reads of 44 060 431 (average length 108) were mapped to the 11 816 contigs for expression analyses, using the advanced RNA-seq analysis plugin in CLC bio genomics workbench (QIAGEN Aarhus A/S, 2015). Briefly, parameters were, total counts expression measurement, maximum number of hits for a read (1) and additional default workbench parameters.

The resulting gene expression tracts were used to create an expression analysis experiment between the night and day temporal samples. DE expressed transcripts were identified by applying filters for log_2 Fold change ≥ 5 (upregulated in the night) or ≤ -5 (downregulated in the night), Bonferroni corrected p-value ≤ 0.01 and FDR, corrected p-values ≤ 0.01. Representation of DE transcripts was visualised in volcano plots. The lists of nocturnally up and downregulated transcripts were exported into BLAST2GO (Biobam, 2015) for Fisher's exact test (FDR and P-value thresholds of 0.01, Gotz et al., 2008) and gene network analysis.

2.5. In silico transcriptome annotation using sequence homology, protein domain information and Gene Ontology (GO) term allocation.

The BLAST2GO (Biobam, 2015) pipeline was used for contig annotation (Conesa et al., 2005; Gotz et al., 2008). The contig subset ≥ 900 (11 816) was annotated through BLASTx analysis against the National Center for Biotechnology Information (NCBI) non-redundant protein database (nr), locally stored (downloaded 08 February 2016) in the CLC bio genomics workbench (QIAGEN Aarhus A/S, 2015). Briefly, BLASTx parameters were, word size of 6, e-value cut-off of 1.0E-5, number of BLAST hits equal to 20, filtering for low complexity and high-scoring segment pairs length cut-off of 33.

The BLASTx analysis xml output file was imported into BLAST2GO followed by InterProScan analysis using default and the optional signalPHMM prediction

parameters. The BLASTx results and protein domain information obtained from the InterProScan analysis were merged. Subsequently, the BLAST2GO annotation pipeline's remaining steps including mapping and annotation were executed. The annotation augmentation by annex feature was employed to retrieve additional annotations and the GO-slim feature to refine annotations for relevance to plant genomes.

2.6. Carbohydrate SULT transcript identification and cloning.

The annotated contigs obtained from the BLAST2GO pipeline were filtered for contig descriptions containing the term "sulfotransferase". The putative SULTs identified were assessed for open reading frames (ORFs) and these ORFs were subsequently converted into protein sequences using CLC bio genomics workbench (QIAGEN Aarhus A/S, 2015). The GenomeNet's protein motif search function through means of the Pfam database, was used to assess SULT superfamily allocation, using a E-value upper limit of 0.01 (Kanehisa et al., 2002). From the 18 SULT transcripts identified, a carbohydrate SULT encoded in contig_623, termed putative carbohydrate sulfotransferase 1 (PCHST1) assigned to sulfotransfer_2 superfamily and was chosen for isolation from *P. corallorhiza* cDNA. In addition, we also attempted to isolate a glycolipid SULT (denoted as GLST1), encoded in contig_4888 and belonging to the sulfotransfer_1 superfamily.

The PCHST1 forward primer (ATGTCGCCTCATTCTGTATCCC) and reverse primer (TCAATATTCCTCCGTTTCCTCAGTT) and the GLST1 forward primer (ATGACTCTCGACGCCAAA) and reverse primer (TTAAAAGTTTGGAAGCTTGTC) were designed in CLC bio genomics workbench (QIAGEN Aarhus A/S, 2015). Primer binding regions were from PCHST1 ORF start (residues 1 to 22) and terminus (residues 1122 to 1146) and GLST1 ORF start (residues 1 to 18) and terminus (residues 1158 to 1178). PCHST1 and GLST1 were amplified from *C. Plocamium* cDNA using the polymerase chain reaction (PCR) with high fidelity Q5 DNA polymerase (New England BioLabs®, Inc.) according to the manufacture's protocol. The resulting PCR amplicons were assessed through a 0.5 % agarose gel electrophoresis, for 1 hour at 100V in 0.5x TBE buffer stained with PRONASAFE (Laboratorios Conda, SA) (Sambrook and Russel, 1989). The resulting DNA bands

were excised under UV light (Alpha Innotech Corporation, USA) and purified. Purification of the excised amplicons was done using the Wizard® SV Gel and PCR Clean-Up kit (Promega, USA) according to the manufacturer's instructions. The purified PCHST1 and GLST1 amplicons were assessed for concentration and purity (absorbance at 260/280) on the Nanodrop lite (Thermo Fisher Scientific, Inc., USA) and subsequently ligated into pJet1.2 blunt (Thermo Fisher Scientific, Inc.) overnight using the manufacture's protocol.

Next, the PCHST1-pJet1.2 and GLST1-pJet1.2 ligation reactions were transformed into chemically competent (Mascher, 2014) DH5α™ *Escherichia coli* cells (Molecular Cloning Laboratories, USA). The transformed *E. coli* were cultured on Luria-Bertani media (LB), supplemented with 100 µg/L ampicillin. PCR colony screening with insert specific primers, using GoTaq® G2 DNA Polymerase (Promega, USA) according to the manufacture's protocol and subsequent 0.5% agarose gel electrophoresis (0.5x TBE buffer) revealed colonies with either the PCHST1::pJet1.2 or the PGLST1::pJet1.2 ligation event present. Positive colonies were cultured in LB broth, supplemented with 100 µg/L ampicillin and the plasmids isolated using the Wizard®Plus SV Minipreps DNA Purification System (Promega, USA) following the manufacture's protocol. The isolated plasmids were Sanger sequenced using pJet1.2 sequencing primers (Thermo Fisher Scientific, Inc.) at Stellenbosch University's Central Analytical Faculty (CAF). Colonies containing the PCHST1::pJet1.2 and GLST1::pJet1.2 plasmids were stored at -80^{o}C in 80% glycerol for future analysis.

Chapter 3: Results

3.1. *P. corallorhiza* as a generic representative of the Rhodophyta phylum, for SULT identification

Marine algae contain a wide range of carbohydrate-based bioactive compounds, which have extra-nutritional health benefits. In marine algae, the most sought after bioactive compounds are sulfoconjugated polysaccharides (SPs) which have the potential to be used as functional ingredients in the nutraceuticals industry. SPs are of particular biological interest to human health and have been shown to exert anti-inflammatory, anti-tumorigenic, immune-stimulatory, anti-ulcer, and more recently anti-viral activities. Using a high-throughput RNA-sequencing approach, we aimed to identify and isolate carbohydrate sulfoconjugating SULTs with the long-term objective of creating SPs with novel bioactivities.

This approach included the diurnal sampling of a South African macro red algal species, *Plocamium corallorhiza,* to identify SULT transcripts and assess the potential differential changes in gene expression, specifically regarding sulfoconjugated galactan biosynthesis. Thallus material from a number of macro algae was sampled from Strand, Western Cape, South Africa (-34.123963, 18.831267) between 12:00 pm to 02:00 pm & 12:00 am to 02:00 am on the second to the third of May 2015, respectively. Algal samples were phenotypically identified (Figure 7).

We examined the presence of SPs in *P. corallorhiza* thallus material by selectively staining polyanionic compounds (SPs) using agarose gel electrophoresis and toluidine blue O staining post-staining (Figure 8). The positive controls, heparan sulfate (lane 1) and sulfoconjugated arabinogalactan (lane 2), stained positive. Conversely, the negative control (non-sulfoconjugated galactan, lane 3), did not. Crude thallus extracts from *P. corallorhiza* (lane 4) stained positive for SPs, while crude extracts from *A. thaliana* leaves (lane 5) did not. Comparative analysis of *P. corallorhiza* native SPs (lane 6) and desulfoconjugated SPs (lane 7), showed that both stained positive with the desulfoconjugated extracts staining more intensely and migrating less on the agarose gel than the native SPs.

Figure 7. A diurnal sampling of macro algae from shallow intertidal zones on the south coast of South Africa. Top row (left to right), *Plocamium corallorhiza, hypnea* sp., *Codium* sp.; Bottom row, *Caulerpa filiformis*, *Sargassum* sp., unidentified algal specimen.

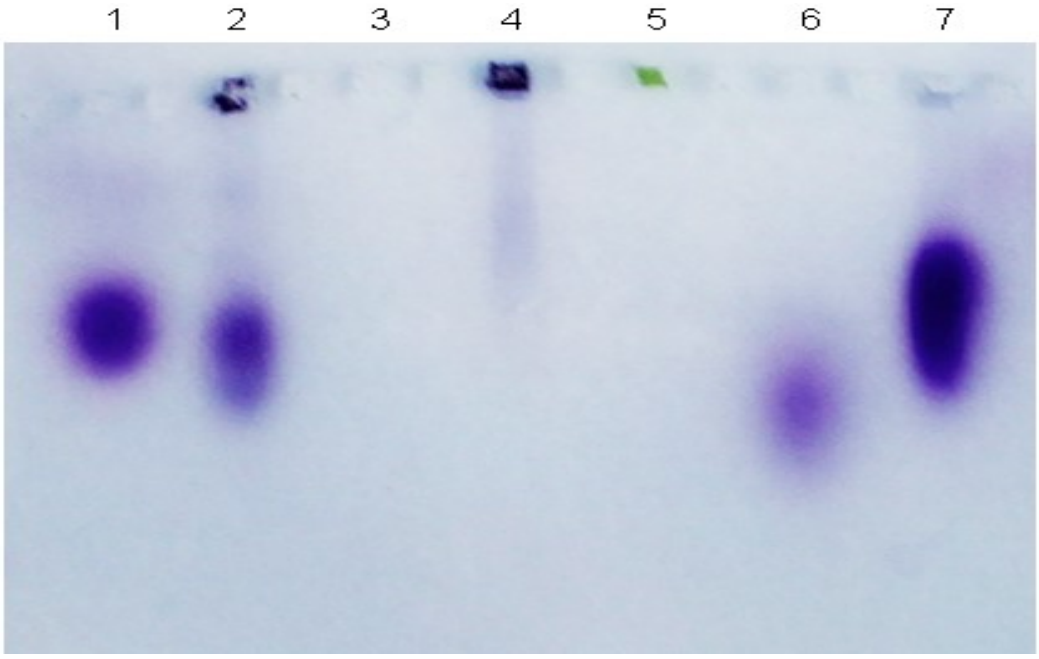
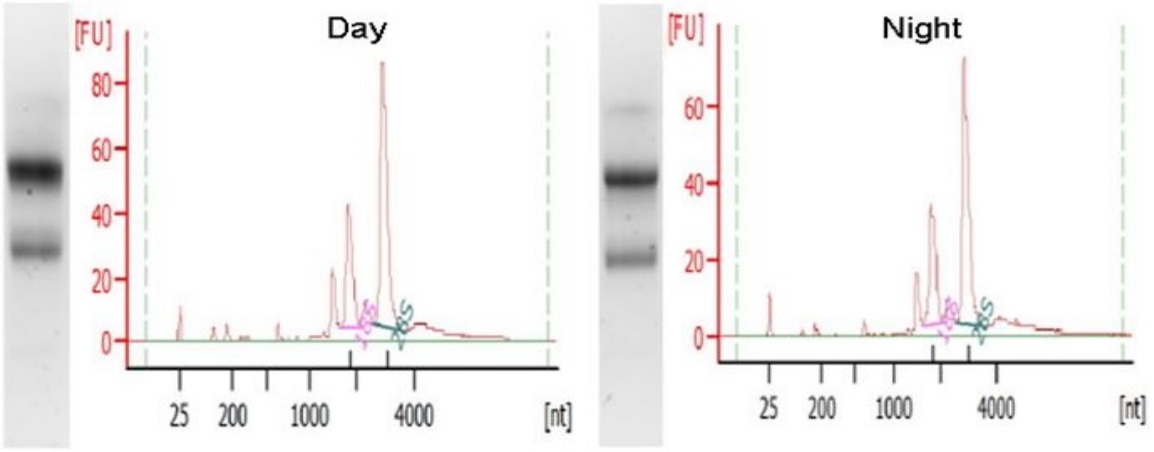


Figure 8. Agarose gel post-stained with toluidine blue O, demonstrate the presence of acidic and/ or sulfoconjugated polysaccharides in the alga *P. corallorhiza* & model system *A. thaliana*. 1. Heparan sulfate, 2. Sulfoconjugated arabinogalactan, 3. Non-sulfoconjugated galactan, 4. *P. corallorhiza* whole algal extract, 5. *A. thaliana* leaf extract, 6. *P. corallorhiza* SPs, 7. *P. corallorhiza* desulfoconjugated SPs.

3.2. RNA quantification and integrity analysis

RNA quantification and integrity analyses through denaturing formaldehyde agarose electrophoresis and analysis on the Agilent 2100 bioanalyzer determined that RNA extractions from the thallus material were of sufficient quality for further experimentation. The day temporal RNA results include a RNA area of 463.2, a concentration of 432 ng/µL, a 28S/18S ratio of 2.1 and a RIN of 6.5. Similarly, the RNA results for the nocturnal sample were as follow: RNA area of 352.1, a concentration of 328 ng/µL, a 28S/18S ratio of 1.9 and a RIN of 6.4 (Figure 9).



RNA area	463.2	352.1
RNA concentration	431 ng/µL	328 ng/µL
rRNA ratio (28S/18S)	2.1	1.9
RNA Integrity Number (RIN) *	6.5	6.4

Figure 9. *P. corallorhiza* RNA integrity analysis of both a day and nighttime sample. Formaldehyde agarose gel electrophoresis and Agilent bioanalyser 2100 analysis showing 18S and 28S rRNA fragments with accompanying RNA integrity number (RIN), sample concentration, and 28S/18S rRNA ratios. Fluorescence (FU) and nucleotides (nt). *(B.02.08, Anomaly Threshold(s) manually adapted).

3.3. RNA pre-processing, transcriptome assembly and annotation statistics

Deep sequencing of *P. corallorhiza* cDNA (one midday and one midnight sample) on the Illumina Hiseq2500 platform produced a total of 90 255 210 raw paired-end reads (Table 3). Sequencing reads were 125bp in length resulting in 22.56 Gb of RNA-seq data. 91 843 0576 reads averaging at 107 to 108 bp in length were obtained after removing defective base pairs at the beginning and ends followed by sequencing adapter removal and quality trimming (reads off poor quality were discarded - Phred scale quality score limit of 0.05 and trimming of ambiguous nucleotides with maximum number of ambiguities equal to two). Respectively 47 782 626 trimmed reads averaging at 107 bp in length and 44 060 431 trimmed reads with an average length of 108 were obtained for the day and the night sample.

De novo assembly of *P. corallorhiza* transcriptome (Table 3): A total of 91 843 057 paired-end reads comprising of the combined diurnal samples (one day sample with 5 technical replicates and one night sample with 5 technical replicates) were used for the de novo assembly in CLC bio genomics workbench (version 8.5.1). Assembly size was 65.46 Mb comprising of 119 675 assembled contigs. The contig size range was from a minimum of 108 bp to a maximum of 25 060 bp, with a N50 of 597 and GC content of 42.66%. Evaluation of the assembled transcriptome was executed by assessing how well the trimmed sequencing reads map back to the assembled contigs. In total, of the 91 843057 trimmed reads, 93.02% (85 047 771) mapped back to the 119 675 contigs either as intact pairs (84.52%) or as broken pairs (8.49%) with an average coverage of 134.78 fold. Contigs were filtered by size, bigger and equal to 900, resulting in a subset of 11 816 (9.87%) with contained a total of 19 037 open reading frames (ORFs).

The BLAST2GO pipeline was used for transcriptome annotation. Comprising of four main steps 1. BLASTx, 2. InterProScan, 3. Mapping and 4. Annotation, the contig subset ≥ 900 bp (11 816) was computationally annotated (Figure 11). BLASTx analysis against the NCBI non-redundant (nr) database (locally stored) with an E-value cut-off of 1E-5 resulted in hits for 77.2% of the contigs, while the remaining 22.80% had no BLAST hits. Using protein domain information housed by the European Bioinformatics Institute (EBI), the InterProScan analysis returned Gene Ontology terms for 97.99%

of the contigs. BLAST hit and protein domain information were merged and mapped, with 68.89% contig allocation.

Table 3. Summary data for Illumina RNA-sequencing, pre-processing, de novo transcriptome assembly, and evaluation.

	SUMMARIZED DATA	ALL SAMPLES
ILLUMINA SEQUENCING AND READ PRE-PROCESSING	Day sample reads (125 bp paired end)	48 000 000
	Night sample reads (125 bp paired end)	44 255 210
	Total Raw reads (125 bp paired end)	92 255 210
	After cleaning reads Day sample (107 bp paired end)	47 782 626
	After cleaning reads Night sample (108 bp paired end)	44 060 431
	Total cleaned reads (107 – 108 bp paired end)	91 843 057
DE NOVO ASSEMBLY AND EVALUATION	De novo assembled contigs	65 460 040 bp
	All contig count	119 675
	Contig count ≥ 900 (9.87%)	11 816
	Reads mapped back to all contigs (93.02%)	85 047 771 bp
	Reads mapped in pairs (84.52%)	77 280 812 bp
	Reads mapped in broken pair (8.49%)	7 766 959 bp
	Average coverage	134.78
	N50	597 bp
	Minimum contig length	108 bp
	Maximum contig length	25,063 bp
	GC content (%)	42.66
	ORF in contigs ≥ 900	19 037

Annotations were refined by annex-based GO term augmentation, which infers relationships between the allocated annotations (molecular function, biological

processes, and cellular components). Additionally, the GO slim function was employed in order to refine the allocated GO terms to a plant-based transcriptome, as algae are evolutionary more closely related to Viridiplantae (Merchant et al., 2010). Ultimately, 54.66% of the contigs were annotated (Figure 11).

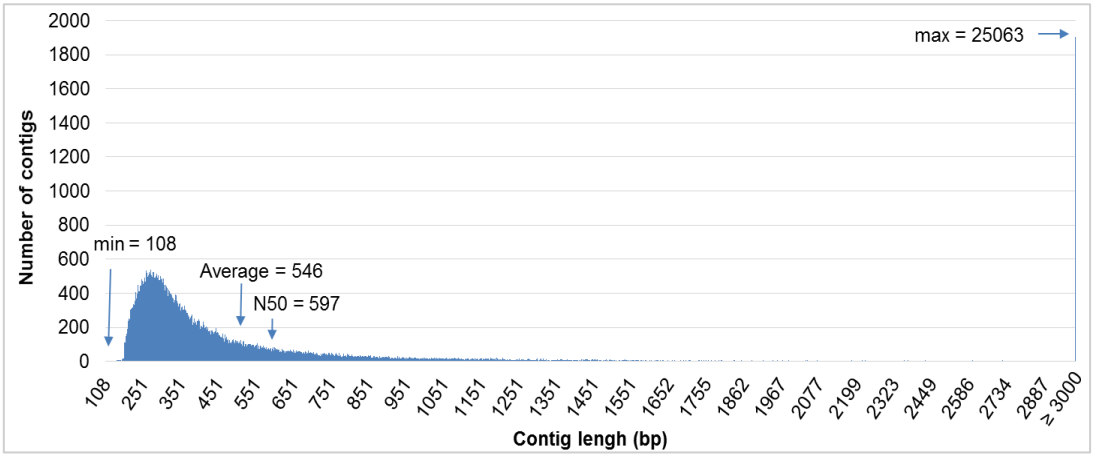


Figure 10. Analyses of contiguous assemblies. The distribution of 119 675 de novo assembled contigs covering 65.46 Mb, using a k-mer size of 24.

