The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES IN THE COMMERCIALLY IMPORTANT AGAROPHYTE, *GRACILARIA GRACILIS*, FOLLOWING NITROGEN DEPRIVATION

BY

Tanya Lebi
September 2006

University of Cape Town

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in the Department of Molecular and Cell Biology, Faculty of Science, University of Cape Town, South Africa.

Department of Molecular and Cell Biology, University of Cape Town,
Private Bag, Rondebosch, 7701, South Africa

TABLE OF CONTENTS

	Acknowledgements	**II**
	Abstract	**III**
	Abbreviations	**V**
CHAPTER 1	**Introduction**	**1**
CHAPTER 2	**Nitrogen deprivation: Acquisition of nitrogen deprived *G. gracilis* thalli for use in microarray**	**35**
CHAPTER 3	**Optimisation of microarray protocols: RNA isolation, cDNA synthesis, labeling and hybridisation**	**50**
CHAPTER 4	**Microarray hybridisation**	**79**
CHAPTER 5	**Identification of differentially expressed cDNA targets: sequencing and bioinformatics**	**117**
CHAPTER 6	**Conclusions and future work**	**136**
APPENDIX A	**Media, buffers and solutions**	**143**
APPENDIX B	**DNA primer sequences and PCR cycle profiles**	**151**
APPENDIX C	**Standard methods**	**153**
APPENDIX D	**pDNR-LIB Cloning vector**	**156**
APPENDIX E	**R interface programming language**	**157**
APPENDIX F	**Sequences**	**159**
Literature Cited		**163**

University of Cape Town

ACKNOWLEDGEMENTS

I would like to thank my supervisor Associate Professor Vernon Coyne, for his invaluable supervision, advice and guidance. In addition to this, I would like to thank him for the financial assistance and for proof-reading my thesis.

Thank you to the National Research Foundation for scholarship support.

Thank you to Wiessner Vos from the UCT, National Bioinformatics Network (NBN) for his help with statistical analysis of the microarray data. Thank you also to Sally Ann Walford, Tsion Tabraha, and Rachel Van Dyk from CAPAR for all their assistance, and Heather Sessions from the Marine and Coastal management for her help with CHNS analysis.

Thank you to all my colleagues in lab 201, for your constant encouragement, support, advice and most of all friendship. It has been a difficult road to travel, and there is no doubt in my mind that my journey would have been impossible without the help of such an amazing unit of people encouraging me along the way. May you all continue to succeed in your careers and lives. Thank you!

To my amazing family: Mom, Dad and Nick, your love and support is invaluable. You have always supported me and allowed me to pursue all my goals in life, and for this I am truly grateful. To my grandparents, granny Audrey, grandpa Arnold and my granny Gabriella your encouragement and love is never ending; I love you all very much.

To all my friends: When John Lennon wrote 'I get by with a little help from my friends', he obviously hadn't met you guys. You have all been an amazing support system, Thank you!

To my Mike, your belief in my capabilities, your constant encouragement and your love has not only been a huge help throughout my studies, but in addition, has helped me develop a strong sense of all my strengths, capabilities and what I want most in life. You are my rock! I love you.

University of Cape Town

ABSTRACT

The red agarophytic alga, *Gracilaria gracilis*, occurs naturally within Saldahna Bay, South Africa. *Gracilaria* species are commercially exploited for their hydrocolloid agar, valued at US$132 million per annum (FAO, 2002). *G. gracilis* is thus a valuable resource for South Africa in terms of generating foreign currency through export. Nitrogen limitation occurring within Saldahna Bay during the summer to autumn months, however, is considered to be the major factor preventing commercial cultivation of *G. gracilis*. Algae possess various survival mechanisms that generally involve the activation or repression of various gene regulatory systems. Such regulatory systems aid in preserving energetic homeostasis at a cellular level, thereby allowing maximal algal survival during adverse environmental conditions. Microarray technology is a high-throughput global gene analysis tool, allowing for the simultaneous identification of thousands of differentially expressed genes in a single assay. Combined with genomic technologies such as genetic engineering, microarray technology has the potential to identify genes that confer enhanced *in vivo* tolerance to a specific stress. In the current investigation, microarray technology was utilized to identify genes differentially expressed in *G. gracilis* in response to nitrogen limitation. A total of 39 differentially expressed genes were identified. Of the 35 genes that were sequenced, 13 were assigned a putative function based on sequence similarity to genes deposited in various databases. The remaining sequences corresponded to either unknown hypothetical proteins or else had no significant homology to any of the sequences in the databases. Overall, the current investigation has provided a foundation for future research into the biological role(s) of

University of Cape Town

each of the identified differentially expressed genes within their true isogenic background. Future studies may thus allow for improved crop tolerance and stability, enhancing aquaculture of this valuable resource in Southern Africa.

University of Cape Town

Abbreviations

α	alpha
β	beta
λ	lambda
γ	gamma
μ	relative growth rate
μg	microgram(s)
μl	microlitre(s)
μm	micromole(s)
μmol	micromolar
ASW	artificial sea water
Asp(AT)	aspartate aminotransferase
ATP	adenosine triphosphate
ABC_ATPase	adenosine triphosphate binding domain_adenosine triphosphatase
aRNA	antisense RNA
BLAST	basic local alignment search tool
BLASTn	basic local alignment search tool nucleotide
BLASTx	basic local alignment search tool nucleotide translated
BSA	Bovine serum albumin
bp	base pair(s)
C	carbon
CAPAR	Cape Array Opportunities
CHD(s)	Chromodomain helicase deoxyribonucleic acid-binding proteins
CHNS	carbon, hydrogen, nitrogen, sulphur
cDNA	complementary DNA
Cm	Chloramphenicol
COT1-DNA	cobalt transport deoxyribonucleic acid
CTAB	hexacetyltrimethyl ammonium bromide
CY-3/5	cyanine (fluorescent dye label)
DEPC	diethyl pyrocarbonate
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxyribonucleoside triphosphates (dATP, dCTP, dTTP and GTP)
dUTP	deoxyuridine triphosphate
EBI	European bioinformatics institute
EDTA	ethylenediaminetetraacetic acid (disodium salt)
EF-3	elongation factor 3
EST	expressed sequence tag

University of Cape Town

g	standard gravitational acceleration (9.81 m/s^2)
GeO_2	Germanium Oxide
GSH	glutathione
GT	glutamyltransferase
HLIP	high light inducible protein
hr(s)	hour(s)
IVT	invitro transcription
kDa	kilodalton(s)
kb	kilobase(s)
l	litre(s)
LB	Luria broth
$m^{-2}s^{-1}$	per square meter per second
MA	marine agar
MB	marine broth
MCS	multiple cloning site
mg	milligram(s)
min	minute(s)
mJ	milli joules
ml	millilitre(s)
MOPS	(3-[N-morpholino]propane-sulfonic acid)
mRNA	messenger RNA
MW	molecular weight
N	nitrogen
NCBI	national centre for biotechnology information
ng	nanogram(s)
OD	optical density
OD_{600}	optical density at 600 nm
O/N	overnight
ORF	open reading frame
PCR	polymerase chain reaction
PMT	photo multiplier tube
PS	photosystem
PVPP	Polyvinylpolypyrrolidinone
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal ribonucleic acid

University of Cape Town

RPEase	ribulose-phosphate 3-epimerase
RT-PCR	real time-polymerase chain reaction
s	second(s)
SDS	sodium dodecyl sulphate (sodium lauryl sulphate)
SSC	sodium chloride tri-sodium citrate buffer
TAE	Tris-acetate-EDTA buffer
TE	Tris-EDTA buffer
Tris	tris(hydroxymethyl)aminomethane
tRNA	transport ribonucleic acid
U	unit(s)
UV	ultraviolet
V	volts
wt	weight

University of Cape Town

CHAPTER 1

INTRODUCTION

1.1 AQUACULTURE 2

1.2 SEAWEED AQUACULTURE 3

1.3 THE GENUS *GRACILARIA* 6

1.4 *GRACILARIA* CULTIVATION 7

1.5 ENVIRONMENTAL FACTORS LIMITING CULTIVATION 10

1.6 ESTABLISHMENT OF *GRACILARIA* FARMING IN SOUTH AFRICA 13

1.7 NITROGEN ASSIMILATION IN PLANTS AND ALGAE 18

1.8 GENERAL 'STRESS' RESPONSE OF PLANTS AND ALGAE TO UNFAVOURABLE CONDITIONS 21

1.9 GENETIC RESPONSES TO NITROGEN LIMITATION IN ALGAE AND HIGHER PLANTS 24

1.10 MICROARRAY TECHNOLOGY 28

1.11 AIMS OF THIS STUDY 33

University of Cape Town

CHAPTER 1

INTRODUCTION

1.1 AQUACULTURE

Utilization of the ocean's vast resources by humans for food and other essential commodities almost certainly dates back to prehistoric times. The marine environment itself is an enormous source of biodiversity, with over 80 % of all life believed to exist beneath its surfaces (Patrzykat *et al.*, 2003). Thus, today like our primordial ancestors, we continue to use the ocean as a central means of survival; however our ability to manipulate this rich resource is far from primitive.

With greater understanding of the biology of various species, commercial exploitation has advanced, giving rise to the establishment of numerous industrial corporations, each manufacturing a diverse range of products. Consequently, with the expansion of such industries, the demand for these 'raw materials' often exceeds the available supply from natural 'stocks'. In general, such an occurrence tends to enhance the economic viability of farming a species. Thus aquaculture, defined as "the farming of aquatic organisms, including fish, molluscs, crustaceans, and aquatic plants" (Edwards and Demaine,1998) was developed as a means of meeting such commercial demands, while additionally counteracting the numerous environmental challenges that natural populations often encounter.

University of Cape Town

Today, aquaculture is a rapidly growing industry, providing a means for intensive seafood production under controllable conditions (Olafsen, 2001). Currently, aquaculture accounts for over 30% of global fish production, and is expected to exceed 50% by 2010 (FAO).

In South Africa total production from aquaculture has accumulated substantially, increasing from 3 000 tons ($51 million) in 1997 to 4 030 tons ($146 million) in 2000 (Aquaculture Association of Southern Africa http://www.aasa-aqua.co.za/). In general South African aquacultural practices predominantly focus on high-value niche – market species, with abalone and oyster production comprising 76% of the total output while seaweed, mussels, finfish and prawn farming make up the remaining 24%. Thus while aquaculture serves as a valuable resource for earning foreign currency through export, it additionally aids in enhancing the livelihood of the country as a whole by both directly and indirectly creating jobs for numerous people of different socio-economic backgrounds. Furthermore, aquaculture provides a means of supporting and supplementing local diets, while additionally enhancing conservation of natural populations.