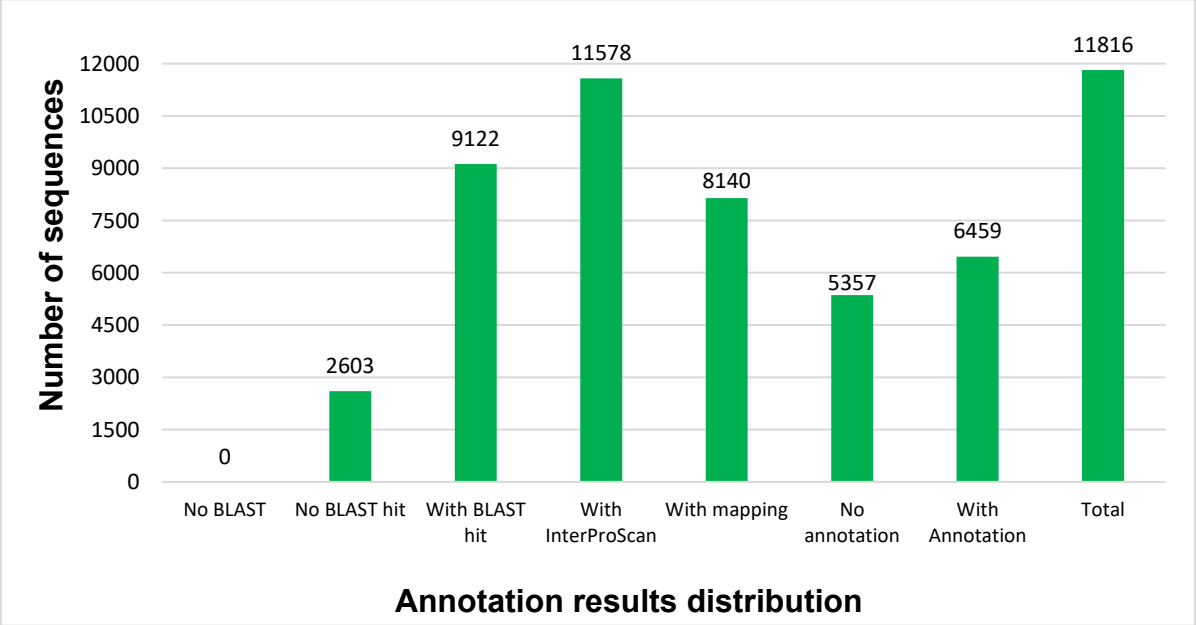


Figure 11. Annotations of contiguous assemblies through Gene Ontology (GO) term transfers. *P. corallorhiza* contig subset ≥ 900 was annotated through the four-step analysis in the BLAST2GO pipeline including BLASTx, InterProScan, mapping, and annotation. Annotations were refined by annex and GOslim execution.

Enzymes codes (EC) retrieved from the BLAST2GO annotation procedure (Figure 11) were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG), giving an

indication of the metabolic pathways represented in the draft *P. corallorhiza* transcriptome. The use of pathway level analysis, aids with the interpretation and integration of large-scale datasets, such those generated from RNA-seq or similar high throughput NGS techniques (Barrett, 1995; Kanehisa et al., 2012). Here, selected metabolic pathways are shown with the number of enzymes allocated to the particular pathway (Table 4). Secondly, the sum of the transcripts coding for the mapped enzymes is also indicated. Pathways relevant to carbohydrate metabolism includes the carbon fixation in photosynthetic organisms, glycolysis/ gluconeogenesis, citrate cycle (TCA cycle), starch and sucrose metabolism, fructose and mannose metabolism, pentose and glucuronate interconversions, galactose metabolism. Vital for sulfur assimilation, the sulfur metabolism, cysteine, and methionine metabolism pathways are also included. Further, the terpenoid backbone biosynthesis pathway is also listed.

3.4. Analysis of the draft *P. corallorhiza* transcriptome: metabolic pathways associated with algal SPs biosynthesis.

Table 4. Selected metabolic pathways associated with the *P. corallorhiza* transcriptome identified through KEGG pathway analysis.

SPECIFIC METABOLIC PATHWAYS	ENZYMES COUNT	TRANSCRIPT COUNT
CARBON FIXATION IN PHOTOSYNTHETIC ORGANISMS	15	38
GLYCOLYSIS/ GLUCONEOGENESIS	21	53
CITRATE CYCLE (TCA CYCLE)	15	37
STARCH AND SUCROSE METABOLISM	14	26
FRUCTOSE AND MANNOSE METABOLISM	11	21
PENTOSE AND GLUCURONATE INTERCONVERSIONS	6	10
GALACTOSE METABOLISM	8	13
SULFUR METABOLISM	6	9
CYSTEINE AND METHIONINE METABOLISM	21	43
TERPENOID BACKBONE BIOSYNTHESIS	13	13

A total of 14 enzymes were identified from the assembled contigs, which mapped to the starch and sucrose metabolic pathway. Enzymes mapped include a beta-amylase, which is responsible for hydrolysis of 1-4 alpha-D-glucosidic linkages in polymers releasing maltose units from the non-reducing terminal of starch and glucose phosphomutase which is responsible for the conversion of D-glucose 1-phosphate to D-glucose 6-phosphate vital for starch and sucrose synthesis. Further, starch phosphorylase, pectin depolymerase, UDP-glucose pyrophosphorylase, endo-1, 4-beta-D-glucanase, exo-1, 4-alpha-glucosidase and glucose-6-phosphate isomerase are among the other enzymes present (Figure 12).

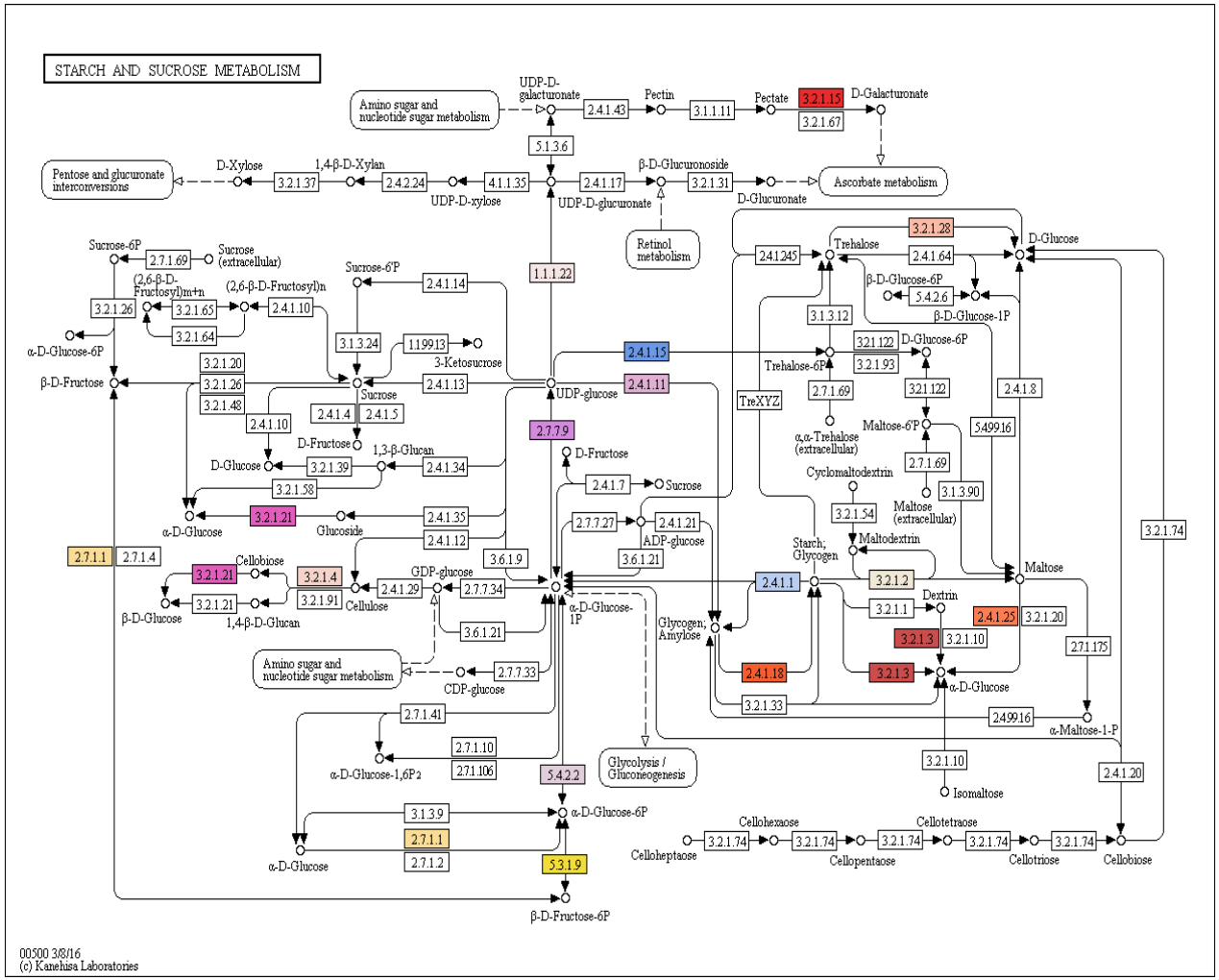
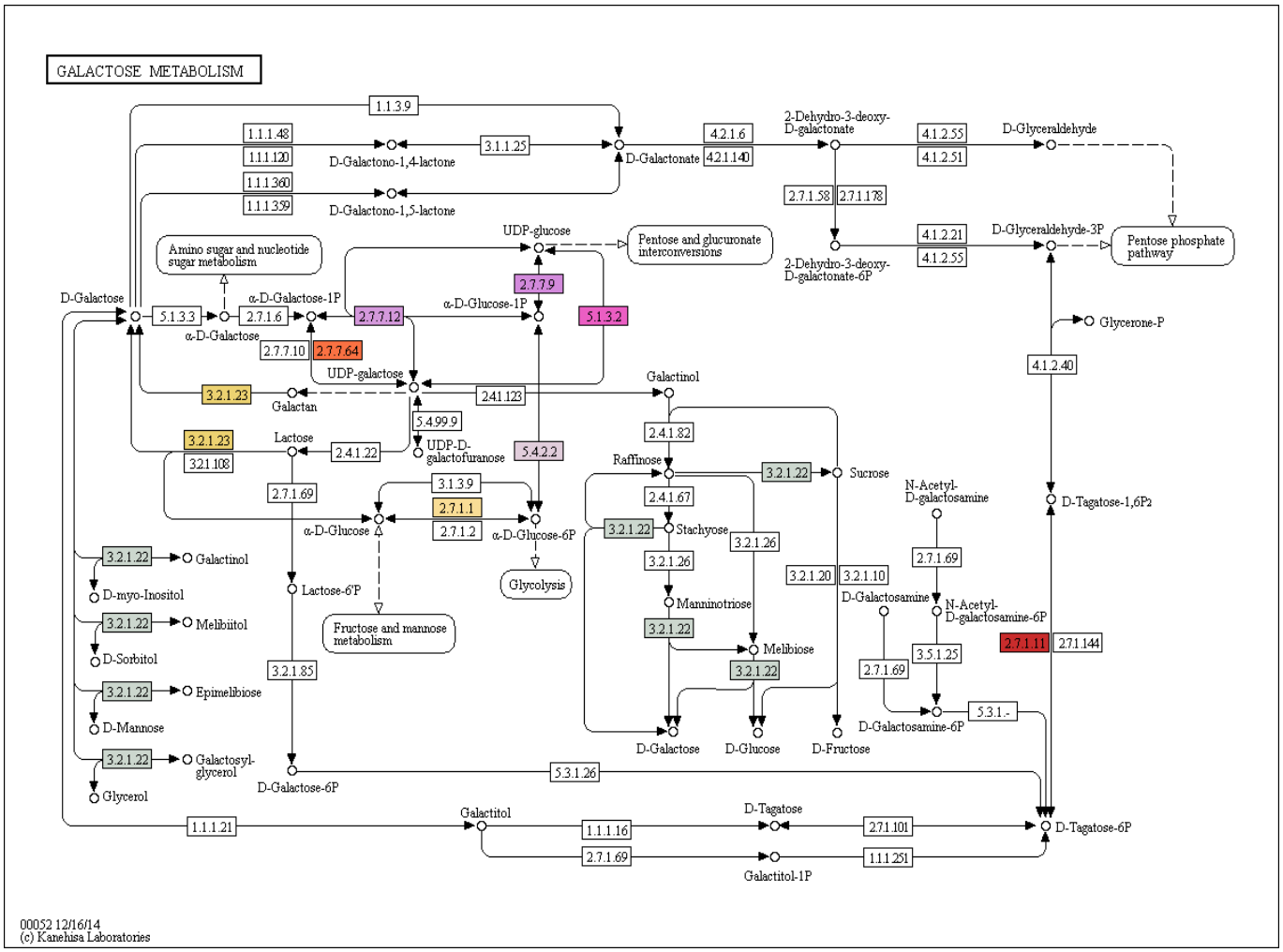


Figure 12. The KEGG starch and sucrose metabolic pathway with Enzyme Code (EC) assigned to contigs identified from thallus RNA. Coloured blocks represent enzymes encoded in specific contigs, with the EC to which the contig mapped.

ENZYME CODE (EC) LEGEND

2.7.1.1	hexokinase type IV glucokinase
2.7.7.9	UDP-glucose pyrophosphorylase
3.2.1.4	endo-1,4-beta-D-glucanase
3.2.1.3	exo-1,4-alpha-glucosidase
3.2.1.2	beta-amylase
5.3.1.9	glucose-6-phosphate isomerase
5.4.2.2	glucose phosphomutase
1.1.1.22	UDP-glucose 6-dehydrogenase
2.4.1.1	starch phosphorylase
2.4.1.18	glucan branching enzyme
2.4.1.15	trehalose 6-phosphate synthase
2.4.1.11	starch synthase
2.4.1.25	disproportionating enzyme
3.2.1.15	pectin depolymerase

The biosynthesis of Floridean starch, floridoside and SPs are believed to share UDP-D-galactose as a common precursor. In this regards, we considered enzymes associated with galactose metabolism as the major constituent sugar of many algal SPs is galactose and its derivatives (Li et al., 2001; Pomin and Mourão, 2008; Wijesekara et al., 2011). Overall, eight enzymes that play a role in galactose metabolism were identified from the annotated contigs (Figure 13). These included hexokinase type IV glucokinase, phosphohexokinase, UDP-glucose 4-epimerase, UDP-glucose pyrophosphorylase, galactose-1-phosphate uridylyltransferase, glucose phosphoglucomutase and alpha-galactosidase. With regard to their enzymatic role in galactose metabolism, UDP-glucose 4-epimerase is responsible for the interconversion of UDP-D-glucose to UDP-D-galactose, galactose-1-phosphate uridylyltransferase is responsible for catalysis of UDP-D-glucose + D-galactose-1-phosphate to D-glucose-1-phosphate + UDP-alpha-D-galactose and alpha-galactosidase is responsible for the hydrolysis of terminal, non-reducing D-galactose components from galactose containing oligosaccharides, polymers and glycolipids.

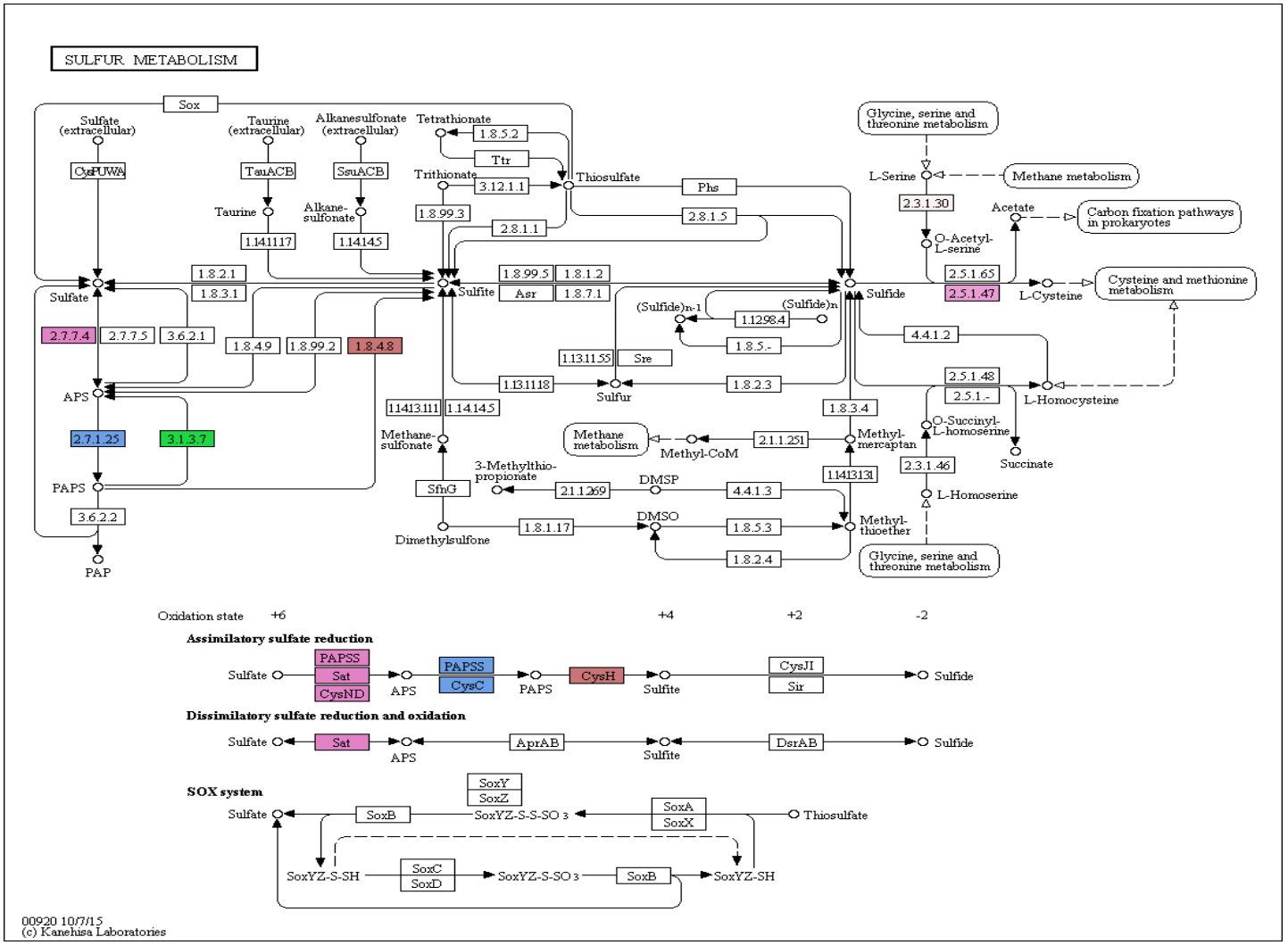


ENZYME CODE (EC) LEGEND

2.7.1.1	hexokinase type IV glucokinase
2.7.1.11	phosphohexokinase
5.1.3.2	UDP-glucose 4-epimerase
2.7.7.9	UDP-glucose pyrophosphorylase
2.7.7.64	UDP-galactose pyrophosphorylase
2.7.7.12	Galactose-1-phosphate uridylyltransferase
5.4.2.2	phosphoglucomutase
3.2.1.22	α-galactosidase

Figure 13. The KEGG galactose metabolic pathway with Enzyme Code (EC) assigned to contigs identified from thallus RNA. Coloured blocks represent enzymes encoded in specific contigs, with the EC to which the contig mapped.

Ultimately, many marine-derived polysaccharides undergo enzymatic modifications, with sulfoconjugation being among the most important, resulting in bioactive SPs (Pomin and Mourão, 2008; Wijesekara et al., 2011). Therefore, we considered enzymes associated with sulfur metabolism (Figure 14). Six such enzymes were identified from the annotated draft *P. corallorhiza* transcriptome. Enzymes comprised of the proteins responsible for the biosynthesis and degradation of the universal sulfate donor, 3'-phosphoadenosyl-5'-phosphosulfate (PAPS). PAPS Biosynthesis enzymes included ATP sulfurylase, responsible for the synthesis of the precursor adenylyl sulfate (APS) through the conversion of SO_4^{-2}- + ATP → APS + PPi and adenylyl-sulfate kinase, which through phosphorylation converts APS to PAPS. PAPS degradation enzymes include 3'-phosphoadenylylsulfate 3'-phosphatase which catalysis PAPS + H2O → adenosine monophosphate (AMP) + phosphate and PAPS reductase responsible for PAPS breakdown to APS and thioredoxin.



ENZYME CODE (EC) LEGEND	
2.7.1.25	adenylyl-sulfate kinase
1.8.4.8	PAPS reductase, thioredoxin-dependent
2.7.7.4	adenylylsulfate pyrophosphorylase
2.3.1.30	O-acetyl-transferase
3.1.3.7	3'-phospho-adenylylsulfate 3'-phosphatase
2.5.1.47	cysteine synthase

Figure 14. The KEGG sulfur metabolic pathway with Enzyme Code (EC) assigned to contigs identified from thallus RNA. Coloured blocks represent enzymes encoded in specific contigs, with the EC to which the contig mapped.

3.5. Mining of annotated contigs for sulfotransferase transcript identification

A total of 18 contigs encoding putative SULTs were identified based on nucleotide homology, protein domain information, and GO term assignments (Table 5). An ORF coding for a full-length SULT protein was absent in contig 21148. However, the remaining 17 contigs had full ORFs, coding for different SULTs. The sulfotransfer_1 family comprises of a range SULTs such as oestrogen SULTs, flavonyl 3-SULTs, and phenol SULTs. In contrast, SULTs represented in the sulfotransfer_2 family include heparan sulfate 2-O-SULT, heparan sulfate 6-O-SULT and chondroitin 6-SULT. The sulfotransfer_3 and Gal-3-O_sulfotr SULT families are represented by several tyrosylprotein SULTs and galactose-3-O-sulfotransferases, respectively. In total, two SULTs had no family allocation, four SULTs grouped only to sulfotransfer_1 family and an additional four SULTs grouped to both sulfotransfer_1 and sulfotransfer_3 families. Five SULTs group to the sulfotransfer_2 family and one SULT grouped to both the sulfotransfer_2 and Gal-3-O_sulfotr families. Only a single SULT grouped to the Gal-3-O_sulfotr family.

Table 5. Putative Sulfotransferase transcripts identified. A subset of contigs identified as SULTs through BLAST2GO annotation pipeline of the de novo assembled *P. corallorhiza* transcriptome.

CONTIG ID.	DESCRIPTION	ORF (BP)	EXPERIMENTAL EVIDENCE ON ORTHOLOGS AND E-VALUE
NO SULT FAMILY ALLOCATION			
21148*	SULT1	Partial ORF,	Zebrafish, SULT2 (Sugahara et al., 2003), 1E-10
1593	SULT2	519	-
***SULFOTRANSFER*_1**			
4888	SULT3 (GLST1)	1029	Human, SULT4A1 (Sidharthan et al., 2014), 3E-11
4938	SULT4	1008	Human, SULT4A1 (Sidharthan et al., 2014), 1e-15

TABLE 5 CONTINUED			
22698	SULT5	963	Human, SULT4A1 (Sidharthan et al., 2014), 2E-17
8436	SULT6	1440	Mouse, GalNAc4S-6ST (Mizumoto et al., 2013), 0.031
SULFOTRANSFER_2			
623	SULT11 (PCHST1)	1146	-
6083	SULT12	1050	-
2815	SULT13	1146	-
890	SULT14	858	Zebrafish, CHST14 (Mizumoto et al., 2009), 9E-26
2552	SULT15	1182	-
3135	SULT16	1554	-
GAL-3-O_SULFOTR			
18266	SULT17	948	Human, Gal3STs (Chandrasekaran et al., 2004), 0.004
SULFOTRANSFER_1* AND *SULFOTRANSFER_3			
8375	SULT7	1122	Canine, cSULT1B1 (Tsoi et al., 2001), 1E-28
7828	SULT8	1146	Zebrafish, SULT1 ST5 (Yasuda et al., 2005), 1E-35
237	SULT9	1023	Squid, GalNAc4S-6ST (Yamaguchi et al., 2007), 2E-11
11561	SULT10	1290	Human, GalNAc4S-6ST (Ohtake et al., 2003), 0.042
SULFOTRANSFER_2* AND *GAL-3-O_SULFOTR			
557	SULT18	1050	Human, HS6ST-2 (Habuchi et al., 2003), 2E-05

* Whole contig nucleotide sequence was subjected to the GenomeNet's motif search, using the Pfam database with a significance cut-off of 0.01.

3.6. ORF cloning of a *P. corallorhiza* glycolipid SULT (GLST1) and a putative carbohydrate SULT (PCHST1), grouping to the sulfotransfer_1 and _2 families, respectively

Following PCR amplification from thallus cDNA, the GLST1 and PCHST1 amplicons were ligated into pJET1.2 and transformed into *E. coli* DH5α (PCHST1-pJet1.2 and GLST1-pJet1.2), respectively. Colonies were PCR screened for the respective gene insertions (Figure 15). We obtained colonies for both the PGLST1::pJet1.2 and PCHST1::pJet1.2 constructs, confirming that the ORFS of both genes had been amplified from thallus cDNA. Both colonies screened for the PCHST1::pJet1.2 recombination event resulted in successful PCR amplifications. Only one of the five colonies screened for the PGLST1::pJet1.2 recombination event did not show positive PCR amplification.

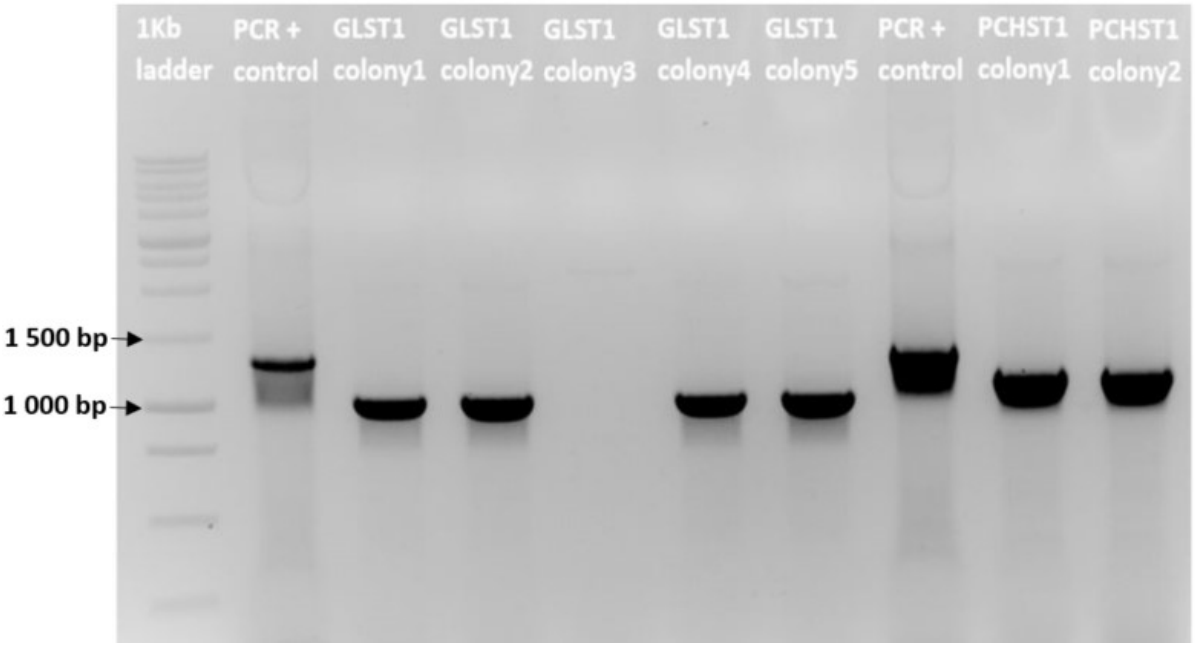


Figure 15. PCR amplification and isolation of *P. corallorhiza* SULT transcripts, GLST1 (1029 bp) and PCHST1 (1146 bp). PCR positive control gpST::pJET1.2*. PCR screening of colonies containing GLST1::pJet1.2 or PCHST1::pJet1.2 plasmid construct. *gpST (1415bp) refers to a SULT gene previously isolated from *Gelidium pristoides*.

3.7. Diurnal differential gene expression analysis: Single gene and gene network analysis.

A total of 636 nocturnal differentially expressed genes (DEGs) were identified, 86 nocturnally upregulated and 553 nocturnally downregulated. Table 6 lists the 40 nocturnal DEGs with highest log_2 fold change (20 upregulated and 20 downregulated) Upregulated DEG with the highest log2 fold change range from a cytochrome c oxidase subunit III, outer arm dynein beta heavy chain, NADH dehydrogenase subunit 3 (mitochondrion) and a 60S ribosomal L4 (partial) gene. Additional upregulated DEGs include an S-adenosylhomocysteine hydrolase, elongation factors, and disulfide-isomerase. Among the nocturnally downregulated DEGs with the greatest log_2 fold change are 5 heat shock proteins, a high light inducible protein, 3 chaperone proteins and a 2 tetratricopeptide repeat proteins. The Bonferroni and FDR corrected p-values are also listed for each DEG.

Table 6. Transcripts with the highest log2 fold change in the RNA-seq differential expression analysis for *P. corallorhiza* diurnal samples. Diurnal gene expression analysis was performed with Wald statistical analysis (log2 Fold change ≥ 5 or ≤ -5, with the Bonferroni and False Discovery Rates (FDR), corrected p-values ≤ 0.01). Transcripts identified using above criteria were considered as favourable candidates for potentially statistically significant DEGs.

UPREGULATED NIGHT TRANSCRIPTS

CONTIG ID	Log_2 fold change	FDR P-value*	Bonferroni p-value*	Description
CONTIG_1915	12.52	0	0	cytochrome c oxidase subunit III
CONTIG_3484	12.48	0	0	Outer arm dynein beta heavy chain
CONTIG_3434	12.37	0	0	NADH dehydrogenase subunit 3 (mitochondrion)
CONTIG_1961	11.39	4.96E-16	1.31E-12	60S ribosomal L4, partial
CONTIG_686	11.29	4.96E-16	1.31E-12	60S ribosomal L19
CONTIG_1556	10.63	0	0	cytochrome c oxidase subunit 1

TABLE 6 CONTINUED

CONTIG_1212	10.42	9.90E-14	2.94E-10	unnamed protein product [Chondrus crispus]
CONTIG_3822	10.18	0	0	28S ribosomal S2-like isoform 3
CONTIG_6456	9.92	1.47E-12	4.67E-09	No BLASTx hit
CONTIG_3230	9.78	0	0	60S ribosomal L3
CONTIG_7980	9.57	9.08E-12	3.02E-08	S-adenosylhomocysteine hydrolase
CONTIG_2205	9.44	0	0	translation elongation factor 2
CONTIG_16	9.42	0	0	unnamed protein product (*Chondrus crispus*)
CONTIG_1849	9.27	0	0	elongation factor 1-alpha
CONTIG_6759	9.23	5.10E-11	1.77E-07	No BLASTx hit
CONTIG_1346	9.18	0	0	No BLASTx hit
CONTIG_4539	9.17	6.63E-11	2.32E-07	Hsp90aa1 , partial
CONTIG_3227	9.06	1.18E-10	4.18E-07	radixin isoform X1
CONTIG_5497	9.01	1.48E-10	5.28E-07	Protein disulfide-isomerase
CONTIG_6736	9	1.55E-10	5.54E-07	V-type proton ATPase catalytic subunit A
	DOWNREGULATED NOCTURNAL TRANSCRIPTS			
CONTIG_65284	-11.09	0	0	heat shock protein 70 kDa 16 isoform X2
CONTIG_65773	-10.66	2.29E-14	6.56E-11	dnaJ homolog subfamily C member 7
CONTIG_30419	-10.26	0	0	heat shock protein 70 kDa
CONTIG_38221	-10.24	2.36E-13	7.15E-10	tetratricopeptide repeat protein
CONTIG_65635	-10.06	0	0	Molecular chaperone DnaJ
CONTIG_63444	-9.58	8.11E-12	2.69E-08	Hypothetical protein

TABLE 6 CONTINUED				
CONTIG_50738	-9.48	0	0	Chaperone dnaJ homolog subfamily B member 2 isoform X1
CONTIG_4874	-9.36	0	0	metalloproteinase inhibitor 2-like
CONTIG_38738	-9.28	0	0	tetratricopeptide repeat protein
CONTIG_6219	-9.16	6.58E-11	2.30E-07	Retrovirus-related Pol polyprotein from transposon TNT 1-94
CONTIG_65339	-9.14	7.36E-11	2.58E-07	Heat shock protein 70 kDa
CONTIG_48102	-9.06	0	0	Chaperone protein DnaK
CONTIG_142	-9.05	0	0	high light inducible protein
CONTIG_32054	-8.9	0	0	No BLASTx hit
CONTIG_65741	-8.8	0	0	No BLASTx hit
CONTIG_65445	-8.59	9.64E-10	3.68E-06	No BLASTx hit
CONTIG_42387	-8.58	1.01E-09	3.86E-06	unnamed protein product (Chondrus crispus)
CONTIG_66209	-8.44	0	0	No BLASTx hit
CONTIG_8226	-8.31	0	0	heat shock protein 70 kDa
CONTIG_26395	-8.29	0	0	Heat shock protein 70

*corrected p-values are not statistically significant, because of a lack of biological replicates for the RNA-seq expression analysis. However, they do provide surrogate estimates of gene expression.

From the 639 nocturnal DEGs, the 553 genes nocturnally downregulated were compared to the complete 11 816 contigs using the Fisher's exact test to search for enriched GO terms that might be present (Gotz et al., 2008). Figure 16 represent the enriched molecular function GO terms for the nocturnally downregulated DEG. GO terms enriched include nucleotide binding, which in regard to the molecular function hierarchy forms part of small molecule binding, which is also enriched and ultimately small molecule binding fall under molecular function through binding. Similarity, nucleotide binding fall under nucleoside phosphate binding, which is enriched here, forming part of organic cyclic and heterocyclic compound binding and ultimately converging on binding which forms part of the GO term molecular function (certain GO terms has been filtered out).

.

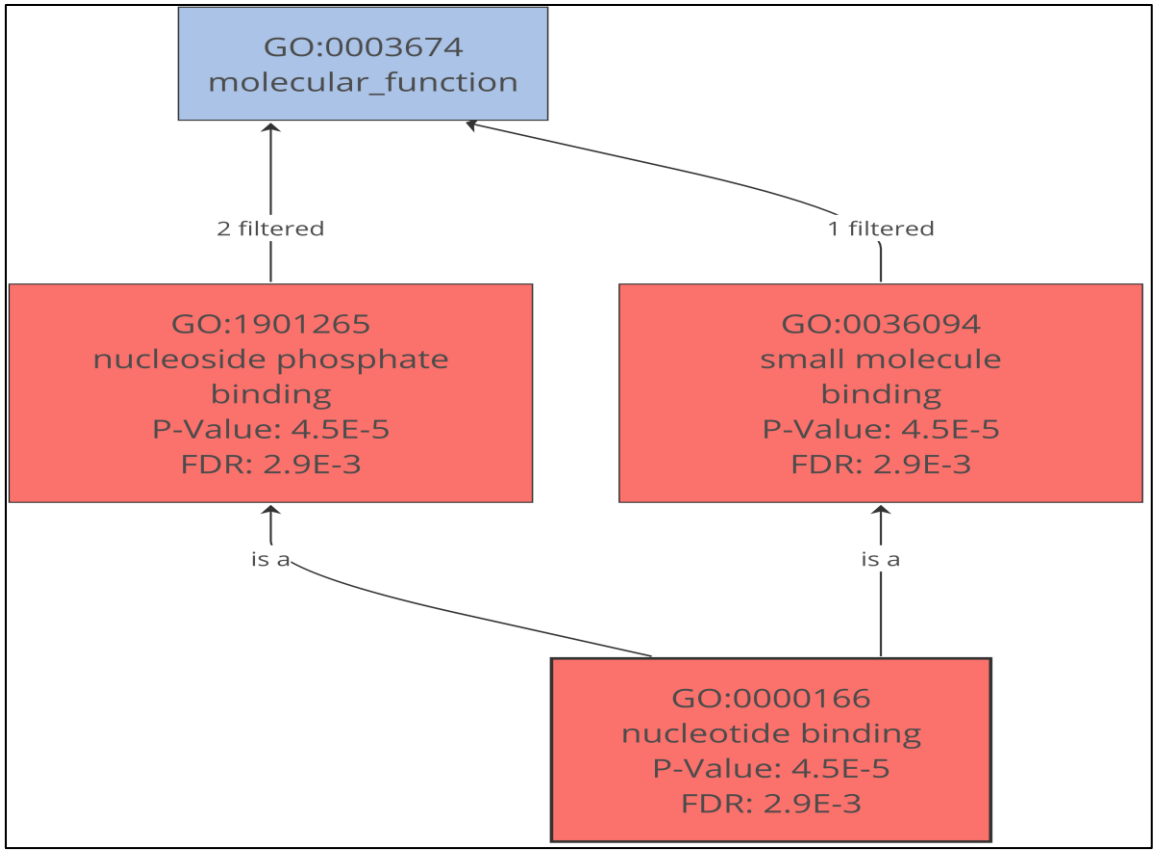


Figure 16. Directed acyclic graph of the Fisher's exact test for enriched Gene Ontology (GO) terms present on the nocturnally downregulated DEGs. Assessment of the GO terms for molecular function in the diurnal differentially expressed genes compared to all GO terms for all the transcripts. The Fishers Exact Test was used with a FDR, corrected p-value cut-off ≤ 0.01.

Of the 639 nocturnal DEGs, 86 DEGs were identified to be nocturnally upregulated in the diurnal differential gene expression analysis. Assessment of these 86 DEGs against the complete contig set i.e. the 11 816 contigs produced, using the Fisher's exact test for enriched Gene Ontology (GO) terms identified the following molecular

function GO terms (Figure 17). Among the enriched GO terms represented in the directed acyclic graph are RNA binding, nucleotide binding, receptor binding, nucleoside phosphate binding, and small molecule binding (certain GO terms has been filtered out).