University of Cape Town

1.2 SEAWEED AQUACULTURE

Ancient records reveal that the collection of seaweed by people as a food and health source can be recorded as far back as 500 B.C. in China (Lobban *et al.*, 1994). As part of the human diet, seaweeds are highly nutritious, providing protein, vitamins, and essential minerals such as iodine and iron. Today, despite the large number of commercial

applications for seaweeds (see below), the principle use of seaweeds, particularly in the Far East (Japan, China and Korea) remains as a source of human food, with the annual consumption estimated to be 400 000 tonnes dry weight (Critchley, 1993).

In general, seaweeds can be broadly classified into three main groups, namely the Rhodophyta (red), Phaeophyta (brown) and Chlorophyta (green). Of these the Chlorophyta serve solely as a source of food, while Rhodophyta and Phaeophyta species, although contributing substantially to the commercial food industry, are additionally exploited for their phycocolloids: Agar, Carrageenan and Alginate. At present the global market for these phycocolloids (long chain algal polysaccharides) is estimated to be worth 585 million US$ annually (McHugh, 2003)

Alginate is extracted from brown seaweeds and is generally used as a gelling, emulsifying and stabilizing agent in the food, paper, textile, welding and pharmaceutical industries (Anderson *et al.*, 1989). The remaining two phycocolloids, agar and carrageenan, are derived from the Rhodophyta and serve mainly as stabilizing and gelling agents. Carrageenan is used widely in confectionery, ice cream, meat products and sauces as well as in the cosmetic and silk industries, while agar is predominantly used as the principal component of bacteriological culture bases as well as a gelling agent in foods such as jellies, mayonnaise, processed cheese and sweets. In addition agar has been used in the pharmaceutical industry as an inert carrier for drug products where slow release of the drug is required (Lee, 1999b).

University of Cape Town

Presently, growth of the agar market has been considerable, increasing at a historical rate of 2 % per year, forming the basis of a multimillion-dollar industry (Oliveira *et al.*, 2000), with the world market value of agar assessed to be at US$ 200 million (Critchley, 1993). Before 1942, the world's primary source of agar was derived from *Gelidium* species sp., due to *Gelidium*'s high gelling strength and low sulphation content. Over-harvesting and exploitation of this species at the beginning of the twentieth century resulted in increased interest in other agarophytic seaweeds as substitutes. Interest turned to agar extraction from fast growing *Gracilaria* sp., however the pre-treatment and extraction techniques then employed did not result in products similar to the classical agar extracted from *Gelidium*. Agar extracted from *Gracilaria* was described as an 'agaroid' having a higher gelification temperature due to methoxylation and a larger sulphate content to that of the 'natural agar' extracted from *Gelidium* (Armisen, 1995). It was only after the discovery of alkaline hydrolysis of sulphates by treatment of *Gracilaria* with sodium hydroxide that the transformation of these 'agaroids' into high quality agar was ascertained.

University of Cape Town

Today, the principal seaweed sources of agar are *Gracilaria* (53 %) and *Gelidium* (44 %), with minor amounts from other agarophytes such as *Gelidiella* and *Pterocladia* (Critchley, 1993). Of these, the main species of economic importance, based on agar yield and quality, are considered to be *Gracilaria chilensis* and *Gracilaria gracilis* (Oliveira *et al.*, 2000). In South Africa it is the latter species that has been and is exploited for its high quality agar.

As with land based plants, successful cultivation of a species requires extensive knowledge of the plant's biology, biochemistry and physiology. Thus *Gracilaria* utilization, cultivation and productivity is constantly advanced through scientific analysis of the morphology, development, biochemistry, physiology, genetic, and phylogenetic characteristics of this seaweed.

1.3 THE GENUS *GRACILARIA*

In general, gracilarioid algae, or *Gracilaria*, are classified as red algae belonging to the phylum Rhodophyta. The algae in this phylum are not necessarily red in colour, and species may manifest in black, yellow, green or red forms. Thalli on the whole are either bushy and rigid with relatively short branches (*Gracilaria beckeri*), or slender with ramifying, stringy branches (*Gracilaria gracilis*) (Fig 1.1) (Branch and Branch, 1981). Taxonomically the genus *Gracilaria* is further classified within the class Florideophyceae, order Gracilariales and family Gracilariaceae.

Thus classification of *Gracilaria gracilis* is as follows:

- Phylum: Rhodophyta
- Class: Florideophyceae
- Order: Gracilariales
- Family: Gracilariaceae
- Genus: *Gracilaria*
- Species: *G. gracilis*

Fig 1.1: *Gracilaria gracilis* thallus fragment

University of Cape Town

On the whole discrimination of *Gracilaria* species into taxonomic categories is hampered by morphologies that are both conservative (shared by a number of species) and variable (subject to considerable alteration by external factors) (Bird, 1995). Today traditional taxonomic approaches are supplemented with studies on hybridisation, chromosome number, characterisation of cell wall polysaccharides as well as various molecular techniques such as 18S rRNA sequencing or Rubisco sequencing, in order to better classify species within their correct taxonomic categories. De Oliviera and Plastino (1994) and Bird (1995) comprehensively discuss these approaches and expand in depth the challenges presently faced in Gracilariaceaen taxonomy.

University of Cape Town

1.4 *GRACILARIA* CULTIVATION

In its natural environment large amounts of *Gracilaria* tend to wash ashore, and in countries such as Chile, New Zealand, Malaysia, Thailand, the Philipines, Indonesia, China, Namibia and South Africa these populations have been, and currently are, exploited for industrial purposes (Santelices and Doty, 1989; Oliveira *et al.*, 2000). Harvesting of these natural populations may involve either passive collection of the seaweed in nets, or active collection by raking of the seafloor bed (Oliveira *et al.*, 2000). Typically, production values of 1-2 g (dry).$m^{-2}.d^{-1}$ have been recorded from these natural *Gracilaria* populations (Critchley, 1993).

Harvesting of natural populations for commercial exploitation is sporadic, and has been described as 'opportunistic' rather than managed (Santelices and Doty, 1989). In countries such as Chile, Taiwan, Brazil, the Philippines and Thailand, absence of

previous management and regulations, together with socio-economic pressure, has resulted in gradual loss and productivity from these natural beds (Santelices and Doty, 1989). In particular, insufficient knowledge regarding the implementation of the raking process employed to collect *Gracilaria* from seafloor beds has resulted in the disruption of underground thallal systems reducing overall productivity and future establishments (Santelices and Doty, 1998).

The importance of managing natural stocks has thus been realised and various harvesting strategies have been adopted. These strategies address questions such as how, when, and with what frequency to harvest, as well as the amount of material that should be left to allow for re-establishment of surviving underground thalli. A time-related harvesting strategy has been adopted in Chile with harvesting authorized every other month. This strategy, although changing the seasonal pattern of wild crop production, has resulted in the collection of a regular yield (Santelices and Doty, 1989). Ultimately, however, the unpredictability of both yield and quality of naturally collected *Gracilaria* prohibits the establishment of a successful agar industry and thus collection of natural crops cannot solely support the market demand for this phycocolloid.

University of Cape Town

The logical successor of wild crop management is artificial cultivation or mariculture. The rationale is that farming produces a crop that is more reliable in volume and quality and can be expanded as the market increases (Santelices and Doty, 1989). Furthermore, greater control of principle variables such as nutrient availability, salinity, pH, temperature and algal density can be achieved (Chapman and Chapman, 1980). Within

the seaweed industry Chile, China, and Taiwan have been forerunners in developing cultivation techniques for large scale *Gracilaria* cultivation (Armisen, 1995; Dawes, 1995). Today the main cultivation techniques used in *Gracilaria* mariculture are: cultivation in tanks (Edding *et al.*, 1987), raceways (Huguenin, 1976), reactors (Lapointe *et al.*, 1976), ponds (Friedlander and Levy, 1995) and the ocean (Dawes, 1995).

Cultivation, however, is limited by the amount of capital required in order to finance a continuous and successful harvest. Each cultivation technique is exceptionally intensive and energy expensive. In general, cultivation is only cost competitive if high-value end uses such as the production of drugs or delicacy foods is achieved, or if the price attainable for the cultivated *Gracilaria* is high. The economic feasibility of cultivation however has been improved through the initiation of a polyculture system whereby gracilarioids are co-cultured with other high-value species such as abalone, shrimp, crabs, molluscs, salmon and/or other fish (Troell *et al.*, 1997; Neori *et al.*, 2000; Rotmann, 1987; Chiang, 1981). In such polyculture systems *Gracilaria* is able to remove excreted ammonia and other toxic by-products from the water, producing a larger seaweed biomass while simultaneously serving as a food source for the co-cultured organism. In this way the overall costs of cultivating *Gracilaria* is covered by the operational costs for the primary farmed organism, making the whole system economically profitable and ecologically friendly. In addition, algae cultivated in this way have been shown to have a higher agar content (Martinez & Buschmann, 1996).

University of Cape Town

1.5 ENVIRONMENTAL FACTORS LIMITING CULTIVATION

As with land based agricultural systems, *Gracilaria* cultivation, whether within a natural or synthetic environment, is susceptible to unfavourable fluctuating conditions. Such conditions generally result in decreased algal growth, decreased agar quality or ultimately crop fatality. In terms of *Gracilaria* cultivation, the major biotic- (organic/living) and abiotic (inorganic/physical) factors considered to be responsible for variable *Gracilaria* yields are grazers, epiphytes, bacterial or fungal diseases, and nutrient availability.

Briefly, grazers are detrimental only when large numbers predominate or when growth and establishment of *Gracilaria* is slow. In 1993, Anderson *et al.* found that grazing of *G. gracilis* by fish in shallow waters, and keyhole limpets and urchins in deeper waters, prevented recovery of this resource in Saldanha Bay, South Africa. Additionally in laboratory based experiments, they determined that invertebrates present in Saldanha Bay in January 1990 were capable of eating more than 150 tons of *Gracilaria* in one month. It is thus necessary to monitor invertebrate levels near *Gracilaria* crops. Buschmann *et al.* (2001) suggest that rotation of farming areas could be a useful strategy in overcoming this problem.

University of Cape Town

The presence of invertebrate grazers can at times be favourable, as is the case with the isopod *Paridotea reticulata* found in association with commercially farmed *G. gracilis* in Saldanha Bay. During certain times of the year large populations of the red algal epiphyte *Ceramium diaphanum* flourish, lowering the yield and quality of the *G. gracilis* farmed in this area. When given the choice, *P. reticulata* will consume *C. diaphanum* rather than

G. gracilis. Thus during most months of the year *P. reticulata* successfully combats the spread of *C. diaphanum*, keeping the levels of this epiphyte under control (Anderson *et al*., 1998; Anderson *et al*., 1999). Similarly, small scale investigations for the use of the snail *Tegula atra* as a biological control agent have been considered in Chile (Buschmann *et al*., 2001).

In general epiphytes exert their negative effects on *Gracilaria* populations by increasing the mechanical drag on plants and acting as a competitor for available light and nutrients. In addition, many epiphytes are able to produce detrimental allelochemicals and/or have rhizoids that damage the host by penetrating into thalli (Oliveira *et al*., 2000). Anderson *et al*. (1996a) reported that the foliose green epiphyte *Ulva* could reduce the light reaching its host by 50 %.

Various physical, chemical and biological methods have been employed to control and reduce epiphyte infestations. The most common are: the introduction of fish or mesoherbivores into the cultivation system; use of agents such as copper chloride or sodium hypochlorite (which can be deleterious for both the epiphyte and *Gracilaria*); exposure to fresh water for 5-10 min before planting; and lastly, manipulation of the supply of essential nutrients such as nitrogen (Oliveira *et al*., 2000; Buschmann *et al*., 2001). Most of these methods are appropriate for tank cultivation, but are difficult to apply to open culture areas.

University of Cape Town

In 1997 Jaffray *et al.* isolated numerous bacterial epiphytes from Saldanha Bay *G. gracilis*. Under normal conditions these bacteria may provide the seaweed with dissolved inorganic nitrogen (DIN) (Bird and Benson, 1987) and may protect the seaweed from other pathogenic bacteria. Under unfavourable conditions however the symbiotic relationship between host and epiphyte can be altered with the result that the previously beneficial bacteria become pathogenic (Jaffray *et al.*, 1997).

The last principal environmental factor influencing *Gracilaria* growth is nutrient availability. All seaweeds require various macro- and micro-nutrients, as well as a range of trace elements in order to survive (Oliveira *et al.*, 2000). During certain periods of the year the availability of one or more of these nutrients in the aquatic environment is low enough to limit primary production (Anderson *et al.*, 1996b; Lignell and Pedersén, 1987). In general, nitrogen, phosphorous and iron appear to be the most critical limiting elements with nitrogen reported most frequently to limit the growth of seaweeds in their natural ecosystems (Hanisak, 1990; Bird *et al.*, 1987; Lignell and Pedersén, 1987). Conversely, excess availability of nutrients favours the growth of opportunistic organisms such as *Ulva* and *Ectocarpus* (Anderson *et al.*, 1996a; Oliveira *et al.*, 2000).

University of Cape Town

Physiologically *G. gracilis*, like many other *Gracilaria* sp., has the ability to store excess nitrogen from the environment as pigment proteins (Lignell and Pedersén, 1987; Bird *et al.*, 1982). Nitrogen uptake is thus uncoupled from the *G. gracilis* growth response revealing a 'strategy' of 'luxury consumption' (Smit, 2002). In addition to this, *Gracilaria* has the ability to absorb nutrients during the night

(Hanisak, 1990). These physiological attributes, together with the realisation that many microalgae and epiphytes are incapable of such tendencies, has allowed farmers to develop a "pulse-feeding" regime to minimize microalgal and epiphyte establishment under culture conditions (Lapointe, 1985).

1.6 ESTABLISHMENT OF *GRACILARIA* FARMING IN SOUTH AFRICA

The South African coastline is unique in that it is separated into warm and cold regions with the warm Agulhas and Mozambique currents on the East coast, and the cool Benguela Current on the west. The coastline therefore supports a continuum of diverse and economically valuable seaweed species.

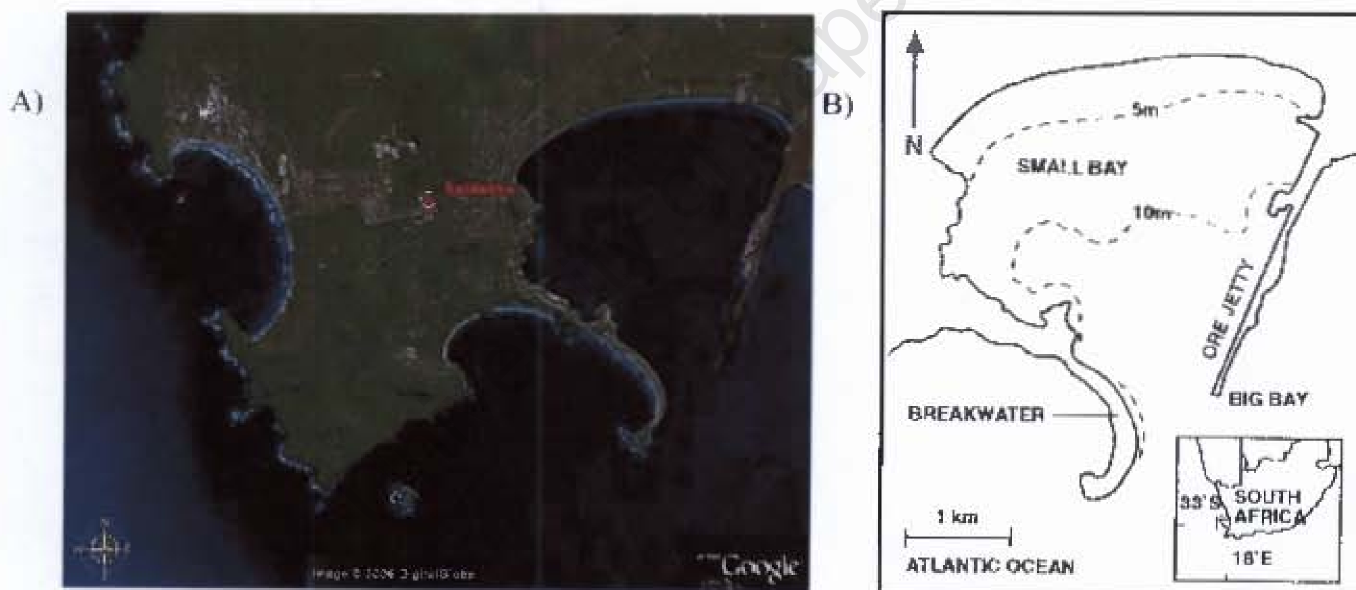


Fig 1.2. Saldanha Bay, South Africa. A) Google Earth Satellite Image; B) diagram of Saldanha Bay (taken from Anderson *et al.*, 1999).

Gracilaria (gracilis) populations occur naturally in the cool temperate waters of the West coast, with accessible commercial beds situated in Saldanha Bay (Fig 1.2). In the past, these populations were seen as a public nuisance as large quantities of *Gracilaria* would wash ashore and would be left to rot in an otherwise recreational and residential area. It was not until the early 1950s, shortly after the conclusion of World War 2, when agar supplies from Japan became unavailable in Britain that commercial interest in this species was aroused (Rotmann, 1990, Anderson *et al.*, 1989).

An agar extraction industry was thus established; with traders in Cape Town promoting *Gracilaria* to agar producers worldwide (Rotmann, 1990). During this period yields of over 1000 tons dry weight (d wt) of beach cast material were typical (Isaac, 1956), encouraging the establishment of two agar processing factories in the Western Cape in the 1960s (Rotmann, 1990; Anderson *et al.*, 1989). The longevity of these factories however was short lived. High establishment costs as well as insufficient technology were initially problematic, but it was the construction of a break water and ore-jetty (Fig. 1.2B) within Saldanha Bay in 1974, that was accredited for their eventual demise (Anderson *et al.*, 1996b).

The breakwater and ore-jetty generated changes in the water flow characteristics within the bay. These changes encouraged the development of strong thermal stratification of the water column, with oligotrophic surface waters prevailing in the summer months. As

University of Cape Town

a result, the unavailability of nutrients, particularly nitrogen, resulted in the collapse of the *Gracilaria* resource (Anderson *et al.*, 1996b).

Since 1974 revival of this resource has been slow, with enormously variable commercial yields attained. Yields began to recover in the early 1980s with 429 dry tons recorded in 1988 (Anderson *et al.*, 1993). This stimulated interest in once again establishing an agar-processing factory in Saldanha Bay, but by the end of 1988 *G. gracilis* populations had once again collapsed. The cause of the collapse remains unclear, but it is believed to have been due to poor attachment of *Gracilaria* as a consequence of rapid growth, coupled with the presence of a substantial number of invertebrate grazers and bacterial epiphytes (Anderson *et al.*, 1993; Jaffray *et al.*, 1996). No further beach casts were obtained until 1992, but since then yields of about 400 t dry wt have persisted (Anderson *et al.*, 1996b).

In the summer of 1993-1994 all beach cast *Gracilaria* had to be discarded due to contamination by a localized bloom of *Ulva* within Saldanha Bay. The bloom resulted due to eutrophication by effluent waste from two fish processing factories in Saldanha Bay (Anderson *et al.*, 1996b; Anderson *et al.*, 1999). These factories discharge large amounts of nitrogen into the water on the west shore of the bay promoting fouling by opportunistic organisms when environmental factors favour their establishment over that of *Gracilaria.*

The unstable yields obtained from the Saldanha Bay *Gracilaria* populations has thus prevented the development of a local agar processing company, but simultaneously has stimulated cultivation attempts. It was thus in 1990 when the natural populations had

University of Cape Town

effectively disappeared that research on suspended *Gracilaria* cultivation was initiated. Cultivation techniques based on design criteria used in Lüderitz, Namibia (Dawes, 1995) were implemented, and growth rates of approximately 5% d^{-1} were obtained (Anderson *et al.*, 1996a). An attractive outcome of this procedure was that detached *Gracilaria* from the cultivation lines aided in replenishing the substratum underneath, maintaining the natural bottom populations in Saldanha Bay (Anderson *et al.*, 1993).

During the summer months however, raft-cultivated *Gracilaria* was adversely affected by the warm and oligotrophic conditions that prevail within Saldanha Bay. This factor, together with the increased fouling by mussels and the tunicate *Ciona*, resulted in subsequent abandonment of commercial suspension cultivation (Anderson *et al.*, 2003).

Currently commercial exploitation of *Gracilaria* in Saldanha Bay relies on the labour intensive collection of beach cast material (Anderson *per comm*). Accordingly, the majority of inhabitants in Saldahna Bay earn their livelihood from seasonal fishing operations (Rotmann, 1990). At present the *Gracilaria* industry in Saldanha Bay employs 60-100 people, with the potential to utilize a greater work force (Anderson *et al.*, 2003). In Chile an estimated 9 000 to 11 000 fishermen in rural and coastal areas depend on algae for part or all of their income (Vâsquez *et al.*, 1993). Further development of this industry in Saldanha Bay would thus act to both augment the livelihood of inhabitants in the area as well as enhance the South African economy as a whole.

University of Cape Town

Currently all high quality dried *Gracilaria* from Saldanha Bay is exported to Japan, Korea and Chile to be used as 'natural agar', while the lower quality product is delivered to Namibia (Rotman, 1990; Anderson *et al.*, 2003). The South African *Gracilaria* industry although small, is thus capable of competing in a highly competitive international market. In order to maintain a healthy industry however, Anderson *et al.* (2003) believe that it is necessary to take the lead in developing new products. They suggest that South Africa would benefit more by creating niche markets and developing high-value products such as cosmetics or plant growth stimulants rather than relying solely on export of this raw material. In addition the water quality along the South African coast is excellent, suggesting that the development of products for human consumption or cosmetic use could possibly enjoy a marketing advantage over similar products from countries where water quality might not be as high (Anderson *et al.*, 2003).