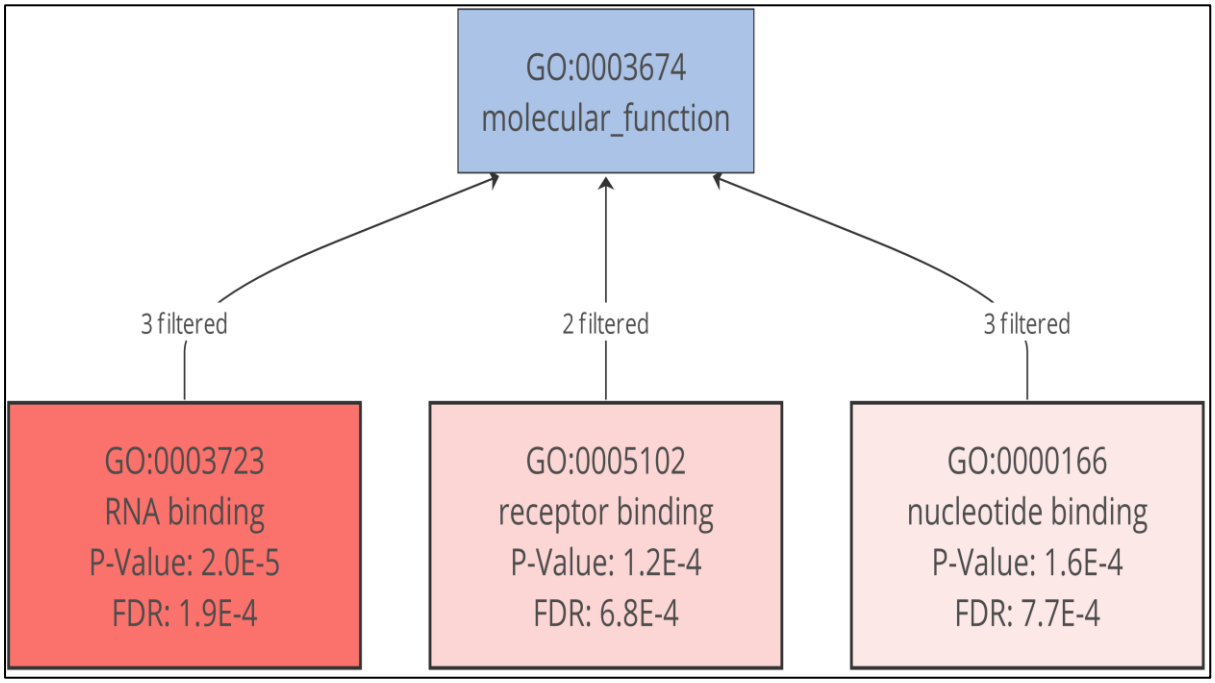


Figure 17. Directed acyclic graph of the Fisher's exact test for enriched Gene Ontology (GO) terms present on the nocturnally upregulated DEGs. Assessment of the GO terms for molecular function in the diurnal differentially expressed genes compared to all GO terms for all the transcripts. The Fishers Exact Test was used with a FDR, corrected p-value cut-off ≤ 0.01.

Figure 18 represents the enriched GO terms retrieved by Fisher's exact test analysis, comparing the nocturnally upregulated 86 DEGs against the complete contig data set of 11 816 contigs, for enriched Gene Ontology (GO) terms. Illustrated are the GO terms for cellular components that are enriched. Enriched GO terms include nucleus, endoplasmic reticulum, mitochondrion, ribosome and intracellular organelle. Ultimately, these enriched GOs converge in their hierarchy into the GO terms cell, membrane, extracellular region, macromolecular complex and finally cellular component (certain GO terms has been filtered out).

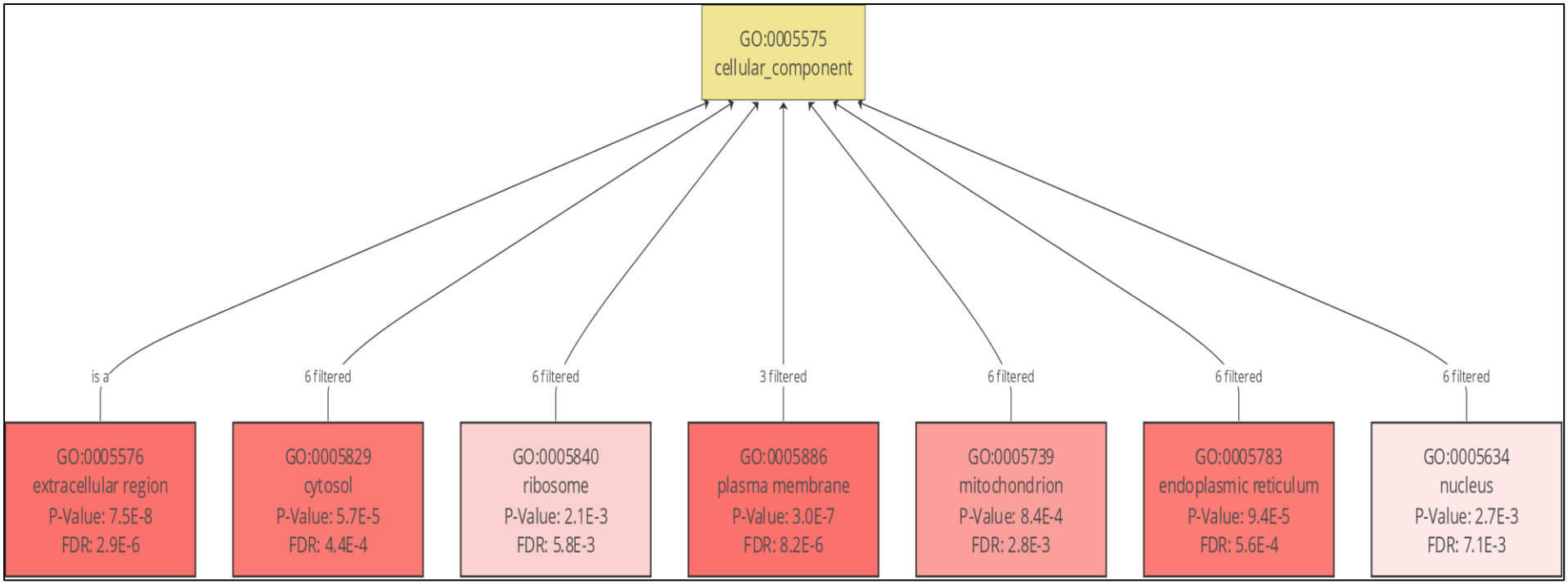


Figure 18. Directed acyclic graph of the Fisher's exact test for enriched Gene Ontology (GO) terms present on the nocturnally upregulated DEGs. Assessment of the GO terms for cellular component in the diurnal differentially expressed genes compared to all GO terms for all the transcripts. The Fishers Exact Test was used with a FDR, corrected p-value cut-off ≤ 0.01.

The nocturnally upregulated genes identified were analysed for enriched GO terms when compared to the complete contig list of 11 816 contigs, from which they were derived i.e. identified as DEGs (Figure 16 and 18). The assessment of DEGs using directed acyclic graphs of enriched GO terms, relating to biological processes provides the opportunity for interpreting the roles these DEGs play at a biological level (Figure 19). Biological processes GO terms enriched in the 86 nocturnally upregulated DEGs include signal transduction, regulation of macromolecule metabolic process, nucleobase-containing compound metabolic process, embryo development etc. All these enriched biological processes converge in the hierarchy onto the following GO terms i.e. metabolic

process, biological regulation, response to stimulus, cellular process, cellular component organization or biogenesis, single-organism process signalling, multicellular organismal process, developmental process and localisation (certain GO terms has been filtered out).

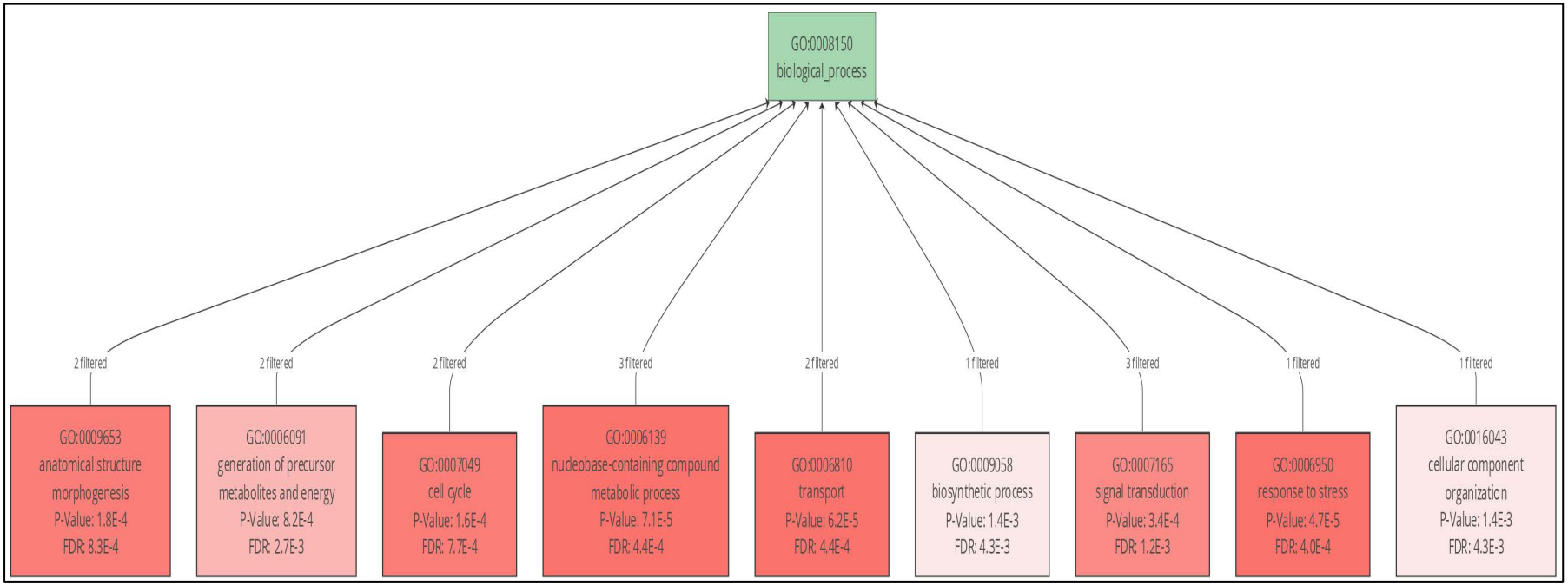


Figure 19. Directed acyclic graph of the Fisher's exact test for enriched Gene Ontology (GO) terms present on the nocturnally upregulated DEGs. *Assessment of the GO terms for biological processes in the diurnal differentially expressed genes compared to all GO terms for all the transcripts. The Fishers Exact Test was used with a FDR, corrected p-value cut-off ≤ 0.01.*

Chapter 4: Discussion

While a number of reports have functionally characterised SULTs from various organisms (Hernández-Sebastià et al., 2008) their isolation (and biochemical characterization) from macro red algae has, surprisingly, not been reported (Ho, 2015). This is despite macro algae been well described to accumulate bioactive SPs (Wijesekara et al., 2011) and the biotechnological and commercial potential of novel bioactive SPs. Only a single patent has thus far described an approach to use a recombinant SULT to produce SPs (WO2011128895A2, Arad et al., 2010). That patent claimed the successful overexpression of a native *Porphyridium sp*. SULT, resulting in a 38% higher sulfur content in SPs, when compared to the untransformed controls.

4.1. *Plocamium corallorhiza*, the generic Rhodophyta species used for SULTs identification, accumulates sulfoconjugated polysaccharides.

The majority of macro red algae studied to date synthesise SPs and frequently these SPs make up the majority of the algal dry biomass (Rhein-Knudsen et al., 2015). Our initial sampling identified six algal specimens in the field of which only two could be identified to species level (Figure 7). We chose to proceed with *Plocamium corallorhiza,* a ubiquitously distributed macro red alga as a generic representative of the phylum Rhodophyta. Since SP accumulation has been associated with diurnal patterns i.e. floridoside and Floridean starch assimilated during the day is converted into SPs during the night (Fournet et al., 1999; Hector, 2013; Li et al., 2001), we obtained thallus samples representing middle-of-day and middle-of-night.

To confirm that *P. corallorhiza* accumulates SPs, we employed a selective staining technique (toluidine blue O, Volpi and Maccari, 2002) to thallus extracts. Both heparan sulfate and sulfoconjugated arabinogalactan (positive controls) stained successfully (Figure 8). From cell extracts of *P. corallorhiza* and Arabidopsis (thallus and leaf, respectively) only those from *P. corallorhiza* stained positive for SPs. The SP content of various algae reaches up to 50% (dry weight basis) (Rhein-Knudsen et al., 2015). While terrestrial plants (like Arabidopsis) are known to accumulate SPs and other sulfoconjugated compounds such as the sulfoconjugated brassinosteroids, flavonols

and glucosinolates (Dantas-Santos et al., 2012; Hirschmann et al., 2014) they do not reach the concentrations described for algae. This could account for the negative result obtained for Arabidopsis leaf extracts.

We followed a desulfoconjugation protocol for *P. corallorhiza* extracts but failed to obtain samples, which did not stain positive. Native *P. corallorhiza* SP extracts and desulphoconjugated SP extracts were compared and differences in the staining intensities and gel migration distances between the samples were observed (Figure 8). The differences in sulfate content due to alkali pre-treatment can possibly explain the differences in gel migration distances, as negative charges are needed for electrophoretic mobility. According to Villanueva, Sousa, Gonçalves, Nilsson, & Hilliou, (2010), the NaOH concentration and treatment duration used in alkali pre-treatment greatly affects agar properties such as yield, sulfate content, 3,6-anhydrogalactose content, gel strength and agar gelling and melting temperatures. The greater staining intensity observed for the desulphoconjugated sample is likely due to an increased extraction efficiency because of the alkali pre-treatment, leading to a higher SP yield – as one microgram was loaded on the gel for all samples. However, the results obtained for the selective staining of algal SPs and polyanionic compounds should be interpreted with caution, as toluidine blue O is not selective to SPs only, but to most anionic compounds including complex glycosaminoglycans. However, in light of literature revealing the presence of SPs in algae from the same genera i.e. *Plocamium* and the ubiquitous nature of SPs in macro red algae, we proceeded with RNA extractions from the *P. corallorhiza* diurnal samples (Falshaw et al., 1999; Harden et al., 2009; Miller, 1999).

4.2. Assembling the transcriptome of *P. corallorhiza.*

The successes of previous similar NGS approaches in the identification of algal SULTs (Table 2) and the difficulties of a high throughput functional screening system for SULTs, led us to employ a high-throughput next-generation RNA sequencing approach for SULT gene identification (Wu et al., 2014, 2010).

Considered as a high-affinity, low-capacity pathway, SULTs forms part of the secondary detoxification pathways in organisms and typically are not expressed at high levels, under normal growth conditions (Gamage et al., 2006; Hirschmann et al.,

2014).Therefore, in order to overcome potential difficulties in identifying full-length SULT transcripts (because of low expression levels) deep sequencing on the Illumina HiSeq2500 platform was used. High-quality RNA-seq data was obtained and is exemplified in that of the 92.25 million paired-end reads obtained, 91.84 million reads remained after stringent quality trimming (Figure 9 and Table 3). Multiple red algal genomes and cDNA collections exist e.g. that of *Chondrus crispus*, *Cyanidioschyzon merolae,* and *Galderia sulphuraria* and the use of such reference sequences greatly facilitates RNA-seq assembly and annotation. However, attempts at using the *C. crispus*, *C. merolae* and *G. sulphuraria* genomes or cDNA collections as reference sequences resulted in assemblies with very low degree of reciprocal read mapping ("mappibility)" i.e. less than 10% of all RNA-seq reads could be mapped to the reference sequences (data not shown).

A de novo assembly strategy was subsequently used with a reciprocal read mapping of 93.02%, of which 84.52% of the reads mapping back in complete pairs. Similar RNA-seq reciprocal read mappibilities have also been obtained for the rubber tree (*Hevea brasiliensis*) and field pea (*Pisum sativum* L.) transcriptomes (Páez et al., 2015; Sudheesh et al., 2015). Ashrafi et al., (2012) and Williams and Baccarella, Parrish, & Kim, (2016) have shown that read trimming can significantly affect reciprocal read mappibility, the number of contigs produced and gene expression estimates. However, here we did not investigate the effects of Phred based read trimming parameters on downstream applications such as the length and number of contigs obtained and subsequent gene expression values (Ewing and Green, 1998). In total, an assembly size of 65.46 Mb comprising of 119 675 assembled contigs was obtained. Further, we also obtained a N50 of 597bp which is similar to the values obtained for 19 red algal species in a high throughput transcriptomic study - where large difference can also be seen between the different algal samples (Wu et al., 2014). In contrast, the contig count obtained here, i.e. 119 675 contigs, differed substantially to those of S. Wu et al., (2014), which obtained a minimum contig count of approximately 10 000 and a maximum of 50 000. These differences are likely due to methodological dissimilarities, especially since we did not employ GapCloser, which facilitates cap closing between scaffolds and ultimately leads to fewer sequences with longer lengths (Wu et al., 2014; Xie et al., 2014). As *A. thaliana* SULTs generally have an amino acid length ranging from 273 to 403, with an average length of 321 amino

acids, it was decided to only proceeded with contigs which have a length ≥ 900bp (Hirschmann et al., 2014). Subsequently, 11 816 (9.87%) contigs remained, which is characteristic of RNA-seq studies, as smaller contigs typically make up the majority of the total contigs produced, (Figure 10, Ashrafi et al., 2012; Páez et al., 2015).

4.3. In silico transcriptome annotation using sequence homology, protein domain information and Gene Ontology (GO) term assignment.

Valuable information can be extracted using a commercial in silico annotation pipeline (Ashrafi et al., 2012; Conesa et al., 2005). Figure 11 illustrates that 56% of 11 816 contigs were successfully annotated, providing the means to interrogate individual transcripts in addition to their combined role in galactan biosynthesis. The BLAST2GO annotation software combines three complementing strategies for annotation i.e. sequence homology, protein domain information and GO functional term assignment. Using this approach, the percentage of annotated contigs are higher in contrast to studies which only employed classical sequence homology techniques such as BLASTx (Gerdol et al., 2014; Gotz et al., 2008). In addition to the restricted amount of algal genomic and transcriptomic data available, RNA-seq read mappibility to either the *C. crispus*, *C. merolae* or *G. sulphuraria* genomic or cDNAs as reference sequences we surprisingly low i.e. less than 10% reciprocal read mappibility. To this end, we choose to use the NCBI nr databases, as it houses more sequence data from more diverse species (Ho, 2015). BLASTx hits were obtained for approximately 12 346 species, ranging from ciliate protozoa e.g. *Oxytricha trifallax* to the air-breathing freshwater snail *Biomphalaria glabrata* (Figure 20.A, Appendix A). The most significant BLASTx hits were obtained from approximately 966 species, with hits derived from *Paramecium tetraurelia*, *C. crispus,* and *Tetrahymena thermophila* being among the most prominent (Figure 20.B, Appendix A). The presence of BLASTx hits derived from various protozoa such as *P. tetraurelia*, *T. thermophile*, *Pseudocohnilembus persalinus* and *Ichthyophthirius multifiliis* are among the top hit species and likely indicates the presence of contamination prior to NGS sequencing. Contamination is likely as algae are known to have various organisms associated with them (Nakamura et al., 2013). Additionally, GC content analysis provides insight into organismal gene regulation and genome thermostability. In this regard, the estimated GC content of the *P. corallorhiza* transcriptome is 42.66% (Figure 21, Appendix A). The GC content of

the *P. corallorhiza* draft transcriptome is substantially lower than other Rhodophyta species such as *C. crispus* (53.50%), *Porphyridium purpuream* (56.50%) and *C. Merolae* (56.60%) (Zhang et al., 2015). Additionally, transcriptome coverage over GC content analysis (Figure 22, Appendix A), which is classically used to investigate the presence of contamination within NGS data, shows the presence of multiple GC content peaks. Three distinct peaks in the lower GC content range i.e. 6 %, 15%, 19%. These low GC content peaks likely indicate the presence of protozoa contamination as these organisms have characteristically low GC contents and are prominent within the BLASTx annotation analysis - where they are among the top-hit species (Xiong et al., 2015). Protozoa contamination may be responsible for the lower GC content observed for *the P. corallorhiza* transcriptome. Additionally, several other peaks are also present in the high GC content range. Possible sources of contamination include *Lingula anatina*, *Crassostrea gigas* and *Capitella teleta* as they are all top-hit species within the BLASTx analysis, Figure 20. B.