Thus, although the *Gracilaria* industry in Saldanha Bay is relatively successful, total *Gracilaria* yields and product development need to be assessed in order for the industry to remain viable.

Of all the environmental factors shown to have impaired *G. gracilis* cultivation within Saldanha Bay, nutrient limitation, and in particular nitrogen limitation, is considered to be the most critical. A clear understanding of how the seaweed endeavours to adapt and tolerate the 'stress' is therefore necessary to improve *G. gracilis* productivity. Stress resistance strategies often involve the activation of an adaptive response through signal transduction. Various genes are activated or repressed, and new proteins are synthesized.

University of Cape Town

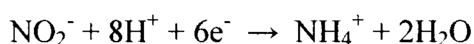
Thus understanding the genetic response of *G. gracilis* to nitrogen deprivation may lend itself to future genetic manipulation and engineering of stress tolerant *G. gracilis* species, ultimately improving *Gracilaria* cultivation.

1.7 NITROGEN ASSIMILATION IN PLANTS AND ALGAE

All living organisms require nitrogen in order to produce various structural and functional macromolecules such as proteins and nucleic acids. In the marine ecosystem the most common forms of inorganic nitrogen are ammonium and nitrate (Lopes *et al.*, 2002). In order to be assimilated into the biosynthesis of nitrogen-containing compounds, nitrate must be reduced to ammonia or amine in a two step reaction that is sequentially catalyzed by the enzymes nitrate and nitrite reductase. The first step of the reaction, i.e. the reduction of nitrate to nitrite, requires NADH or NADPH as an electron donor. This step occurs in the cytosol of algal cells, and is generally considered to be the rate-limiting step in the nitrate assimilation process (Lopes *et al.*, 2002). The reaction occurs as follows:

$$NO_3^- + NAD(P)H + H^+ \rightarrow NO_2^- + NAD(P)^+ + H_2O$$

The resulting nitrite is transported to chloroplasts where the second step of the reaction is initiated (Lobban and Harrison, 1994). Thus nitrite reductase catalyses the reduction of nitrite to ammonium as follows:

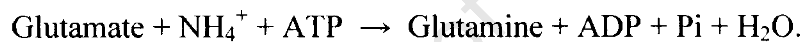


$$NO_2^- + 8H^+ + 6e^- \rightarrow NH_4^+ + 2H_2O$$

University of Cape Town

Ammonium is generally toxic to plant cells because of its ability to uncouple respiration at low concentrations. Thus strict regulatory control is required in the nitrogen assimilation pathway in order to rapidly assimilate ammonium into non-toxic organic compounds (Temple *et al.*, 1998). In the past, glutamate dehydrogenase (GDH) was considered to be the key enzyme governing this process. This concept however was revised in the 1970s with the discovery of the enzymes glutamine synthetase (GS) and glutamate synthetase (GOGAT) (Temple *et al.*, 1998; Inokuchi *et al.*, 2002). The joint action of these two enzymes is collectively referred to as the GS/GOGAT cycle and is now believed to be the primary route of nitrogen assimilation.

Glutamine synthetase catalyzes the first reaction in this cycle, with the amination of glutamate to yield glutamine in an ATP dependent reaction:



$$\text{Glutamate} + NH_4^+ + \text{ATP} \rightarrow \text{Glutamine} + \text{ADP} + \text{Pi} + H_2O.$$

The amide group of the glutamine molecule is then transferred by a reaction catalyzed by GOGAT to α-ketoglutarate (2-oxoglutarate (2-OG)) to yield two molecules of glutamate:

$$\text{Glutamine} + \text{2-OG} + \text{NAD(P)H} \rightarrow 2\ \text{Glutamate} + \text{NAD(P)}^+$$

Synthesised glutamate is then able to donate its amino group to form other nitrogen-containing compounds, or it is reduced to replenish the pool of glutamate for subsequent glutamine synthetase catalysis.

University of Cape Town

Transfer of the amino group from glutamate is catalyzed by the enzyme aspartate aminotransferase (AspAT), and requires oxaloacetate from the tricarboxylic acid (TCA) cycle in order to form aspartate. The reaction is thus as follows:

Glutamate + Oxaloacetate ↔ Aspartate + 2-OG

The aspartate so produced can then be directed to one of two pathways. Aspartate can be transferred to the matrix of mitochondrial cells where it is re-transaminated to oxaloacetate, or it can be used together with glutamine and ATP to form asparagine through catalysis by asparagine synthetase (AS). This reaction is reversible and occurs as follows:

Glutamine + Aspartate + ATP ↔ Glutamate + Asparagine + AMP + PPi

Glutamine, glutamate, aspartate and asparagine thus provide the starting points for the synthesis of other organic compounds, initiating a wide variety of biosynthetic reactions. Overall, the GS/GOGAT pathway cannot occur without the TCA intermediates, α-ketoglutarate and oxaloacetate. Nitrogen assimilation and carbon metabolism are thus strongly co-ordinated through the requirement for these intermediates (Fig. 1.3).

G. gracilis ammonium-nitrogen (NH_4^+-N) uptake has been described as a linear, rate-unsaturated response, with the slope increasing with nitrogen limitation (Smit, 2002). Nitrate-nitrogen (NO_3^- -N) uptake on the other hand is energetically costly since it is

University of Cape Town

somewhat dependent on photosynthesis, and is best approximated by the Michaelis-Menten equation. Generally *G. gracilis* shows a higher affinity for NH_4^+-N than for NO_3^--N, however under nitrogen limited conditions the affinity for both forms of dissolved inorganic nitrogen is enhanced.

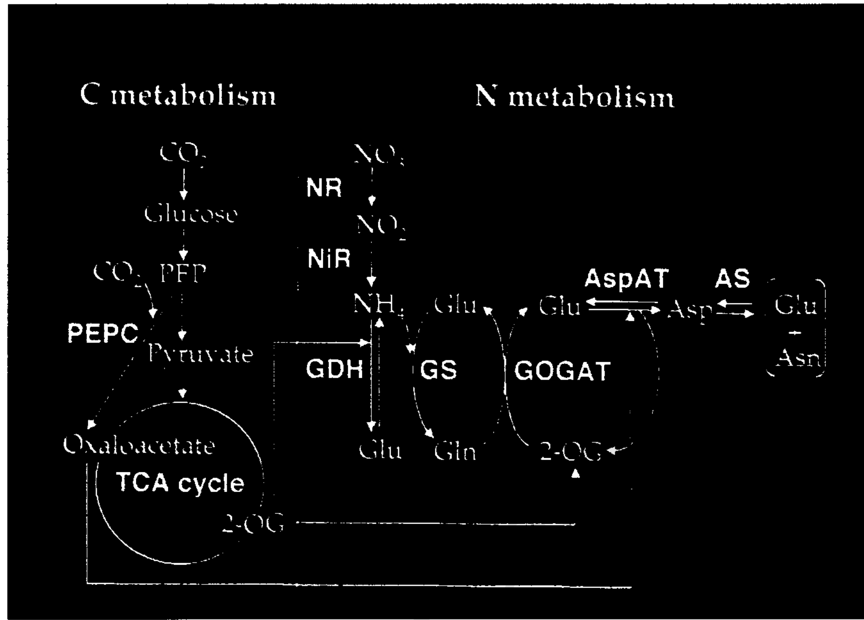


Fig 1.3 Flow chart showing the co-ordination between carbon and nitrogen metabolism in plant cells (Inokuchi *et al.*, 2002).

University of

1.8 GENERAL 'STRESS' RESPONSE OF PLANTS AND ALGAE TO UNFAVOURABLE CONDITIONS

As with land-based plants, *G. gracilis* is non-motile and thus unable to avoid unfavourable fluctuating environmental conditions. In order to survive, plants and algae have developed various defence responses that require regulatory pathways or systems

for the activation of multiple gene pathways. Such responses may be either reversible (elastic strain) or irreversible (plastic strain) in nature, but generally favour plant or algal survival at the expense of overall productivity (Treshow, 1970). In addition, the swiftness and sensitivity of the response to the 'stress' plays a fundamental role in plant or algal viability, as does its capacity to recover after the stress has abated.

The primary factor in plant or algal acclimation or 'tolerance' to an apparent stress is the ability of the plant or algae to perceive the stress through various internally regulated signals. For example, concentrations of the cytoplasmic pool of cycling amino acids, or various intermediary products such as 2-oxoglutarate, have been found to serve as internal regulatory signals in sensing the nitrogen status of plants and cyanobacteria (Crawford *et al.*, 1998; Muro-Pastor *et al.*, 2001). These signals allow for transcriptional activation or repression of various genes which in turn results in increased or decreased protein synthesis, respectively.

Although plant 'stress' responses are primarily regulated through altered gene expression, plant hormones may play an additional role in allowing the plant to respond to an environmental cue (Morgan, 1990). Plant hormones have been described as 'chemical messengers regulating the normal progression of developmental changes as well as responses to environmental signals' (Morgan, 1990). Thus an increase or decrease in these hormone levels under adverse conditions through altered gene expression may result in one of three processes outlined in Figure 1.4. Briefly, hormones may serve as transcriptional regulators inducing the synthesis of various "shock proteins"; alternatively

University of Cape Town

they may act at a post-translational level by either altering protein phosphorylation reactions thereby shifting enzyme activity or by direct interaction with membrane transporters altering ion/substrate fluxes.

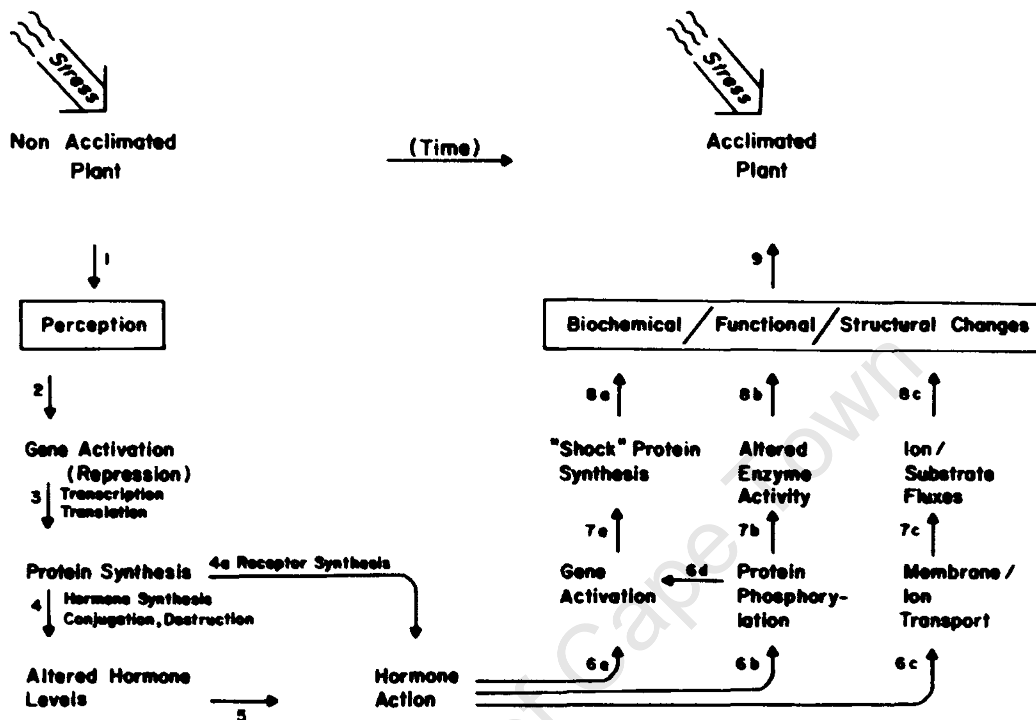


Fig. 1.4 Proposed model for the role of plant hormones in acclimation to stress (Morgan, 1990). The numbered arrows 1 – 9 represent the sequence of events involved in plant acclimatisation. Numbers with accompanying under-case letters represent alternate pathways that may be followed within the acclimatisation process.

University of Cape Town

Understanding the intricate signalling systems in plants is highly complex. Often a cell uses the same signalling system or 'universal' gene set to respond to more than one type of stress (Morgan, 1990). A clear example of this is the identification of altered 'pool' levels of several plant hormones in response to a range of abiotic stresses (Table 1.1).

Thus identification of a unique gene set activated in response to a specific stress is particularly complicated and at the same time enormously desirable.

Table 1.1 Summary of the effects of various abiotic stresses on levels of expression of several plant hormones (Morgan, 1990). Each arrow represents either an increase or decrease in the hormone level relative to basal concentrations.

Hormones	Drought	Salt	Heat	Cold	Chilling	LowN/P
Abscisic acid	↑	↑	↑	↑	↑	↑
Cytokinins	↓	↓	↓			↓
Gibberellin	↓			↓		
Indole acetic acid	↓					

1.9 GENETIC RESPONSES TO NITROGEN LIMITATION IN ALGAE AND HIGHER PLANTS

Various gene products such as NblR (non-bleaching), NtcA (nuclear transcription factor) and GlnN (glutamine synthetase III) have been found to participate significantly in plant or algal survival under nitrogen limiting conditions (Luque *et al*., 2001). Phenotypically, nitrogen-deficiency is characterised by stunted growth, sparse foliage, and a loss of colour due to the breakdown of various pigment proteins (phycobilisomes) for nitrogen recycling and retrieval. The latter process has been described as 'chlorosis' and is mediated by a family of *nbl* genes (Sauer *et al*., 2000).

University of Cape Town

When plant or algal cell metabolism is slow, excessive absorption of excitation energy is superfluous, detrimental and uneconomical. In response to this, plants have the ability to control their energy expenditure through the proteolysis of their energy absorbing pigment proteins through the activation of the *nbl*A and *nbl*B gene products. Such control allows for enhanced cell metabolism and survival. Under nutrient sufficient conditions, the expression of the *nbl*A gene is very low and, although its product is not active in phycobilisome degradation, the *nbl*B gene seems to be constitutively expressed (Sauer *et al.*, 2000). At the onset of stress conditions, the *nbl*A gene is rapidly upregulated and it in-turn activates degradation by the *nbl*B gene product. Thus, in *Gracilaria* and other red algal systems the loss of thallus pigmentation could be an indication that the *nbl* system has been activated.

The *nbl*A gene product itself is under the control of two response regulators, NblR and NtcA (Luque *et al.*, 2001). NblR is a general regulator of *nbl*A expression and is ubiquitous in that it is activated in response to a variety of stresses. NtcA on the other hand is a DNA binding protein that is considered to be a global response regulator specific for regulating the plant's response to nitrogen. NtcA synthesis is upregulated in response to low ammonium levels, regulating the expression of its own gene as well as those required for the utilization of nitrate and nitrite (Lindell *et al.*, 2002). Thus NtcA targets include all genes involved in nitrogen assimilation (see section 1.6), including those in the *nbl* gene family.

University of Cape Town

The *gln*N gene product, glutamine synthetase III, is considered to be important in plant recovery following prolonged periods of nitrogen starvation. *gln*N is a candidate for NtcA control, as NtcA binding motifs within the promoter regions of the gene have been identified (Sauer *et al.*, 2000). Synthesis of GlnN is strongly induced under conditions of nitrogen starvation (Sauer *et al.*, 2000), and acts to metabolise and assimilate the nitrogen released from phycobilisome degradation (Luque *et al.*, 2001).

In addition to *gln*N activation, NtcA has been shown to regulate transcription of the *gln*B gene product, P_{II}, and influence its phophorylation level in response to nitrogen and carbon supplies (Lee *et al.*, 1999a). P_{II} is considered to be one of the most ancient signalling proteins thus far identified, with examples of this protein having been found in bacteria, archae, algae and higher plants (Ninfa and Jiang, 2005). P_{II} commonly serves as a sensor for α-ketoglutarate levels and the regulation of GS activity (Ninfa and Jiang, 2005). In addition P_{II} is believed to control/regulate arginine biosynthesis in Arabidopsis through interaction with, and regulation of, N-acetyl-glutamate kinase (Ferrario-Méry *et al.*, 2006). Lastly P_{II} is believed to play a role in modifying the activity of nitrate/nitrite transporters as well as high affinity bicarbonate transporters in cyanobacteria (Ferrario-Méry *et al.*, 2006).

University of Cape Town

Plants have evolved regulated, energy dependent systems for the uptake of nitrate using both high and low affinity transporters. In general, transport systems are induced in the presence of a substrate if the cell requires it, and induction continues until cellular requirements are met (Hildebrand and Dahlin, 2000). Two nitrate transporter gene

families have been identified in plants to date: the NRT1 and NRT2 families. Unlike NRT1, NRT2 is specifically nitrate-inducible, and constitutes a high affinity nitrate uptake system. NRT2 has additionally been found to be conserved amongst fungi, diatoms, algae and higher plants (Hildebrand and Dahlin, 2000; Crawford and Glass, 1998). Activation of these nitrate transporters ultimately results in an increase in the possible influx of nitrate from the surrounding environment and thus increased activity of the genes involved in nitrate assimilation, namely nitrate reductase (NR) and nitrite reductase (NiR).

On the whole, various endogenous and environmental stimuli have been found to play a role in regulating the gene expression and enzyme activity of various proteins involved in the nitrate assimilatory pathway (Rothstein and Sivasankar, 1999). Despite this, understanding of the molecular basis for the regulation of nitrate assimilation is limited. For instance, the NiR gene promoter has been extensively studied and a number of promoter elements responsible for nitrate inducible expression have been identified (Reynolds, 1999). Regardless of this the mechanism of glutamine and asparagine repression is still relatively unknown (Reynolds, 1999).

University of Cape Town

Overall, a number of genes have been identified that play a significant role in nitrogen metabolism and transport. In order to fully understand the complexity of nitrogen specific plant and algal gene and gene pathway regulation however, a great deal of research is still required. In the long term, an improved understanding of those genes specifically activated or repressed in response to nitrogen starvation in *G. gracilis* could possibly

enhance our ability to engineer nitrogen metabolism for improved growth and productivity.

1.10 MICROARRAY TECHNOLOGY

Over the past decade various genomic technologies have emerged that have been useful in elucidating the changes in plant cellular metabolism that are induced by abiotic stresses. Such technologies have provided insights and answers on how plant genes, proteins, protein activities, and metabolic type and flux respond to external factors (Bohnert *et al.*, 2006). Microarray technology is one such emerging genomic tool that has helped to provide important insights into the dynamics of the transcriptional changes that accompany abiotic stress treatments.

University of Cape Town

In general microarray technology allows for the rapid parallel surveillance of the expression of thousands of genes in a single assay. It is therefore extremely advantageous in that both individual gene and multi-gene expression patterns can be obtained, providing clues to the various regulatory mechanisms and biochemical pathways activated in response to an applied stress.

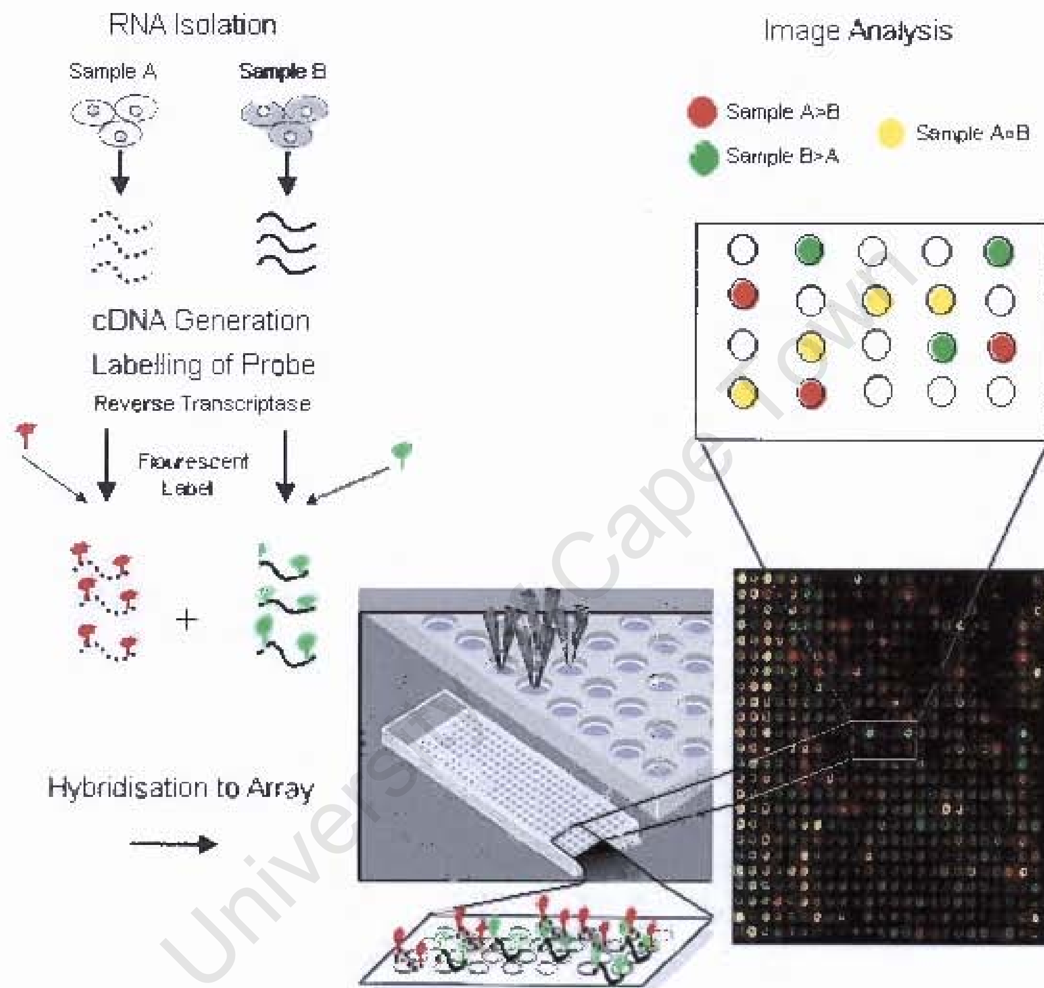
The microarray technique itself is performed by hybridising fluorescent probes, derived from messenger RNA (mRNA) samples from an organism subjected to either stressful or normal environmental conditions, to a glass slide containing various cDNA sequences from the organism in question (Fig 1.5). These cDNA sequences are spotted and fixed onto the surface of the slide, and serve as the 'target'. Each target may be spotted once or