According to Conesa et al., (2016), the typical extent of annotation for RNA-seq studies ranges from 50% to 80% when complimenting strategies of annotation i.e. sequence homology, protein domain information and GO functional term assignment are used. Employing the BLAST2GO annotation pipeline, annotation results can be divided into either molecular function, cellular component or biological processes (Figure 23, Appendix A). Heterocyclic and organic cyclic compound binding are represented by the highest transcript count. Subsequently, for the category of cellular components, membrane-bound organelles are relatively more represented compared to non-membrane-bounded organelles. The prominent biological process represented within the *P. corallorhiza* transcriptome was primary metabolic processes, organic substance metabolic processes and cellular metabolic processes. Interestingly, the biological processes for establishment of location and response to stress were also very prominent (Figure 23, Appendix A). GO functional term distribution analysis i.e. whether a transcript has a particular molecular function, participate in certain broad biological processes or forms part of a specific cellular component, provides a very broad overview of the *P. corallorhiza* transcriptome. However, it does not provide specific information about key biosynthesis pathways e.g. galactan biosynthesis. Therefore, enzymes encoded within the transcriptome and their roles within biochemical pathways relevant to galactan metabolism were investigated. The

proposed Rhodophyta SP biosynthesis pathway i.e. agar biosynthesis (Figure 3). Key metabolic pathways represented within the *P. corallorhiza* transcriptome include carbon fixation in photosynthetic organisms, glycolysis and gluconeogenesis, the citrate cycle, starch and sucrose metabolism, fructose and mannose metabolism, pentose and glucuronate interconversions, galactose metabolism, sulfur metabolism and cysteine and methionine metabolism. According to Barbier et al., (2005) and Hector, (2013), the breakdown products of Floridean starch i.e. linear α-1,4-glucans provides the sugar monomers for subsequent algal SP biosynthesis. The carbon fixation pathway in photosynthetic organisms and the glycolysis and gluconeogenesis pathways play a pivotal role in Floridean starch biosynthesis and are represented by 21 enzymes encoded in 53 transcripts and 15 enzymes encoded by 38 transcripts, respectively (Table 4). Furthermore, the citric acid cycle is also prominently represented within the *P .corallorhiza* transcriptome. The citric acid cycle is responsible for the synthesis of GTP, the nucleotide-triphosphate needed for sugar activation, which is required for algal SP biosynthesis. Collectively, these carbohydrate metabolic pathways provide the sugars monomers needed for agar backbone synthesis, enabled by UTP and GTP-mediated activation (Barbier et al., 2005). Sulfur assimilation (Table 4) for the synthesis of the universal sulfate donor, 3'-Phosphoadenosine-5'-phosphosulfate (PAPS), needed for sugar backbone sulfoconjugation is achieved through the sulfur metabolism pathway and under sulfate starvation conditions, to a lesser extent through cysteine, but not methionine metabolism (Keidan et al., 2006). Interestingly, the family Plocamiaceae is well recognised for the synthesis of bioactive linear and cyclic halogenated monoterpenes and here we identified 13 enzymes responsible for terpenoid backbone biosynthesis (De Oliveira et al., 2015; Knott, 2015). KEGG mapping of enzymes involved in starch and sucrose metabolism, galactose metabolism and sulfur metabolism (responsible for providing key metabolites needed for agar biosynthesis, Figures 12 to 14).

4.4. Putative sulfotransferase transcript identification and subsequent cloning and validation.

Here we were able to identify 18 putative SULT transcripts encoded within the *P. corallorhiza* transcriptome. Although this represents a higher SULT count for any single alga observed presently e.g. *C. crispus* (12), *P. cruetum* (13), *E. Siliculosus*

(14). Large differences in SULT gene count have been previously observed among different alga species e.g. *Chlamydomonas reinhardtii* (8), *Ostreococcus lucimarinus* (2) and *Micromonas* sp.RCC299 (4) (Ho, 2015). Generally expressed at low levels under normal growth conditions and to a higher extent under stress conditions. SULTs are thought to play pivotal roles in organismal secondary metabolism and broad stress responses (Chen et al., 2012; Fallon et al., 2016). In order to counteract the low expression level of SULTs, deep sequencing (134x coverage) was used, as a means of obtaining full-length SULT transcripts. Deep sequencing along with the inadvertent stressing (illustrated by the prominence of the biological processes: establishment of location and response to stress) possibly contributed to the successful identification of these 18 putative SULTs.

The Pfam database houses approximately 538 plant SULT sequences. At 459, the majority of the SULT entries have a *Sulfotransfer*_1 domain. Subsequently, the Sulfotransfer_2 domain has 49 entries and the *Sulfotransfer*_3 domain has 16 entries. A similar trend in the number of SULTs identified with specific SULT domains was also observed here. Eight of the SULTs identified from the *P. corallorhiza* transcriptome had a *Sulfotransfer*_1 domain. However, four of the eight *Sulfotransfer*_1 SULTs also had a *Sulfotransfer*_3 domain. Additionally, seven SULTs had a *Sulfotransfer*_2 domain, with 1 also possessing a *Gal-3-O_sulfotr domain*. The overlapping catalytic domains observed for these SULTs likely highlights the inadequate algal SULT repositories currently available (Ho, 2015). However, the inability exclusive categorise these SULTs may also indicate incorrect data assemblies. Therefore, the SULTs identified with overlapping domains were not exclusively assigned to any one SULT family, as they had statistically significant homology (<0.05) to either the *Sulfotransfer*_1 and *Sulfotransfer*_3 or the *Sulfotransfer*_2 and the *Gal-3-O_sulfotr* domain containing SULT entries, respectively.

Hypothesised to play key roles in algal SP biosynthesis, Ho, (2015) compared algal sequences encoding SULTs, formylglycine-dependent sulfatases (FGly-SULFs) and putative sulfatase modifying factors (SUMFs), from 16 different algae. SULTs possessing *Sulfotransfer*_1 and *Sulfotransfer*_2 domains could be identified for all algae investigated, which included green, brown and red algal species. However, only one SULT containing the *Gal-3-O_sulfotr* domain, though to be characteristic of algae, was identified from the *C. crispus* genome (Collén et al., 2013; Ho, 2015). Surprisingly,

here we were also able to identify 2 SULTs which possessed the *Gal-3-O_sulfotr* domain, considered to be a rare SULT in algae and represented by only 11 entries within the UniProt/SwissProt database (Collén et al., 2013). Additionally, Formylglycine-dependent sulfatases are postulated to be responsible for the modification of the sulfate moieties on algal SPs, requiring the translational modification of cysteine residues to formylglycine for function, which is catalysed by SUMFs (Ho, 2015). In Ho, (2015), neither FGly-SULFs nor SUMFs could be identified for the red algae species investigated, but were identified for the green and brown algae species (Ho, 2015). Collén et al., (2013), concluded that the absence sulfatases from the carrageenophyte *C. crispus* genome suggest that carrageenan is not modified post-biosynthesis or may indicate the presence of sulfatases belonging to novel families within this alga. However, in addition to the SULTs identified for the *P. corallorhiza* transcriptome, we were also able to identify six sulfatases, possibly belonging to the aryl-sulfatase family, and three SUMFs, suggesting that *P. corallorhiza* SPs are modified, post-biosynthesis and sulfoconjugation (Data not shown). Of the 18 SULTs identified, two SULTs contained no conserved SULT domain and may represent pseudogenes (Table 5). Ultimately, we chose to PCR isolate two SULTs i.e. GLST1 and PCHST1 (Figure 15). As a result of technical issues or inadequate assemblies we were unable to isolate GLST1, but proceeded to PCR isolated PCHST1 followed by validation by Sanger sequencing. The retrieved sequencing data for PCHST1 was identical to the in silico predicted nucleotide sequence, except for one nucleotide residue. The in silico analyses predicted an adenosine residue (position 71 upstream); however, a guanine residue was observed for the Sanger sequencing data (data not shown). This discrepancy may be the result of a low coverage area or a single nucleotide polymorphism (SNP). However, the discrepancy does not seem to be unfavourable, because it is not present within the SULT catalytic domain and the resulting amino acid change was within the same group (polar amino acids with uncharged side chains, asparagine to serine)

4.5. Potential differences in diurnal gene expression estimates: Analysis of differentially expressed genes (DEGs) at gene and gene network level.

Only genes with a log2 fold change ≥ 5 (upregulated) or ≤ -5 (downregulated) and with Bonferroni and False Discovery Rates (FDR) corrected p-values ≤ 0.01, we considered differentially expressed. Our stringent parameters are due to the lack of biological replicates (increased possibility for false positive results), which is a prerequisite for differential gene expression studies (Robles et al., 2012). However, although the genes identified do not represent statistical significant DEGs, they provide surrogate estimates of gene expression and give insight into algal SP biosynthesis. In order to identify DEGs, the statistical distribution property of the underlying RNA-seq count data is vital. We employed the Wald test with log-transformed data (log2), as the underlying read counts possessed a Poisson distribution, typical of most RNA-seq studies (Chen et al., 2011; Gierliński et al., 2015). Furthermore, in addition to the log2 fivefold change considered for DEGs, we also employed Bonferroni and FDR corrected p-values as a conservative measurement for DEGs identification (Benjamin and Hochberg, 1994).

Table 6, lists the top 20 nocturnally upregulated and downregulated transcripts. The nocturnally upregulated DEGs includes various structural proteins such as ribosome proteins RPL3, RPL4 and RPL19 which are essential for protein synthesis (Provost et al., 2013). Increased protein synthesis is further corroborated by DEGs such as the translational elongation factor 1 and 2, which are needed for docking of the aminoacyl-tRNA to the A-site of ribosomes during protein synthesis. The high level expression, i.e. Log_2 fold change ≥ 9, of protein synthesis genes in the presence of S-adenosylhomocysteine hydrolase (responsible for the breakdown of S-adenosine homocysteine, which is considered a riboswitch target for subsequent gene expression regulation) and protein disulfide-isomerase (responsible for protein disulphide bond formation and correct protein 3D conformational folding); Highlights increase protein production nocturnally (Barnewitz et al., 2004; Wang and Breaker, 2008). The majority of the nocturnally downregulated transcripts code for heat shock and chaperone proteins, indicating a general stress response in the day temporal sample (Liu et al., 2014). This assumption is positively reinforced by the prominence

of the GO term “response to stress” being represented by 448 transcripts (Figure 23, Appendix A). No DEGs encoding enzymes that play a role in galactan biosynthesis through pathways such as starch and sucrose metabolism, galactose metabolism and sulfur metabolism were observed for the respective metabolic pathways (Figure 12 to 14).

In addition to simply investigating individual DEGs, gene set enrichment analysis (GSEA) for molecular functions, cellular components, and biological processes was employed on the 639 nocturnal DEGs (553 downregulated and 86 upregulated, Figures 16 to 19). GSEA provides the means to gain insight where single-gene analysis fails to find differences/similarities between conditional studies e.g. diurnal transcriptional changes (Gotz et al., 2008; Subramanian et al., 2005). Molecular functions that were significantly downregulated nocturnally includes nucleotide binding, nucleoside phosphate binding and small molecule binding (Figure 16). However, similar molecular functions were also significantly upregulated nocturnally i.e. nucleotide binding, nucleoside phosphate binding and small molecule binding (Figure 17). This overlap suggests Prominent cellular components that specific cellular process involving the selective interaction and non-covalent binding of nucleoside phosphates, small molecules and esterified nucleosides (ribose and deoxyribose) were diurnally differentially influenced.

It hard to reconcile Figure 16 with any specific biological processes. In contrast, corresponding to the upregulated DEGs listed in Table 6, the absence of RNA binding and receptor binding GO terms suggest increase protein biosynthesis nocturnally. Further, the 86 nocturnally upregulated DEGs were assessed with regard to the cellular components they code for (Figure 18). Prominent cellular components include the endoplasmic reticulum, mitochondrion, ribosome, and cytosol. Copious processes may lead to the biosynthesis of these cellular components making it difficult to link their biosynthesis to galactan metabolism. Further, carbohydrate SULTs typically have a type 2 transmembrane domain and are associated with the Golgi apparatus. However, the Golgi apparatus was not significantly represented within the nocturnally upregulated DEGs (Figure 18). In the end, a multitude of biological processes was diurnally influenced, according to GSEA for the nocturnally upregulated DEGs (Figure 19). However, no genes involved in starch and sucrose metabolism, galactose metabolism and sulfur metabolism were differentially expressed. These metabolic

pathways all contribute vital metabolites towards the proposed galactan biosynthesis pathway and their absence from the enriched gene sets makes it difficult to link galactan biosynthesis to any of the biological processes that are diurnally differentially influenced. Additionally, the GO terms used in GSEA are well established, with substantial experimental evidence (Bada et al., 2004; Schuurman and Leszczynski, 2008). In contrast, research on proposed galactan biosynthesis pathways is still in its infancy, resulting in a limited number of GO terms and enriched gene sets relating to galactan metabolism being available.

Chapter 5: Conclusion

With over 18 000 marine-derived natural products, and 4 900 patents associated with marine genetic resources, the increasing use of marine-derived products and genetic resources illustrates the biotechnological potential of one of earth's last remaining and largely untapped natural resources. Here, using high throughput next-generation RNA-sequencing technologies we identified 18 putative SULT transcripts and subsequently cloned one, putative carbohydrate SULT. Additionally, the assembly and annotation of a draft transcriptome for a ubiquitous macro red alga, *Plocamium corallorhiza* is also reported on here. Moreover, the inherent diverse substrate specificity of SULT enzymes characterised presently makes them very attractive tools for the synthesis of novel bioactive SPs. Phylogenetic analyses done on the *Galactose-3-O-sulphotransferase* genes identified from the macro red alga *C. crispus* and the brown alga *E. siliculosus* revealed that they form distinct clades when compared to their mammalian counterparts (Collén et al., 2013). This led the authors to propose that algal SULTs possibly have new enzymatic activities (Collén et al., 2013). To date, however, no algal SULTs have been isolated or biochemically characterised, especially with regard to their substrate specificity. This leaves a copious number of possible compounds that could be appropriate algal SULT substrates, some potentially displaying novel bioactivities. Here, we identified 18 SULTs from a macro red alga, *P. corallorhiza*, and cloned one for downstream functional characterization in pursuit of novel bioactive SPs.

Further work in this regard involved us heterologously expressing the carbohydrate SULTs obtained from this study in a yeast model (*Yarrowia lipolytica*) and, using recombinant protein *in vitro* to determine if we are able to sulfate various carbohydrate oligosaccharides (e.g. fructo-, galacto- and gluco-oligosaccharides) thereby rendering the bioactive.

Literature cited

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–10. doi:10.1016/S0022-2836(05)80360-2

Amado, I.R., Vázquez, J.A., González, M.P., Murado, M.A., 2013. Production of antihypertensive and antioxidant activities by enzymatic hydrolysis of protein concentrates recovered by ultrafiltration from cuttlefish processing wastewaters. Biochem. Eng. J. 76, 43–54. doi:10.1016/j.bej.2013.04.009

Aneiros, A., Garateix, A., 2004. Bioactive peptides from marine sources: Pharmacological properties and isolation procedures. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. doi:10.1016/j.jchromb.2003.11.005

Aquino, R.S., Grativol, C., Mourão, P.A.S., 2011. Rising from the sea: Correlations between sulfated polysaccharides and salinity in plants. PLoS One 6, e18862. doi:10.1371/journal.pone.0018862

Arad, S., Levy-Ontman, O., 2010. Red microalgal cell-wall polysaccharides: Biotechnological aspects. Curr. Opin. Biotechnol. 21, 358–364. doi:10.1016/j.copbio.2010.02.008

Arad, S., Plesser, L., Yaakov, W., 2010. Sulfotransferase of a red microalga and uses thereof [WWW Document]. doi:US 20100322867A1

Arrieta, J.M., Arnaud-Haond, S., Duarte, C.M., 2010. What lies underneath: conserving the oceans' genetic resources. Proc. Natl. Acad. Sci. U.S.A. 107, 18318–24. doi:10.1073/pnas.0911897107

Ashrafi, H., Hill, T., Stoffel, K., Kozik, A., Yao, J., Chin-Wo, S.R., Van Deynze, A., 2012. De novo assembly of the pepper transcriptome (Capsicum annuum): a benchmark for in silico discovery of SNPs, SSRs and candidate genes. BMC Genomics 13, 1–15. doi:10.1186/1471-2164-13-571

Bada, M., Stevens, R., Goble, C., Gil, Y., Ashburner, M., Blake, J.A., Cherry, J.M., Harris, M., Lewis, S., 2004. A short study on the success of the Gene Ontology. Web Semant. 1, 235–240. doi:10.1016/j.websem.2003.12.003

Barbier, G., Oesterhelt, C., Larson, M.D., Halgren, R.G., Wilkerson, C., Garavito, R.M., Benning, C., Weber, A.P.M., 2005. Comparative Genomics of Two Closely Related Unicellular Thermo-Acidophilic Red Algae, Galdieria sulphuraria and Cyanidioschyzon merolae, Reveals the Molecular Basis of the Metabolic Flexibility of Galdieria sulphuraria and Significant Differences in Carbo. Plant Physiol. 137, 460–474. doi:10.1104/pp.104.051169.460

Barnewitz, K., Guo, C., Sevvana, M., Ma, Q., Sheldrick, G.M., Soling, H.-D., Ferrari, D.M., 2004. Mapping of a Substrate Binding Site in the Protein Disulfide Isomerase-related Chaperone Wind Based on Protein Function and Crystal Structure. J. Biol. Chem. 279, 39829–39837. doi:10.1074/jbc.M406839200

Barrett, A.J., 1995. Enzyme Nomenclature. Recommendations 1992. Supplement 2: Corrections and Additions (1994). Eur. J. Biochem. 232, 1–1. doi:10.1111/j.1432-1033.1995.tb20774.x

Benjamin, Y., Hochberg, Y., 1994. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. doi:10.2307/2346101

Biobam, 2015. Blast2GO User Manual [WWW Document]. URL http://resources.qiagenbioinformatics.com/manuals/isv/Blast2GO/Blast2GO_PRO_User_Manual.pdf (accessed 10.10.16)

Bogdanich, W., 2008. The Drug Scare That Exposed a World of Hurt [WWW Document]. New York Times. URL http://www.nytimes.com/2008/03/30/weekinreview/30bogdanich.html (accessed 10.10.16)

Boucher, H.W., Talbot, G.H., Bradley, J.S., Edwards, J.E., Gilbert, D., Rice, L.B., Scheld, M., Spellberg, B., Bartlett, J., 2009. Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. Clin. Infect. Dis. 48, 1–12. doi:10.1086/595011

Branch, G., Griffiths, C., Beckley, L., 2013. Two Oceans: A guide to the marine life of southern Africa. Penguin Random House South Africa. doi:109876543

Brownlee, I. a, Allen, A., Pearson, J.P., Dettmar, P.W., Havler, M.E., Atherton, M.R., Onsøyen, E., 2005. Alginate as a source of dietary fiber. Crit. Rev. Food Sci.