multiple times per array. Probes are created through reverse transcription of the mRNA samples to form complementary DNA (cDNA), with the dye label incorporated either during (direct-labelling) or after (indirect-labelling) the reverse transcription reaction. Probes are labelled such that one sample is labelled with a 'green' fluorescent dye (Cy3), while the other is labelled with a 'red' fluorescent dye (Cy5). These samples are then combined and injected onto the array to allow for hybridisation. The labelled probes are able to bind to their complimentary sequences on the array and unhybridised cDNA is removed. The fluorescent emissions of specifically bound probe are then detected using an appropriate scanner. The light emitted from the Cy3 and Cy5 dyes during the scanning process is captured by a photomultiplier tube (PMT) detector and converted into electric current. The red (R) and green (G) fluorescence emissions at each pixel location on the microarray are quantified and saved as a 16 bit tagged image file (Simon *et al.*, 2003a). This serves as the raw data for further analysis.

University of Cape Town

Essentially the R and G fluorescent intensities are a representation of the level of hybridisation of the two probe samples to the target cDNA sequences spotted on the slide, and thus a quantitative estimate of gene expression. However, various anomalies such as non-specific binding of the labelled probe to the array, dust or even autofluorescence from the target DNA may contribute to the overall intensity value for a specific target spot on the array. As a result, a number of data cleaning steps or 'low level analyses' are required in order to extract the numerical red and green foreground (*Rf*, *Gf*) and background intensities (*Rb*, *Gb*) for each spot on the array. This process is referred to as image analysis and involves the subtraction of the local background intensities from

the foreground intensities ($R = Rf \quad Rb$, $G = Gf \quad Gb$) for each spot on the array. In this way any intensity measurement that is not specifically due to the hybridisation of probe to target is eliminated, and a more accurate quantification of transcript abundance for each sample is achieved.



University of Cape Town

Fig. 1.5 Flow diagram of microarray procedure (taken from Krutovskii and Neale, 2001)

During the microarray experiment various sources of systematic variation arise from technical foundations such as differences in print-tip or spatial effects, differences in the

labelling efficiencies and scanning properties of the Cy3 and Cy5 dyes, as well as differences in the PMT voltages used to scan the arrays (Smyth and Speed, 2003). Therefore, before the gene expression profiles of treatment and control RNA samples can be analyzed and interpreted, the red and green intensities must be normalized relative to one another so that the red/green ratios are, as far as possible, an unprejudiced representation of relative expression (Smyth and Speed, 2003).

Following image analysis, the data for each gene is typically reported as an 'expression ratio' of Cy3/Cy5. Induction or repression of a gene is therefore seen to have values of different magnitude such that a two-fold induction has an expression ratio of two, while a two-fold repression has a value of one-half. Normalization however, is generally applied to the $\log_2$ transformed expression ratios, allowing induction and repression of genes to be treated as values of identical magnitude. Therefore, a twofold induction would have a $\log_2$(ratio) of 1, while a two-fold repression would have a numerical value of -1. Similarly, genes that are expressed at a constant level will have a $\log_2$(ratio) of 0.

There are a number of different normalization algorithms that can be applied to normalize the intensities of the hybridised probes within and across arrays (Smyth and Speed, 2003; Yang *et al.*, 2001). In general, the normalization algorithm chosen depends on both the design of the experiment as well as the results obtained after scanning the arrays (Yang *et al.*, 2001). In addition, graphical representation of the raw- and normalized data can help guide the choice of analysis tools by highlighting any specific problems (Smyth *et al.*, 2002). For example, density plots of the Cy3 and Cy5 intensities for each

University of Cape Town

of the arrays performed enables the visualisation of possible dye bias on and across arrays. Alignment and thus overlapping of the two dyes a-priori of normalization indicates successful normalization within arrays and/or across arrays. Additionally, MA-plots, which are essentially plots of the M-values or $\log_2$ (expression ratio) values on the X-axis and A-values or log-intensities (of fluorescence of each spot) on the Y-axis, can be drawn in order to display the relationship between differential gene expression and intensity. MA-plots of un-normalized data will tend to show any dye biases occurring at high and low intensities. Once normalized, the majority of the data should lie around zero. Any outlying data points therefore represent genes that may be differentially expressed.

The simplest and most widely used within-array normalization method is Global normalization, and is based on the assumption that the red-green bias is constant on the log-scale across an array (Smyth *et al.*, 2002). The log-ratios of expression (M-values) are corrected by subtracting a global constant c estimated from the mean or median M-value, to get normalized values $M = M - c$. Red-green intensities however are usually not constant across the spots within and across arrays, and can vary according to overall spot intensity, location on the array, slide origin, as well as various other variables (Smyth and Speed, 2003). In order to account for both spatial and spot intensity variations, alternative normalization methods such as print tip loess (Smyth and Speed, 2003) and robust-spline normalization (Soler *et al.*, 2004) have been developed. In addition, housekeeping genes can be used to normalize arrays if originally spotted onto the array itself (Yang *et al.*, 2001).

University of Cape Town

The last step involved in microarray analysis is statistical assessment of the normalized data to distinguish which genes are statistically differentially expressed. Generally when two samples are compared, statistical analysis involves *t*-test statistics and variations thereof (Kooperberg *et al.*, 2002). The merit of the statistical procedure employed depends on the ability to successfully identify differentially expressed genes while avoiding the elimination of genes that are differentially expressed (Type 1 error) and/or classifying genes that are unchanged as differentially expressed (Type 2 error). Overall, the success of a microarray experiment in detecting differentially expressed genes is dependant on its reproducibility following repetitive experimentation (Naderi *et al.*, 2004).

University of Cape Town

1.11 AIMS OF THIS STUDY

The objective of the current study was to use microarray technology to identify those genes that are differentially expressed in response to nitrogen starvation in the red alga, *G. gracilis*. In order to achieve this it was necessary to prepare and hybridise 'probe' samples to *G. gracilis* microarrays; acquire and normalise data from the images obtained following hybridisation and scanning; perform various statistical analyses to determine which genes on the array are differentially expressed, and lastly to putatively identify the function of those genes that are found to be differentially expressed with the aid of gene sequencing and bioinformatic analysis.

Before this could be achieved however, it was necessary to first determine when *G. gracilis* becomes sufficiently nitrogen limited during cultivation under nitrogen-limited conditions so that samples could be collected for probe synthesis. In addition, as *G. gracilis* is a non-model organism it was necessary to synthesize 'target' cDNA and construct arrays specific for nitrogen limited *G. gracilis*. Lastly, it was necessary to optimise each of the technical preparatory steps used for microarray analysis. As a result, a reproducible RNA isolation protocol for *G. gracilis* was required, and optimisation of the cDNA synthesis, labelling and hybridisation steps was necessary.

University of Cape Town

CHAPTER 2

NITROGEN DEPRIVATION: ACQUISITION OF NITROGEN DEPRIVED *G. GRACILIS* THALLI FOR USE IN MICROARRAY

2.1 INTRODUCTION 36

2.2 MATERIALS AND METHODS 37

2.2.1 Sample collection 37
2.2.2 Nitrogen deprivation experiment 37
2.2.3 Total thallus nitrogen and carbon content 40
2.2.4 Statistical analysis of RGR (%), nitrogen content and C:N ratio data 40

2.3 RESULTS AND DISCUSSION 40

2.3.1 Induction of nitrogen deprivation 40
2.3.2 Total thallus nitrogen content 45

University of Cape Town

CHAPTER 2

NITROGEN DEPRIVATION: ACQUISITION OF NITROGEN DEPRIVED *G. GRACILIS* THALLI FOR USE IN MICROARRAY

2.1 INTRODUCTION

One of the most fundamental aspects of any microarray experiment, like all other experiments, is for the researcher to have a clear idea of the overall objective of the study in question. Generally, the objective of a DNA microarray experiment falls into one of three categories, namely: class comparison, class prediction or class discovery. In a class comparison experiment two mRNA samples exposed to different experimental conditions are compared in order to detect differentially expressed genes. Alternatively, a class prediction experiment is performed to allow for the development of a multigene formula that can be applied to expression profiles of samples whose class is unknown, and in so doing, allow for the prediction of the class of the new samples (Simon and Dobbin, 2003). Lastly, class discovery experiments are employed when the objective is to discover classes of genes that are co-regulated (Simon and Dobbin, 2003).

University of Cape Town

In the current study, the biologically relevant question proposed was whether there is a difference in gene expression between healthy nitrogen enriched and nitrogen limited *G. gracilis* thalli, and if so what are the genes that are differentially expressed. Thus the objective of the study falls into the category of a class comparison experiment, where the nitrogen limited and nitrogen enriched thalli are defined classes, and the contrast in the expression profile between the samples is the comparison to be assessed.

G. gracilis tissue from both nitrogen limited and nitrogen enriched conditions was thus required in order to obtain test and reference samples for microarray hybridization, respectively. Previous documentation suggests that *G. gracilis* can tolerate conditions of nitrogen limitation for one week after which the alga is considered to be nitrogen limited (Smit *et al.*, 1997; Anderson *et al.*, 1996b). In order to confirm this and to establish when samples should be collected for use in mRNA extraction for microarray, it was essential to determine when the nitrogen limiting conditions specified in this study induced a state of nitrogen deprivation in *G. gracilis* thalli. Thalli cultivated in nitrogen limited media were thus monitored for a three week period with thalli exposed to nitrogen enriched media serving as a control. Samples were removed at specific times and assessed according to their appearance, relative growth rate (RGR), total nitrogen content and carbon to nitrogen (C:N) ratios.

University of Cape Town

2.2 MATERIALS AND METHODS

2.2.1 Sample collection

Gracilaria gracilis plants were obtained from Jacob's Bay abalone farm in Jacob's Bay, Cape Town, South Africa. Deionized water was used to remove sediment and visible epiphytes from the collected plants, and healthy thalli were selected for experimental use.

2.2.2 Nitrogen deprivation experiment

In order to determine when *G. gracilis* thalli become nitrogen deficient, sixteen thallus fragments weighing 2.0 ± 0.05 g fresh weight were transferred to sixteen 1 L Erlenmeyer flasks (Fig 2.1) containing 250 ml artificial seawater (ASW) (Appendix A.1) with a

salinity of 3.6 %. Erlenmeyer flasks were divided such that eight flasks contained control, nitrogen enriched thallus samples, while the remaining eight flasks housed experimental, nitrogen limited thallus samples. The ASW in control flasks was enriched according to Provasoli (1968) with PES medium (1/3 strength) (Appendix A.1), whereas ASW in the experimental flasks was supplemented with PES medium lacking $NaNO_3$, and amended with $Fe_2(SO_4)$ instead of $Fe(NH_4)_2SO_4$. The latter supplementary medium is hereafter referred to as PES-N (Appendix A.1). The respective culture media was replaced every 3 days in order to prevent nutrient depletion (other than nitrogen limitation).

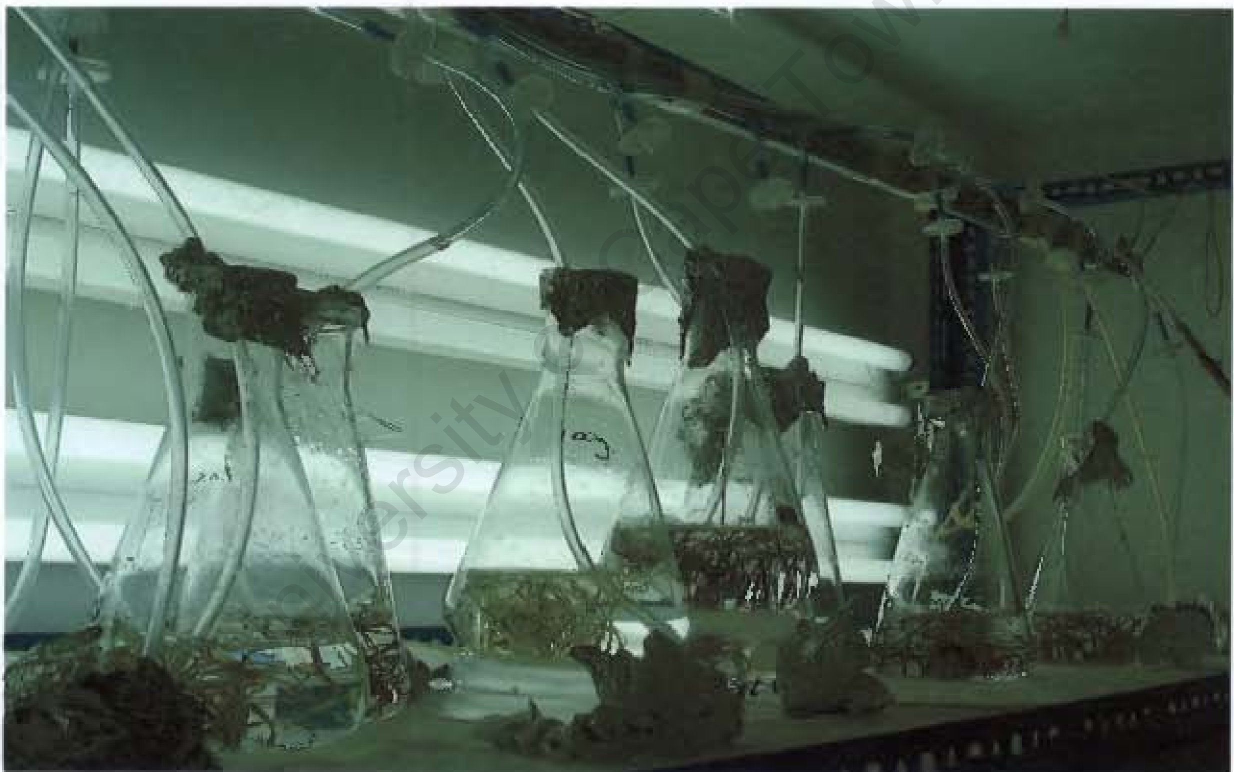


Fig 2.1 Experimental design of the nitrogen limitation experiment showing *G. gracilis* thalli in nitrogen limited and nitrogen enriched ASW.

Aeration and water movement was achieved by passing compressed air through a 0.22 μm Millipore filter attached to plastic airlines supplied to each flask. Incubation temperature was maintained at 15-17 °C, and illumination was provided by cool white fluorescent tubes at 100 μmol photons m^{-2} s^{-1} with a 16:8 light-dark cycle.

In order to determine the relative growth rate (μ) of the experimental and control *G. gracilis* thalli, samples were weighed on days 0, 1, 7, 12, 14, 16, 17, 18, 19, 20 and 28. The Relative Growth Rate (μ), which is the percentage increase in fresh weight per day, was thus calculated using the following equation:

$$\mu = [100 (\ln (N_t / N_o))] / t$$

where: t is the time in days, N_o is the initial weight of the thallus fragment at day 0, and N_t is the weight of the thallus after t days (Smit and Bolton, 1999).

Samples for CHNS analysis were removed on days 0, 1, 7, 14, 17, 18, 19, 21 and 28. From each sample a 0.5 g fragment was removed for use in CHNS analysis, wrapped in foil, flash frozen and stored at -70 °C until all samples were collected. Each thallus sample was thawed forty eight hours prior to CHNS analysis and oven dried at 60 °C.

In order to verify the results, the entire experiment was performed in duplicate.

University of Cape Town

2.2.3 Total thallus nitrogen and carbon content

CHNS analysis was performed on thallus fragments (approximately 1.5 ± 0.2 mg) from each sample acquired in section 2.2.2 above. Measurements for each sample were performed in duplicate with both total carbon and nitrogen content recorded. All measurements were carried out using a CHNS-932 analyser (Leco Co-orp, st Joseph, MI, USA), and were performed by Heather Sessions at the Marine and Coastal Management laboratory, Cape Town, South Africa.

2.2.4 Statistical analysis of RGR (%), nitrogen content and C:N ratio data

The RGR, nitrogen content and C:N ratio data for the nitrogen enriched and nitrogen limited thalli (collected over the 21 day sampling period) are presented as means and standard errors. Differences between the means of the nitrogen enriched and nitrogen limited RGR data were compared using the Mann-Whitney Rank Sum Test. Nitrogen content and C:N ratio data was analysed using a two way analysis of variance (ANOVA) computation. When the effects of the ANOVA were significant, the Holm-Sidak method was used to test for significant differences between the sample means. All statistical analyses were performed using Sigma Stat (version 3.1, Systat software, Inc). In all cases the level of significance was set at $P < 0.05$.

University of Cape Town

2.3 RESULTS AND DISCUSSION

2.3.1 Induction of nitrogen deprivation

Nitrogen deprivation experiments were established in order to determine when *G. gracilis* is sufficiently nitrogen limited and thus optimal for use as a treatment sample for

microarray analysis. Previous studies suggest that *G. gracilis* is able to grow at non-limited rates under nitrogen limitation conditions for one week using only internally stored nitrogen to sustain growth, after which the growth rate decreases substantially (Smit *et al.*, 1997). Thus in the current study *G. gracilis* thalli were incubated in either nitrogen enriched or nitrogen limited ASW for a period of three weeks with sampling times spread over this period.

G. gracilis typically displays the strategy of luxury consumption having the ability to store nitrogen in excess of the required concentration during periods of high nitrogen availability (Lignell and Pedersén, 1987). The nitrogen so acquired is stored in a number of ways, with proteins and phycoerythrins comprising the largest constituents in the internal nitrogen pool, while chlorophyll-*a*, carotenoids and DNA are very small contributors (Smit *et al.*, 1997). One of the first visual indicators of nitrogen deprivation is decreased growth rates, as well as changes in thallus colour due to amino acid and pigment (phycoerythrin) breakdown, respectively. Thus, both thallus colour and relative growth rate (RGR) were monitored over the entire experimental period in order to visually assess the nitrogen status of the *G. gracilis*.

University of Cape Town

In the current study, light green patches on the *G. gracilis* thalli started to appear after seven days exposure to the nitrogen limited media, indicating potential phycoerythrin breakdown. This phenomenon exacerbated over time, with thalli (in the nitrogen limited media) becoming almost completely green by day 14. Thalli became progressively greener (Fig 2.2), with thallus fragmentation occurring by day 28 in some of the samples.

Thalli from day 28 samples were generally brittle and straw-white in colour, suggesting that the adaptive response by *G. gracilis* to nitrogen stress may have already occurred. Due to fragmentation, the day 28 thallus samples were not included in RGR or CHNS analysis. Overall, based on thallus pigmentation and integrity, the optimal sampling period for obtaining nitrogen limited samples for the microarray analysis appeared to be between days 14 and 21.

University of Cape Town

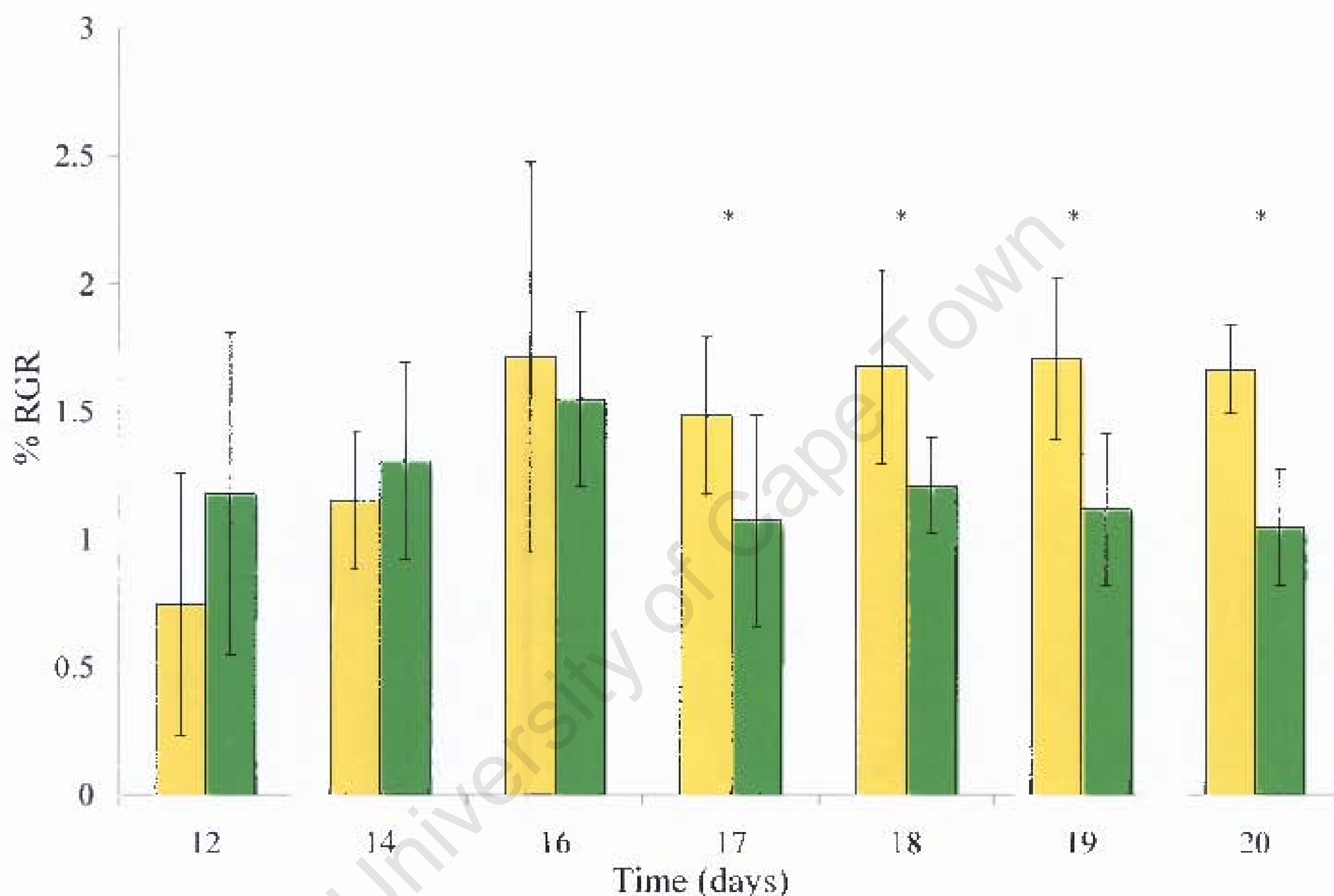
Fig 2.2 Morphology of *G. gracilis* after 18 days of cultivation in nitrogen enriched ASW (A and C) and nitrogen limited ASW (B and D). Note the greenish colour of the algal thallus in the latter sample.