Nutr. 45, 497–510. doi:10.1080/10408390500285673

Cadwallader, A.B., Yost, H.J., 2006. Combinatorial expression patterns of heparan sulfate sulfotransferases in zebrafish: II. The 6-O-sulfotransferase family. Dev. Dyn. 235, 3432–3437. doi:10.1002/dvdy.20990

Cao, X., Liao, Y., Rong, S., Hu, C., Zhang, X., Chen, R., Xu, Z., Gao, X., Li, L., Zhu, J., 2016. Identification and characterization of a novel abiotic stress responsive sulphotransferase gene (OsSOT9) from rice. Biotechnol. Biotechnol. Equip. 30, 227–235. doi:10.1080/13102818.2015.1136237

Carmely, S., Roll, M., Loya, Y., Kashman, Y., 1989. The structure of Eryloside A, a new antitumor and antifungal 4-methylated steroidal glycoside from the sponge Erylus lendenfeldi. J. Nat. Prod. 52, 167–170. doi:10.1021/np50061a022

Chandrasekaran, E. V., Lakhaman, S.S., Chawda, R., Piskorz, C.F., Neelamegham, S., Matta, K.L., 2004. Identification of Physiologically Relevant Substrates for Cloned Gal: 3-O-Sulfotransferases (Gal3STs) DISTINCT HIGH AFFINITY OF Gal3ST-2 and LS180 SULFOTRANSFERASE FOR THE GLOBO H BACKBONE, Gal3ST-3 FOR N-GLYCAN MULTITERMINAL Galbeta1,4GlcNAcbeta UNITS AN. J. Biol. Chem. 279, 10032–10041. doi:10.1074/jbc.M311989200

Chang, L., Sui, Z., Fu, F., Zhou, W., Wang, J., Kang, K.H., Zhang, S., Ma, J., 2014. Relationship between gene expression of UDP-glucose pyrophosphorylase and agar yield in Gracilariopsis lemaneiformis (Rhodophyta). J. Appl. Phycol. 1–7. doi:10.1007/s10811-014-0277-7

Chapman, E., Best, M.D., Hanson, S.R., Wong, C.H., 2004. Sulfotransferases: Structure, mechanism, biological activity, inhibition, and synthetic utility. Angew. Chemie - Int. Ed. 43, 3526–3548. doi:10.1002/anie.200300631

Chen, R., Jiang, Y., Dong, J., Zhang, X., Xiao, H., Xu, Z., Gao, X., 2012. Genome-wide analysis and environmental response profiling of SOT family genes in rice (Oryza sativa). Genes Genomics 34, 549–560. doi:10.1007/s13258-012-0053-5

Chen, Z., Liu, J., Ng, H.K.T., Nadarajah, S., Kaufman, H.L., Yang, J.Y., Deng, Y., 2011. Statistical methods on detecting differentially expressed genes for RNA-

seq data. BMC Syst. Biol. 5 Suppl 3, S1. doi:10.1186/1752-0509-5-S3-S1

Chikhi, R., Medvedev, P., 2014. Informed and automated k-mer size selection for genome assembly. Bioinformatics 30, 31–37. doi:10.1093/bioinformatics/btt310

Collén, J., Porcel, B., Carré, W., Ball, S.G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K., Elias, M., Artiguenave, F., Arun, A., Aury, J.-M., Barbosa-Neto, J.F., Bothwell, J.H., Bouget, F.-Y., Brillet, L., Cabello-Hurtado, F., Capella-Gutiérrez, S., Charrier, B., Cladière, L., Cock, J.M., Coelho, S.M., Colleoni, C., Czjzek, M., Da Silva, C., Delage, L., Denoeud, F., Deschamps, P., Dittami, S.M., Gabaldón, T., Gachon, C.M.M., Groisillier, A., Hervé, C., Jabbari, K., Katinka, M., Kloareg, B., Kowalczyk, N., Labadie, K., Leblanc, C., Lopez, P.J., McLachlan, D.H., Meslet-Cladiere, L., Moustafa, A., Nehr, Z., Nyvall Collén, P., Panaud, O., Partensky, F., Poulain, J., Rensing, S.A., Rousvoal, S., Samson, G., Symeonidi, A., Weissenbach, J., Zambounis, A., Wincker, P., Boyen, C., 2013. 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. 110, 5247–5252. doi:10.1073/pnas.1221259110

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676. doi:10.1093/bioinformatics/bti610

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., Mortazavi, A., 2016. A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 13. doi:10.1186/s13059-016-0881-8

Dantas-Santos, N., Gomes, D.L., Costa, L.S., Cordeiro, S.L., Costa, M.S.S.P., Trindade, E.S., Franco, C.R.C., Scortecci, K.C., Leite, E.L., Rocha, H.A.O., 2012. Freshwater plants synthesize sulfated polysaccharides: Heterogalactans from water hyacinth (Eicchornia crassipes). Int. J. Mol. Sci. 13, 961–976. doi:10.3390/ijms13010961

Daugherty, B.K., Bird, K.T., 1988. Salinity and temperature effects on agar

production from Gracilaria verrucosa Strain G-16. Aquaculture 75, 105–113. doi:10.1016/0044-8486(88)90025-7

Dawczynski, C., Schubert, R., Jahreis, G., 2007. Amino acids, fatty acids, and dietary fibre in edible seaweed products. Food Chem. 103, 891–899. doi:10.1016/j.foodchem.2006.09.041

De Oliveira, L.S., Tschoeke, D.A., De Oliveira, A.S., Hill, L.J., Paradas, W.C., Salgado, L.T., Thompson, C.C., Pereira, R.C., Thompson, F.L., 2015. New insights on the terpenome of the red seaweed Laurencia dendroidea (florideophyceae, rhodophyta). Mar. Drugs 13, 879–902. doi:10.3390/md13020879

Dzein Studio, 2010. Taurus Chemicals (Cape Kelp) (Pty) Ltd [WWW Document]. URL http://www.tcck.co.za/products.html (accessed 7.13.16)

Ehrhardt, D.W., Atkinson, E.M., Faull, K.F., Freedberg, D.I., Sutherlin, D.P., Armstrong, R., Long, S.R., 1995. In vitro sulfotransferase activity of NodH, a nodulation protein of Rhizobium meliloti required for host-specific nodulation. J. Bacteriol. 177, 6237–6245

Ernst, B., Magnani, J.L., 2009. From carbohydrate leads to glycomimetic drugs. Nat. Rev. Drug Discov. 8, 661–677. doi:10.1038/nrd2852

Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8, 186–194. doi:10.1101/gr.8.3.186

Fallon, T.R., Li, F., Vicent, M.A., Weng, J., 2016. Sulfoluciferin is Biosynthesized by a Specialized Luciferin Sulfotransferase in Fireflies. Biochemistry 55, 3341–3344. doi:10.1021/acs.biochem.6b00402

Falshaw, R., Furneaux, R.H., Miller, I.J., 1999. The backbone structure of the sulfated galactan from Plocamium costatum. Bot. Mar. 42, 431–435. doi:10.1515/bot.1999.049

Fournet, I., Zinoun, M., Deslandes, E., Diouris, M., Floc'h, J.Y., 1999. Floridean Starch and Carrageenan Contents as Responses of the Red Alga Solieria chordalis to Culture Conditions. Eur. J. Phycol. 34, 125–130. doi:10.1080/09670269910001736182

Freile-Pelegrin, Y., Murano, E., 2005. Agars from three species of Gracilaria (Rhodophyta) from Yucatan Peninsula. Bioresour. Technol. 96, 295–302. doi:10.1016/j.biortech.2004.04.010

Gamage, N., Barnett, A., Hempel, N., Duggleby, R.G., Windmill, K.F., Martin, J.L., McManus, M.E., 2006. Human sulfotransferases and their role in chemical metabolism. Toxicol. Sci. 90, 5–22. doi:10.1093/toxsci/kfj061

Gaull, G., 1989. Taurine in pediatric nutrition: review and update. Pediatrics 83, 433–442.

Ge, B., Qin, S., Han, L., Lin, F., Ren, Y., 2006. Antioxidant properties of recombinant allophycocyanin expressed in Escherichia coli. J. Photochem. Photobiol. B Biol. 84, 175–180. doi:10.1016/j.jphotobiol.2006.02.008

Gerdol, M., De Moro, G., Manfrin, C., Milandri, A., Riccardi, E., Beran, A., Venier, P., Pallavicini, A., 2014. RNA sequencing and de novo assembly of the digestive gland transcriptome in Mytilus galloprovincialis fed with toxinogenic and non-toxic strains of Alexandrium minutum. BMC Res. Notes 7, 722. doi:10.1186/1756-0500-7-722

Ghosh, T., Chattopadhyay, K., Marschall, M., Karmakar, P., Mandal, P., Ray, B., 2009. Focus on antivirally active sulfated polysaccharides: From structure-activity analysis to clinical evaluation. Glycobiology 19, 2–15. doi:10.1093/glycob/cwn092

Gierliński, M., Cole, C., Schofield, P., Schurch, N.J., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G., Owen-Hughes, T., Blaxter, M., Barton, G.J., 2015. Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment. Bioinformatics 31, 3625–3630. doi:10.1093/bioinformatics/btv425

Glatt, H., 2000. Sulfotransferases in the bioactivation of xenobiotics. Chem. Biol. Interact. 129, 141–170. doi:10.1016/S0009-2797(00)00202-7

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36,

3420–3435. doi:10.1093/nar/gkn176

Habuchi, H., Miyake, G., Nogami, K., Kuroiwa, A., Matsuda, Y., Kusche-Gullberg, M., Habuchi, O., Tanaka, M., Kimata, K., 2003. Biosynthesis of heparan sulphate with diverse structures and functions: two alternatively spliced forms of human heparan sulphate 6-O-sulphotransferase-2 having different expression patterns and properties. Biochem. J. 371, 131–142. doi:10.1042/BJ20021259

Hamed, I., Özogul, F., Özogul, Y., Regenstein, J.M., 2015. Marine Bioactive Compounds and Their Health Benefits: A Review. Compr. Rev. Food Sci. Food Saf. 14, 446–465. doi:10.1111/1541-4337.12136

Harden, E.A., Falshaw, R., Carnachan, S.M., Kern, E.R., Prichard, M.N., 2009. Virucidal activity of polysaccharide extracts from four algal species against herpes simplex virus. Antiviral Res. 83, 282–289. doi:10.1016/j.antiviral.2009.06.007

Harnedy, P.A., Fitzgerald, R.J., 2011. Bioactive proteins, peptides, and amino acids from macroalgae. J. Phycol. 47, 218–232. doi:10.1111/j.1529-8817.2011.00969.x

Hayashi, K., Hayashi, T., Kojima, I., 1996. A natural sulfated polysaccharide, calcium spirulan, isolated from Spirulina platensis: In vitro and ex vivo evaluation of anti-herpes simplex virus and anti-human immunodeficiency virus activities. AIDS Res. Hum. Retroviruses 12, 1463–1471. doi:10.1089/aid.1996.12.1463

Hayes, M., 2012. Marine bioactive compounds: Sources, characterization and applications, Marine Bioactive Compounds: Sources, Characterization and Applications. doi:10.1007/978-1-4614-1247-2

Hayes, M., Tiwari, B.K., 2015. Bioactive carbohydrates and peptides in foods: An overview of sources, downstream processing steps and associated bioactivities. Int. J. Mol. Sci. 16, 22485–22508. doi:10.3390/ijms160922485

Hector, S., 2013. Molecular studies of galactan biosynthesis in red algae.

Hernández-Sebastià, C., Varin, L., Marsolais, F., 2008. “Sulfotransferases from plants, algae and phototrophic bacteria,” in Sulfur Metabolism in Phototrophic Organisms., in: Hell, R., Dahl, C., Knaff, D., Leustek, T. (Eds.),.

Amsterdam:Springer, pp. 146–538. doi:10.1016/S0065-2911(08)00002-7

Hernàndez-Sebastiá, C., Varin, L., Marsolais, F., 2008. Sulfotransferases from Plants, Algae and Phototrophic Bacteria as in Sulfur Metabolism in Phototrophic Organisms, in: Sulfur Metabolism in Phototrophic Organisms. pp. 111–130. doi:10.1007/978-1-4020-6863-8_6

Hirschmann, F., Krause, F., Papenbrock, J., 2014. The multi-protein family of sulfotransferases in plants: composition, occurrence, substrate specificity, and functions. Front. Plant Sci. 5, 556. doi:10.3389/fpls.2014.00556

Ho, C., 2015. Phylogeny of Algal Sequences Encoding Carbohydrate Sulfotransferases, Formylglycine-Dependent Sulfatases, and Putative Sulfatase Modifying Factors. Front. Plant Sci. 6, 1057. doi:10.3389/fpls.2015.01057

Holsclaw, C.M., Sogi, K.M., Gilmore, S.A., Schelle, M.W., Leavell, M.D., Bertozzi, C.R., Leary, J.A., 2008. Structural Characterization of a Novel Sulfated Menaquinone produced by stf3 from Mycobacterium tuberculosis. ACS Chem. Biol. 3, 619–624. doi:10.1021/cb800145r

Hooper, L., Martin, N., Abdelhamid, A., Davey Smith, G., 2015. Reduction in saturated fat intake for cardiovascular disease (review), in: Hooper, L. (Ed.), Cochrane Database of Systematic Reviews. John Wiley & Sons, Ltd, Chichester, UK, pp. 6–9. doi:10.1002/14651858.CD011737

Illumina, I., 2013. TruSeq Stranded mRNA Sample Preparation ® [WWW Document]. URL http://www.utsouthwestern.edu/labs/next-generation-sequencing-core/assets/truseq-stranded-mrna-sample-prep-guide.pdf (accessed 10.10.16)

Illumina Inc., 2013. HiSeq ® 2500 Sequencing System [WWW Document]. URL https://www.illumina.com/documents/products/datasheets/datasheet_hiseq2500.pdf (accessed 10.10.16)

Itokawa, H., Morris-Natschke, S.L., Akiyama, T., Lee, K.-H., 2008. Plant-derived natural product research aimed at new drug discovery. J. Nat. Med. 62, 263–280. doi:10.1007/s11418-008-0246-z

Jančová, P., Šiller, M., 2012. Phase II Drug Metabolism. Top. Drug Metab. 35–60.

doi:10.5772/29996

Jiao, G., Yu, G., Zhang, J., Ewart, H.S., 2011. Chemical structures and bioactivities of sulfated polysaccharides from marine algae. Mar. Drugs 9, 196–233. doi:10.3390/md9020196

Kanehisa, M., Goto, S., Kawashima, S., Nakaya, A., 2002. The KEGG databases at GenomeNet. Nucleic Acids Res. 30, 42–6. doi:10.1093/nar/30.1.42

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M., 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40, 1–6. doi:10.1093/nar/gkr988

Keidan, M., Broshy, H., van Moppes, D., Arad, S., 2006. Assimilation of sulphur into the cell-wall polysaccharide of the red microalga Porphyridium sp. (Rhodophyta). Phycologia 45, 505–511. doi:10.2216/05-57.1

Kitazawa, K., Tryfona, T., Yoshimi, Y., Hayashi, Y., Kawauchi, S., Antonov, L., Tanaka, H., Takahashi, T., Kaneko, S., Dupree, P., Tsumuraya, Y., Kotake, T., 2013. Beta-Galactosyl Yariv Reagent Binds to the Beta-1,3-Galactan of Arabinogalactan Proteins. Plant Physiol. 161, 1117–1126. doi:10.1104/pp.112.211722

Knott, M.G., 2015. A review of secondary metabolites isolated from Plocamium species worldwide 75–93

Kraan, S., 2012. Algal Polysaccharides , Novel Applications and Outlook. doi:10.5772/51572

Kris-Etherton, P.M., Harris, W.S., Appel, L.J., 2002. Fish consumption, fish oil, omega-3 fatty acids, and cardiovascular disease. Circulation 106, 2747–2757. doi:10.1161/01.CIR.0000038493.65177.94

Lee, J.B., Ohta, Y., Hayashi, K., Hayashi, T., 2010. Immunostimulating effects of a sulfated galactan from *Codium fragile*. Carbohydr. Res. 345, 1452–1454. doi:10.1016/j.carres.2010.02.026

Lee, S., Lee, Y.S., Jung, S.H., Kang, S.S., Shin, K.H., 2003. Anti-oxidant activities of fucosterol from the marine algae *Pelvetia siliquosa*. Arch. Pharm. Res. 26, 719–722. doi:10.1007/BF02976680

Lee, Y.S., Shin, K.H., Kim, B.-K., Lee, S., 2004. Anti-diabetic activities of fucosterol from Pelvetia siliquosa. Arch. Pharm. Res. 27, 1120–1122. doi:10.1007/BF02975115

Lehmann, U., Robin, F., 2007. Slowly digestible starch - its structure and health implications: a review. Trends Food Sci. Technol. 18, 346–355. doi:10.1016/j.tifs.2007.02.009

Leslie, I., 2016. The sugar conspiracy [WWW Document]. URL http://www.theguardian.com/society/2016/apr/07/the-sugar-conspiracy-robert-lustig-john-yudkin (accessed 6.18.16)

Li, J.-J., Huang, C.J., Xie, D., 2008. Anti-obesity effects of conjugated linoleic acid, docosahexaenoic acid, and eicosapentaenoic acid. Mol. Nutr. Food Res. 52, 631–645. doi:10.1002/mnfr.200700399

Li, S.Y., Lellouche, J.P., Shabtai, Y., Arad, S., 2001. Fixed carbon partitioning in the red microalga porphyridium sp. (Rhodophyta). J. Phycol. 37, 289–297. doi:10.1046/j.1529-8817.2001.037002289.x

Liu, F., Wang, W., Sun, X., Liang, Z., Wang, F., 2014. RNA-Seq revealed complex response to heat stress on transcriptomic level in Saccharina japonica (Laminariales, Phaeophyta). J. Appl. Phycol. 26, 1585–1596. doi:10.1007/s10811-013-0188-z

Lordan, S., Ross, R.P., Stanton, C., 2011. Marine bioactives as functional food ingredients: Potential to reduce the incidence of chronic diseases. Mar. Drugs 9, 1056–1100. doi:10.3390/md9061056

Lustig, R.H., Schmidt, L. a., Brindis, C.D., 2012. Public health: The toxic truth about sugar. Nature 482, 27–29. doi:10.1038/482027a

Manley, S.L., Burns, D.J., 1991. Formation of nucleoside diphosphate monosaccharides (NDP-sugars) by the agarophyte *Pterocladia capillacea* (Rhodophyceae). J. Phycol. 27, 702–709. doi:10.1111/j.0022-3646.1991.00702.x

Mao, W., Zang, X., Li, Y., Zhang, H., 2006. Sulfated polysaccharides from marine green algae Ulva conglobata and their anticoagulant activity. J. Appl. Phycol. 18,

9–14. doi:10.1007/s10811-005-9008-4

Marinho-Soriano, E., 2001. Agar polysaccharides from Gracilaria species (Rhodophyta, Gracilariaceae). J. Biotechnol. 89, 81–84. doi:10.1016/S0168-1656(01)00255-3

Marinho-Soriano, E., Bourret, E., De Casabianca, M.L., Maury, L., 1999. Agar from the reproductive and vegetative stages of Gracilaria bursa-pastoris. Bioresour. Technol. 67, 1–5. doi:10.1016/S0960-8524(99)00092-9

Marsolais, F., Varin, L., 1995. Identification of amino acid residues critical for catalysis and cosubstrate binding in the flavonol 3-sulfotransferase. J. Biol. Chem. 270, 30458–30463. doi:10.1074/jbc.270.51.30458

Mascher, T., 2014. Competent E. Coli DH5a [WWW Document]. URL http://www.mclab.com/Dh5-Alpha-Competent-E.-Coli.html (accessed 5.23.15)

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, J., Witman, G.B., Terry, A., Salamov, A., Fritz-laylin, L.K., Maréchal-drouard, L., Marshall, W.F., Qu, L., Nelson, D.R., Sanderfoot, A., Spalding, M.H., Kapitonov, V. V, Ren, Q., Cardol, P., Cerutti, H., Chanfreau, G., Ferris, P., Fukuzawa, H., González-ballester, D., Mueller-roeber, B., Rajamani, S., Sayre, R.T., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y.W., Luo, Y., Martínez, D., Chi, W., Ngau, A., Otillar, B., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Igor, V., Rokhsar, D.S., Grossman, A.R., 2010. The Chlamydomonas Genome Reveals the Evolution of Key Animal and Plant Functions. Science (80). 318, 245–250. doi:10.1126/science.1143609

Miller, I.J., 1999. Further evaluation of the structure of the polysaccharide from Plocamium costatum with the use of set theory. Hydrobiologia 398/399, 385–389. doi:10.1023/A:1017051405258

Mizumoto, S., Mikami, T., Yasunaga, D., Kobayashi, N., Yamauchi, H., Miyake, A., Itoh, N., Kitagawa, H., Sugahara, K., 2009. Chondroitin 4-O-sulfotransferase-1 is required for somitic muscle development and motor axon guidance in zebrafish. Biochem. J. 419, 387–99. doi:10.1042/BJ20081639

Mizumoto, S., Watanabe, M., Yamada, S., Sugahara, K., 2013. Expression of N-

acetylgalactosamine 4-sulfate 6-O-sulfotransferase involved in chondroitin sulfate synthesis is responsible for pulmonary metastasis. Biomed Res. Int. 2013, 656319. doi:10.1155/2013/656319

Nakamura, Y., Sasaki, N., Kobayashi, M., Ojima, N., Yasuike, M., Shigenobu, Y., Satomi, M., Fukuma, Y., Shiwaku, K., Tsujimoto, A., Kobayashi, T., Nakayama, I., Ito, F., Nakajima, K., Sano, M., Wada, T., Kuhara, S., Inouye, K., Gojobori, T., Ikeo, K., 2013. The First Symbiont-Free Genome Sequence of Marine Red Alga, Susabi-nori (Pyropia yezoensis). PLoS One 8, 1–12. doi:10.1371/journal.pone.0057122

Ohtake, S., Kimata, K., Habuchi, O., 2003. A unique nonreducing terminal modification of chondroitin sulfate by N-Acetylgalactosamine 4-sulfate 6-O-sulfotransferase. J. Biol. Chem. 278, 38443–38452. doi:10.1074/jbc.M306132200

Páez, U.A.H., Romero, I.A.G., Restrepo, S.R., Gutiérrez, F.A.A., Castaño, D.M., 2015. Assembly and analysis of differential transcriptome responses of Hevea brasiliensis on interaction with Microcyclus ulei. PLoS One 10, 1–21. doi:10.1371/journal.pone.0134837

Park, I.Y., Park, C.B., Kim, M.S., Kim, S.C., 1998. Parasin I , an antimicrobial peptide derived from histone H2A in the cat fish , Parasilurus asotus. FEBS Lett. 437, 258–262.

Paul, P., Suwan, J., Liu, J., Dordick, J.S., Linhardt, R.J., 2012. Recent advances in sulfotransferase enzyme activity assays. Anal. Bioanal. Chem. 403, 1491–1500. doi:10.1007/s00216-012-5944-4

Pi, N., Hoang, M.B., Gao, H., Mougous, J.D., Bertozzi, C.R., Leary, J.A., 2005. Kinetic measurements and mechanism determination of Stf0 sulfotransferase using mass spectrometry. Anal. Biochem. 341, 94–104. doi:10.1016/j.ab.2005.02.004

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., Matsunaga, S., 2004. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei. Proc. Natl. Acad. Sci. U. S. A. 101, 16222–16227. doi:10.1073/pnas.0405976101

Plaza, M., Cifuentes, A., Ibanez, E., 2008. In the search of new functional food ingredients from algae. Trends Food Sci. Technol. 19, 31–39. doi:10.1016/j.tifs.2007.07.012

Pomin, V.H., Mourão, P.A.S., 2008. Structure, biology, evolution, and medical importance of sulfated fucans and galactans. Glycobiology 18, 1016–1027. doi:10.1093/glycob/cwn085

Prather, B., Ethen, C.M., Machacek, M., R, Z.L.W., Systems, D., Ne, M.P., 2012. Phosphatase-Coupled Sulfotransferase Assay [WWW Document]. URL https://www.rndsystems.com/resources/posters/phosphatase-coupled-sulfotransferase-assay (accessed 9.19.16)

Provost, E., Weier, C.A., Leach, S.D., 2013. Multiple ribosomal proteins are expressed at high levels in developing zebrafish endoderm and are required for normal exocrine pancreas development. Zebrafish 10, 161–9. doi:10.1089/zeb.2013.0884

Qiagen, 2001. RNeasy Mini Kit Third Edition [WWW Document]. URL https://www.qiagen.com/za/resources/resourcedetail?id=14e7cf6e-521a-4cf7-8cbc-bf9f6fa33e24&lang=en (accessed 10.9.16)

QIAGEN Aarhus A/S, 2015. CLC Genomics Workbench [WWW Document]. URL http://resources.qiagenbioinformatics.com/manuals/clcmainworkbench/current/User_Manual.pdf (accessed 10.9.16)

Rangel-Huerta, O.D., Pastor-Villaescusa, B., Aguilera, C.M., Gil, A., 2015. A systematic review of the efficacy of bioactive compounds in cardiovascular disease: Phenolic compounds, Nutrients. doi:10.3390/nu7075177

Rhein-Knudsen, N., Ale, M., Meyer, A., 2015. Seaweed Hydrocolloid Production: An Update on Enzyme Assisted Extraction and Modification Technologies. Mar. Drugs 13, 3340–3359. doi:10.3390/md13063340

Robles, J. a, Qureshi, S.E., Stephen, S.J., Wilson, S.R., Burden, C.J., Taylor, J.M., Robles, A., Qureshi, S.E., Stephen, S.J., Wilson, S.R., Burden, C.J., Taylor, J.M., 2012. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. BMC Genomics 13,

484. doi:10.1186/1471-2164-13-484

Sakakibara, Y., Suiko, M., Pai, T.G., Nakayama, T., Takami, Y., Katafuchi, J., Liu, M.-C., 2002. Highly conserved mouse and human brain sulfotransferases: molecular cloning, expression, and functional characterization. Gene 285, 39–47. doi:10.1016/S0378-1119(02)00431-6

Sambrook, J., Russel, D., 1989. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Schaeffer, D.J., Krylov, V.S., 2000. Anti-HIV activity of extracts and compounds from algae and cyanobacteria. Ecotoxicol. Environ. Saf. 45, 208–227. doi:10.1006/eesa.1999.1862

Schieber, A., 2012. Functional Foods and Nutraceuticals. Food Res. Int., Food Science Text Series 46, 437. doi:10.1016/j.foodres.2012.02.009

Schuurman, N., Leszczynski, A., 2008. Ontologies for bioinformatics. Bioinform. Biol. Insights 2, 187–200. doi:doi: 10.1007/978-3-642-30574-0_27

Sidharthan, N.P., Butcher, N.J., Mitchell, D.J., Minchin, R.F., 2014. Expression of the orphan cytosolic sulfotransferase SULT4A1 and its major splice variant in human tissues and cells: Dimerization, degradation and polyubiquitination. PLoS One 9, 1–8. doi:10.1371/journal.pone.0101520

Silbert, J.E., Sugumaran, G., 2002. Biosynthesis of Chondroitin /Dermatan Sulfate. IUBMB Life 54, 177–186. doi:10.1080/15216540290114450

Spellberg, B., Powers, J.H., Brass, E.P., Miller, L.G., Edwards, J.E., 2004. Trends in Antimicrobial Drug Development: Implications for the Future. Clin. Infect. Dis. 38, 1279–1286. doi:10.1086/420937

Stead, P., Hiscox, S., Robinson, P.S., Pike, N.B., Sidebottom, P.J., Roberts, A.D., Taylor, N.L., Wright, A.E., Pomponi, S.A., Langley, D., 2000. Eryloside F, a novel penasterol disaccharide possessing potent thrombin receptor antagonist activity. Bioorganic Med. Chem. Lett. 10, 661–664. doi:10.1016/S0960-894X(00)00063-9

Su, Y., 2011. Isolation and identification of pelteobagrin, a novel antimicrobial peptide from the skin mucus of yellow catfish (*Pelteobagrus fulvidraco*). Comp. Biochem. Physiol. - B Biochem. Mol. Biol. 158, 149–154. doi:10.1016/j.cbpb.2010.11.002

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. 102, 15545–15550. doi:10.1073/pnas.0506580102

Sudheesh, S., Sawbridge, T.I., Cogan, N.O., Kennedy, P., Forster, J.W., Kaur, S., 2015. De novo assembly and characterisation of the field pea transcriptome using RNA-Seq. BMC Genomics 16, 611. doi:10.1186/s12864-015-1815-7

Sugahara, T., Yang, Y., Liu, C., Pai, T.., Liu, M., 2003. Sulphonation of dehydroepiandrosterone and neurosteroids: molecular cloning, expression, and functional characterization of a novel zebrafish SULT2 cytosolic sulphotransferase. Biochem. J. 375, 785–91. doi:10.1042/BJ20031050

Talero, E., García-Mauriño, S., Ávila-Román, J., Rodríguez-Luna, A., Alcaide, A., Motilva, V., 2015. Bioactive compounds isolated from microalgae in chronic inflammation and cancer. Mar. Drugs 13, 6152–6209. doi:10.3390/md13106152

Tobergte, D.R., Curtis, S., 2013. The Dictionary of CELL AND MOLECULAR BIOLOGY, Journal of Chemical Information and Modeling. doi:10.1017/CBO9781107415324.004

Tsoi, C., Falany, C.N., Morgenstern, R., Swedmark, S., 2001. Molecular Cloning, Expression, and Characterization of a Canine Sulfotransferase That Is a Human ST1B2 Ortholog. Arch. Biochem. Biophys. 390, 87–92. doi:10.1006/abbi.2001.2373

UniProt Consortium, 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43, D204-12. doi:10.1093/nar/gku989

van der Kamp, J.W., Asp, N.-G., Miller Jones, J., Schaafsma, G., 2004. Dietary fibre, bio-active carbohydrates for food and feed. Wageningen Academic Pub.

doi:10.3920/978-90-8686-662-5

Varin, L., Chamberland, H., Lafontaine, J.G., Richard, M., 1997. The enzyme involved in sulfation of the turgorin, gallic acid 4-O-(beta-d-glucopyranosyl-6'-sulfate) is pulvini-localized in *Mimosa pudica*. Plant J. 12, 831–837. doi:10.1046/j.1365-313X.1997.12040831.x

Villanueva, R.D., Sousa, a. M.M., Gonçalves, M.P., Nilsson, M., Hilliou, L., 2010. Production and properties of agar from the invasive marine alga, Gracilaria vermiculophylla (Gracilariales, Rhodophyta). J. Appl. Phycol. 22, 211–220. doi:10.1007/s10811-009-9444-7

Viola, R., Nyvall, P., Pedersen, M., 2001. The unique features of starch metabolism in red algae. Proc. R. Soc. B Biol. Sci. 268, 1417–1422. doi:10.1098/rspb.2001.1644

Volpi, N., Maccari, F., 2002. Detection of submicrogram quantities of glycosaminoglycans on agarose gels by sequential staining with toluidine blue and Stains-All. Electrophoresis 23, 4060–4066. doi:10.1002/elps.200290021

Wagner, H and Kraus, S., 2000. Bioactive Carbohydrate Polymers. Springer Netherlands, Dordrecht. doi:10.1007/978-94-015-9572-8

Wall, R., Ross, R.P., Fitzgerald, G.F., Stanton, C., 2010. Fatty acids from fish: The anti-inflammatory potential of long-chain omega-3 fatty acids. Nutr. Rev. 68, 280–289. doi:10.1111/j.1753-4887.2010.00287.x

Wang, J., Liu, L., Zhang, Q., Zhang, Z., Qi, H., Li, P., 2009. Synthesized oversulphated, acetylated and benzoylated derivatives of fucoidan extracted from Laminaria japonica and their potential antioxidant activity in vitro. Food Chem. 114, 1285–1290. doi:10.1016/j.foodchem.2008.10.082

Wang, J., Zhang, Q., Zhang, Z., Song, H., Li, P., 2010. Potential antioxidant and anticoagulant capacity of low molecular weight fucoidan fractions extracted from Laminaria japonica. Int. J. Biol. Macromol. 46, 6–12. doi:10.1016/j.ijbiomac.2009.10.015

Wang, J.X., Breaker, R.R., 2008. Riboswitches that sense S -adenosylmethionine and S -adenosylhomocysteineThis paper is one of a selection of papers

published in this Special Issue, entitled CSBMCB — Systems and Chemical Biology, and has undergone the Journal's usual peer review process. Biochem. Cell Biol. 86, 157–168. doi:10.1139/O08-008

Wijesekara, I., Pangestuti, R., Kim, S.K., 2011. Biological activities and potential health benefits of sulfated polysaccharides derived from marine algae. Carbohydr. Polym. 84, 14–21. doi:10.1016/j.carbpol.2010.10.062

Williams, C.R., Baccarella, A., Parrish, J.Z., Kim, C.C., 2016. Trimming of sequence reads alters RNA-Seq gene expression estimates. BMC Bioinformatics 17, 103. doi:10.1186/s12859-016-0956-2

Wu, S., Sun, J., Chi, S., Wang, L., Wang, X., Liu, C., Li, X., Yin, J., Liu, T., Yu, J., 2014. Transcriptome sequencing of essential marine brown and red algal species in China and its significance in algal biology and phylogeny. Acta Oceanol. Sin. 33, 1–12. doi:10.1007/s13131-014-0435-4

Wu, Z.L., Ethen, C.M., Larson, S., Prather, B., Jiang, W., 2010. A versatile polyacrylamide gel electrophoresis based sulfotransferase assay. BMC Biotechnol. 10, 11. doi:10.1186/1472-6750-10-11

Wynne, M.J., 2002. A description of Plocamium fimbriatum sp. nov. (Plocamiales, Rhodophyta) from the Sultanate of Oman, with a census of currently recognized species in the genus. Nov. Hedwigia 75, 333–356. doi:10.1127/0029-5035/2002/0075-0333

Xia, W., Liu, P., Zhang, J., Chen, J., 2010. Biological activities of chitosan and chitooligosaccharides. Food Hydrocoll. 25, 170–179. doi:10.1016/j.foodhyd.2010.03.003

Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T.W., Li, Y., Xu, X., Wong, G.K.S., Wang, J., 2014. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30, 1660–1666. doi:10.1093/bioinformatics/btu077

Xiong, J., Wang, G., Cheng, J., Tian, M., Pan, X., Warren, A., Jiang, C., Yuan, D., Miao, W., 2015. Genome of the facultative scuticociliatosis pathogen Pseudocohnilembus persalinus provides insight into its virulence through

horizontal gene transfer. Sci. Rep. 5, 15470. doi:10.1038/srep15470

Yabuta, Y., Fujimura, H., Kwak, C.S., Enomoto, T., Watanabe, F., 2010. Antioxidant Activity of the Phycoerythrobilin Compound Formed from a Dried Korean Purple Laver (Porphyra sp.) during in Vitro Digestion. Food Sci. Technol. Res. 16, 347–352. doi:10.3136/fstr.16.347

Yamaguchi, T., Ohtake, S., Kimata, K., Habuchi, O., 2007. Molecular cloning of squid N-acetylgalactosamine 4-sulfate 6-O-sulfotransferase and synthesis of a unique chondroitin sulfate containing E-D Hybrid tetrasaccharide structure by the recombinant enzyme. Glycobiology 17, 1365–1376. doi:10.1093/glycob/cwm103

Yasuda, S., Kumar, A.P., Liu, M.Y., Sakakibara, Y., Suiko, M., Chen, L., Liu, M.C., 2005. Identification of a novel thyroid hormone-sulfating cytosolic sulfotransferase, SULT1 ST5, from zebrafish: Molecular cloning, expression, characterization and ontogenic study. FEBS J. 272, 3828–3837. doi:10.1111/j.1742-4658.2005.04791.x

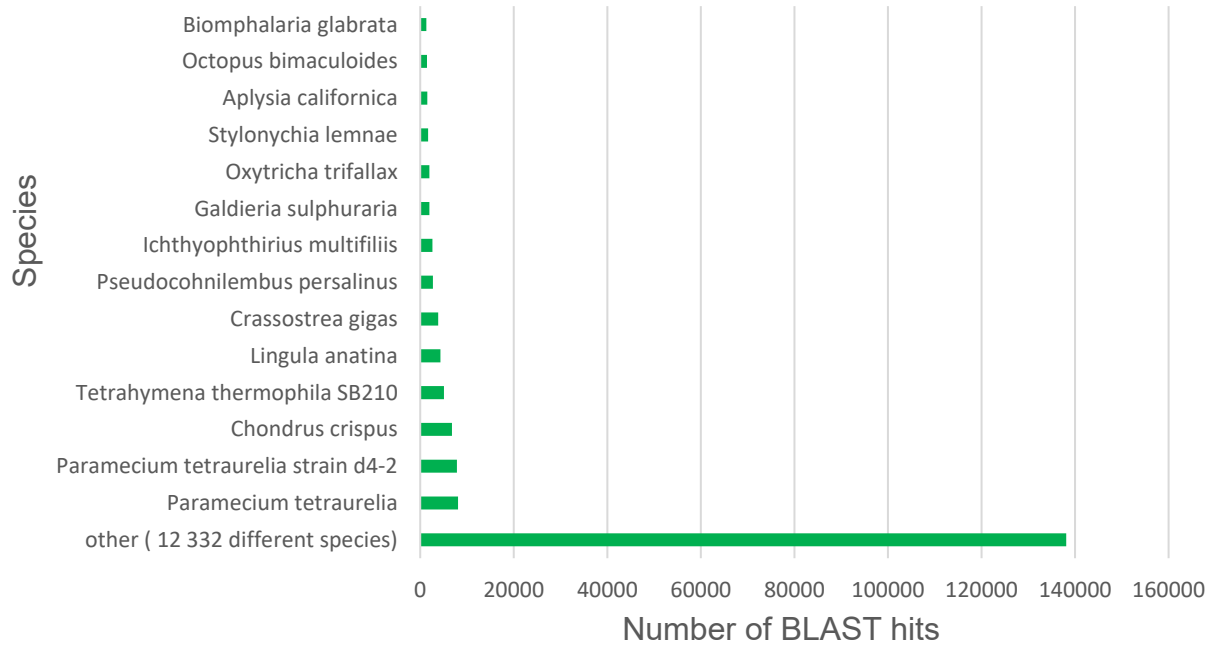
Zhang, Z., Pang, T., Li, Q., Zhang, L., Li, L., Liu, J., 2015. Transcriptome sequencing and characterization for *Kappaphycus alvarezii*. Eur. J. Phycol. 50, 400–407. doi:10.1080/09670262.2015.1069403

Zhang, Z., Zhang, Q., Wang, J., Song, H., Zhang, H., Niu, X., 2010. Regioselective syntheses of sulfated porphyrans from Porphyra haitanensis and their antioxidant and anticoagulant activities in vitro. Carbohydr. Polym. 79, 1124–1129. doi:10.1016/j.carbpol.2009.10.055

Appendix A

Species distribution for annotations derived through BLASTx sequence homology of contigs to the NCBI nr database is summarised in Figure 20. A and B. The top 10 general species distribution of all the BLASTx hits used for annotation inferences are exclusively eukaryotes (Figure 20.A). A species overlap can be observed between Figure 20. A and B, with *Chondrus crispus* (6784 hits), *Tetrahymena thermophila* SB210 (5062 hits), *Paramecium tetraurelia* (8072 hits) and *Paramecium tetraurelia* strain d4-2 (7847 its). The top hit BLASTx results of 3099 were from *Chondrus crispus*, commonly termed Irish moss, a macro red alga species. Additionally, other species with the best BLASTx homology alignments were *Tetrahymena thermophila* SB210 (869 hits), *Paramecium tetraurelia* (685 hits), *Lingula anatine* (670 hits) and *Paramecium tetraurelia* strain d4-2 (668 hits). The top 10 best BLASTx homology alignments, were all eukaryotes such as *Crassostrea gigas* (the Pacific oyster) and *Galdieria sulfuraria* (a red alga) with approximately 50% present in either marine or freshwater environments (Figure 20.B).

(A)



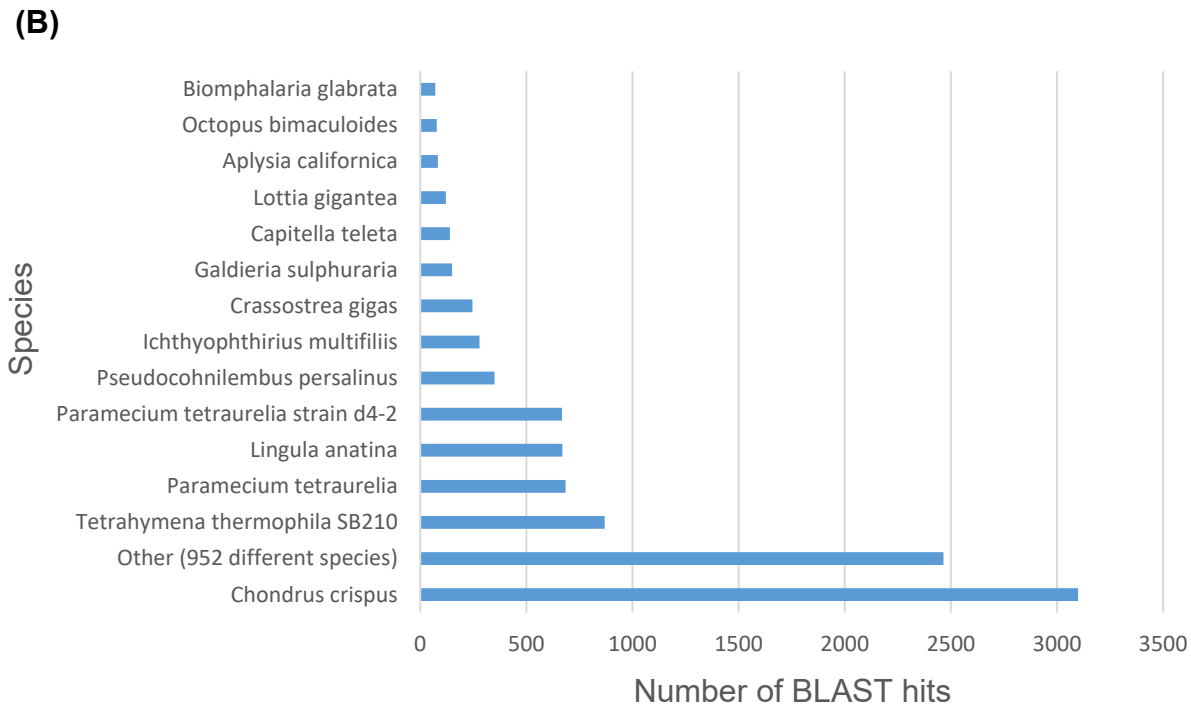


Figure 20. BLASTx hits with regard to species distribution for assembled contigs ≥ 900. (A) Total species distribution derived from BLASTx alignments **(B)** Top-hit species distribution for all BLASTx hits used for annotation inferences.

The GC-content of the assembled contigs (119 675 contigs) calculated as the number of GC bases compared to the total bases, including ambiguous nucleotides (Figure 21).

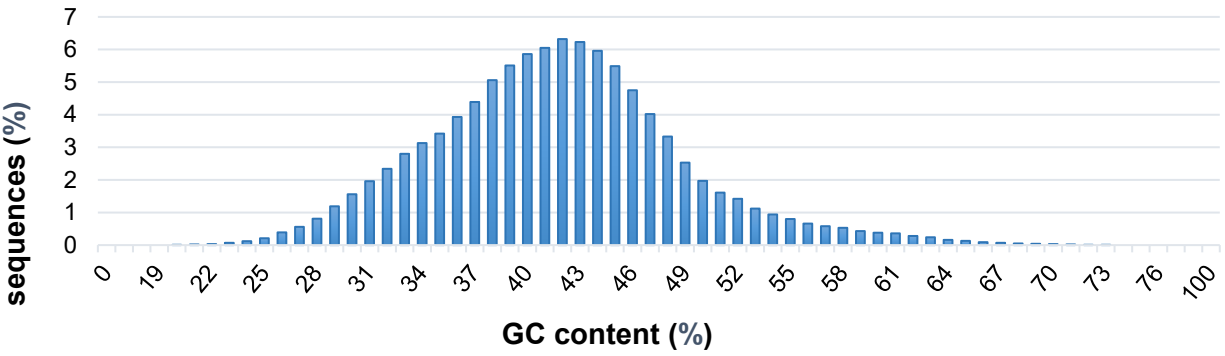


Figure 21. GC-content distribution for the P. corallorhiza transcriptome. The X-axis represents the relative GC-content in percentages of a contig. The Y-axis represents the number of contigs containing particular GC- percentages normalised to the total number of contigs.

For each GC content level (0-100%), the mean read coverage of 100bp reference segments with that GC content is displayed. Three peaks can be observed for the lower GC content range i.e. 5% to 10%. The dominant peak at approximately 55% is further followed by at least 10 other sequence peaks in the upper GC content range i.e. 60% to 85% (Figure 22)

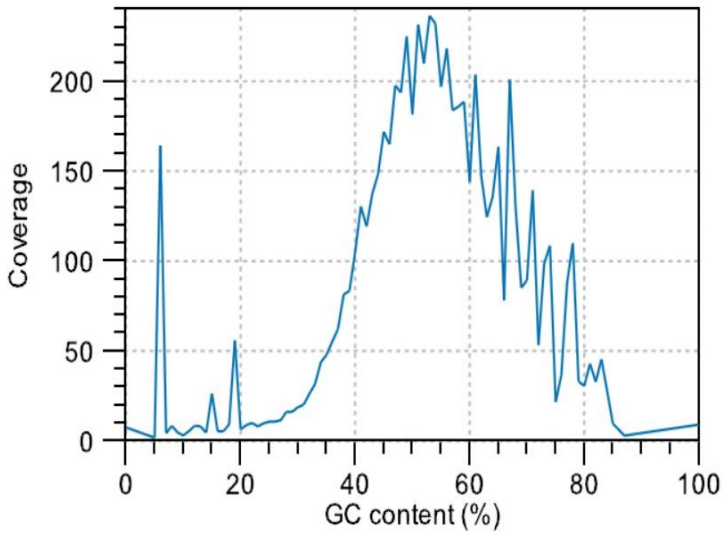


Figure 22. Coverage over GC content analysis of the *P. corallorhiza* draft transcriptome.

Annotation was retrieved for 54.66% of the 11 816 contigs subjected to the in silico annotation procedure. GO terms assigned to the annotated contigs are divided into molecular function (MF), cellular component (CC) and biological processes (BP) (Figure 23). Jointly, these GO terms provide a broader perspective on the functional consequences of gene expression. Figure 23, illustrates that heterocyclic and organic cyclic compound binding are the MFs represented by the greatest number contigs. Other MF included hydrolase, transferase, small molecule binding and protein binding activities. The transferase activity which houses enzymatic functions including the transfer of methyl group, glycosyl group and sulfuryl group, is represented 1228 contig sequences.

Cellular components, encompassing any part of a cell or its extracellular location where a gene product is found, is represented by 3 554 contigs. The CCs with the highest number of contigs are intracellular parts, which is part of the GO term intracellular cell components. Intracellular components include the nucleus and cytoplasm, however, it excludes structures such as the plasma membrane, large vacuoles, secretory masses and indigested material. Further, other CC components found in the GO terms assigned to the *P. corallorhiza* draft transcriptome include non-membrane and membrane-bound organelles implied by 550 and 1312 contigs, respectively.

Lastly, GO terms relating to biological process encompass primary metabolic process, organic substance metabolic process and cellular metabolic process. Interestingly, the response to stress BP is also very highly represented by 448 contigs. Other BP present includes, nitrogen compound metabolic process, cellular component process, biosynthetic process and catabolic process.

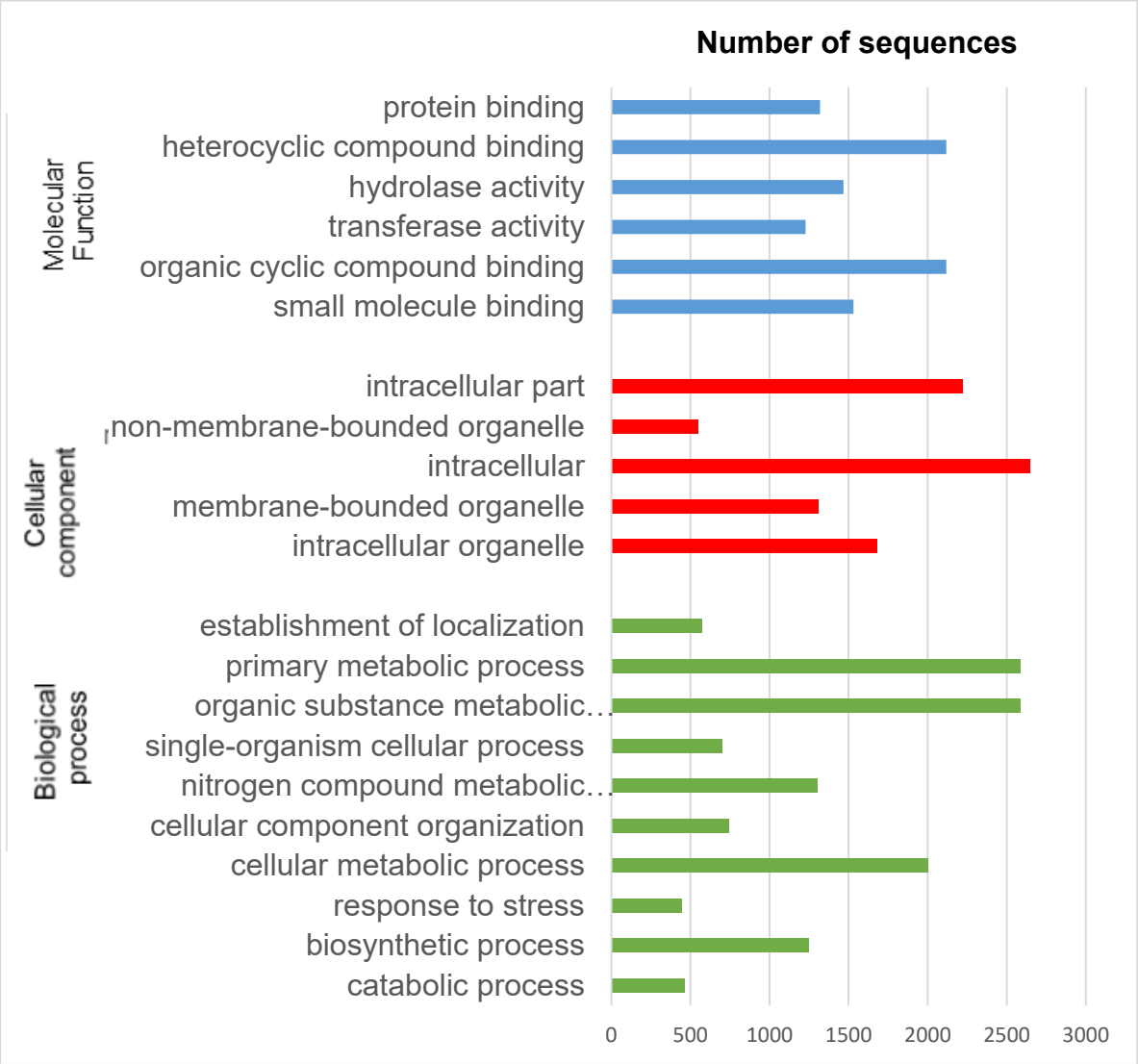


Figure 23. GO term annotation distribution from BLAST2GO pipeline for P. corallorhiza transcriptome. GO terms relating to biological processes, molecular functions and cellular components were assigned to the *P. corallorhiza* transcriptome using the BLAST2GO pipeline, including BLASTx and InterProScan analysis.

Diurnal differential gene expression analysis was executed using the Wald statistical analysis with log2 fold change ≥ 5, signifying nocturnally upregulated genes, or ≤ -5, signifying nocturnally downregulated genes, combined with the Bonferroni and False Discovery Rates (FDR) corrected p-values ≤ 0.01 (Chen et al., 2011). The analysis

produced 639 differentially expressed genes (DEGs) of which 86 genes were nocturnally upregulated and 553 genes nocturnally downregulated. Figure 24, visually represents the differential gene expression analysis, using volcano plots, with red dots indication the DEGs.

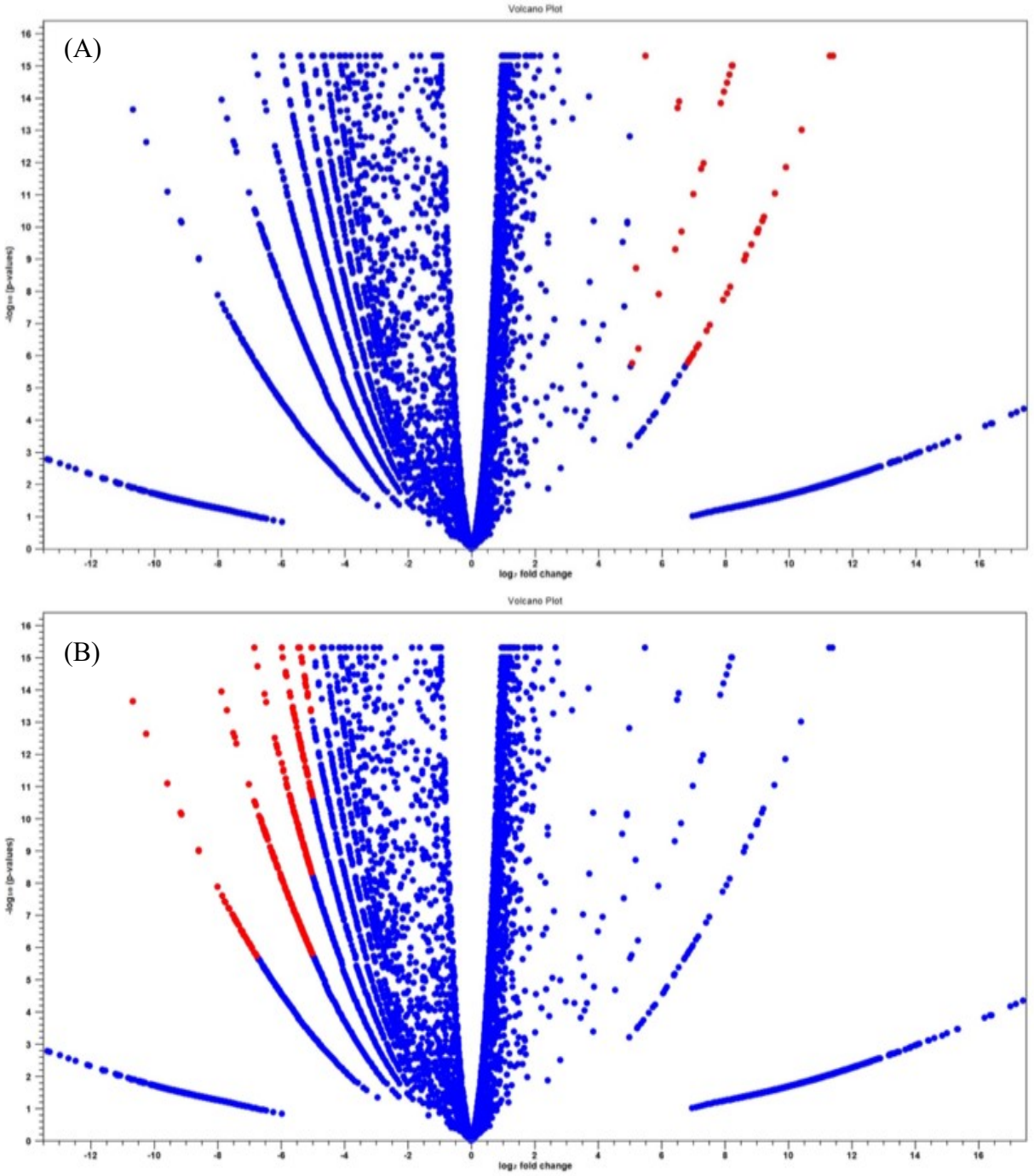


Figure 24. Volcano plot visually presenting the DEGs for the night compared to day RNA-seq sample. DEGs (red dots) were identified through Wald statistical analysis (log2 Fold change ≥ 5 or ≤ -5, with the Bonferroni and False Discovery Rates (FDR), corrected p-values ≤ 0.01)*. **(A)** Upregulated transcripts (red dots) in the Night compared to the Day temporal sample. **(B)** Downregulated transcripts in the Night, when compared to the day temporal sample

*corrected p-values not statistically significant, because of a lack of biological replicates for the RNA-seq analysis.