G. gracilis thalli in the nitrogen enriched media retained its red brown colour over the entire 28 day period with only slight discolouration occurring towards the concluding days of the experiment. This suggests that either the PES medium or the experimental conditions themselves may have been sub-optimal. The literature suggests that nitrogen uptake is influenced by a number of environmental conditions including light, water movement, CO_2 build-up and temperature (Hanisak, 1990; Smit, 2002). Thus, although thallus colour is a relatively good indicator of nitrogen status, phycoerythrin catabolism may be regulated by alternative factors and should not be used as the sole means of determining nitrogen deprivation.

The RGR of each of the thallus samples was determined in order to asses the ability of the *G. gracilaria* to grow under the experimental conditions (Fig 2.3). From day 0 (data not shown) to day 16, the RGR of both the nitrogen enriched and nitrogen limited thallus samples increased. After day 16, the RGR of the nitrogen enriched samples remained relatively stable whereas the RGR of the nitrogen limited samples decreased. Nitrogen deprivation experiments on a *Gracilaria secundata* spp., in New Zealand, revealed that both the pigment content and pool of free amino acids increased after exposure to nitrogen limiting conditions (Lignell and Pedersén, 1987). This increase was noted at 17 days post-inoculation after which time the pigment and amino acid concentrations rapidly decreased (Lignell and Pedersén, 1987). Thus in the current investigation the colour change from red-brown at day 0 to almost completely green by day 14 suggested that nitrogen from phycoerythrin pigment was being broken down for use as a nitrogen reserve in nitrogen limited thalli. The decrease in the RGR after day 16 suggests that after

University of Cape Town

this point the nitrogen reserve had become exhausted and was thus unable to sustain a normal growth rate. Contrary to this, the stable RGR of the nitrogen enriched samples suggested that these samples were not nitrogen limited. Furthermore, statistical analysis revealed that there was a significant difference between the RGR of the nitrogen enriched and nitrogen limited thallus samples on days 17, 18, 19 and 20 (Fig 2.3).



University of Cape Town

Fig 2.3 % RGR of nitrogen enriched (▭) and nitrogen limited (▬) *G. gracilis* thalli on days 12 – 20 of the experimental period. Bars represent the average % RGR from two experiments and the vertical lines indicate standard error. * indicates a significant difference ($P < 0.05$) between the nitrogen enriched and nitrogen limited thallus samples on the day of sampling, determined using the Mann-Whitney Rank Sum test.

Thallus decolouration and RGR measurements were used as a general means of observing when the *G. gracilis* thalli become nitrogen limited. In order to obtain a more accurate quantitative estimate of the nitrogen content of the *G. gracilis* thalli, CHNS analysis was performed on thalli sampled over the course of the experiment.

2.3.2 Total thallus nitrogen content

CHNS analysis was performed in order to determine more precisely when the nitrogen limited samples become nitrogen deprived during the experimental period. In CHNS analysis the samples are combusted in a heated oxygen-rich environment to release CO_2, H_2O, SO_x and N_2. These products of combustion are then measured and displayed as weight percentage carbon, hydrogen, sulphur and nitrogen, giving an indication of the percentage of these components in the original tissue sample.

Total thallus nitrogen content was thus determined for both the nitrogen enriched and nitrogen limited *G. gracilis* thalli. Although total thallus nitrogen content decreased from day 0 in both sample types, the percentage nitrogen loss in the nitrogen limited thalli following 7 days exposure to the experimental conditions, was far greater and significantly different to that of the nitrogen enriched samples (Fig 2.4). The nitrogen content of the thalli exposed to the nitrogen enriched media decreased over the initial 24 hrs of the experiment, and subsequently remained relatively stable. This adjustment indicated that thalli required an initial 24 hr period to acclimatise to the experimental conditions. As a result, the inclusion of an initial acclimatisation period, whereby thalli

University of Cape Town

are placed in ASW and PES medium for two days prior to initiating the nitrogen deprivation experiment, may aid in preventing the initial loss in nitrogen content observed in this study.

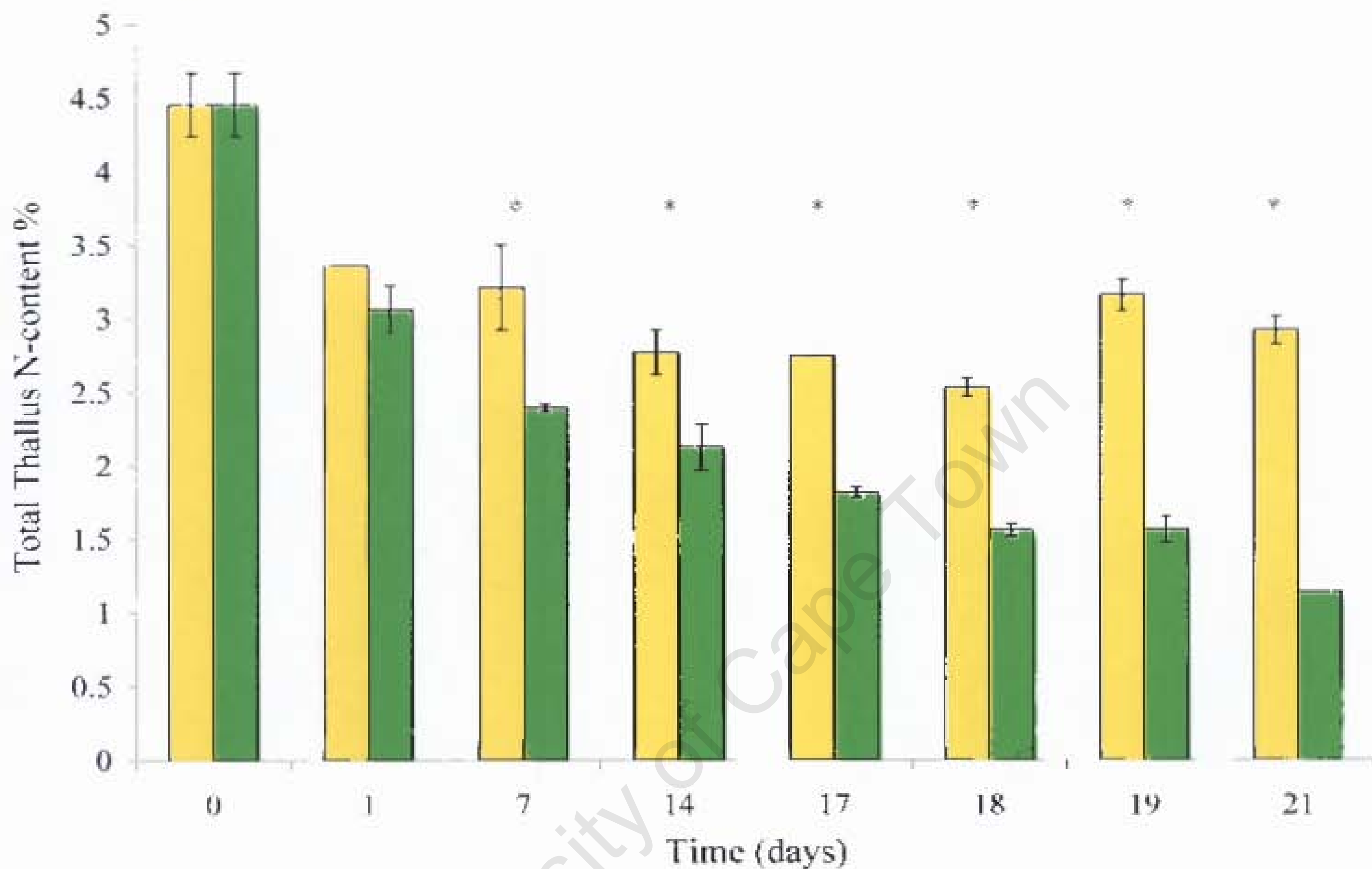


Fig 2.4 Total thallus nitrogen content of nitrogen enriched (▭) and nitrogen limited (▬) *G. gracilis* samples over a 21 day sampling period. Bars represent the average of two experiments and the vertical lines indicate standard error. * indicates a significant difference between the nitrogen enriched and nitrogen limited thallus samples on the day of sampling, determined using a Holm-Sidak multiple comparison method, ($P < 0.05$). (All samples were found to be significantly different from the day 0 sample).

University of Cape Town

Nevertheless, the nitrogen enriched thalli had an average total nitrogen content of 3 % which is considered to be nitrogen replete (Smit *et al.*, 1997). Contrary to this, total nitrogen content of the nitrogen limited thalli decreased over the experimental period with a total nitrogen content of 1.7 % recorded on day 17. Previous documentation suggests that *G. gracilis* is nitrogen limited when the total thallus nitrogen content is between 1.7 – 1.5 % (Smit *et al.*, 1997). Thus *G. gracilis* was considered to be nitrogen deprived after 17 days exposure to the nitrogen limited media.

Analysis of the C:N ratio of *G. gracilis* thalli is an additional means of determining the nitrogen status of the thalli. Numerous studies with *Gracilaria* have shown that as the nitrogen content of the alga decreases, there is an increased flow of photosynthates towards polysaccharide synthesis (Smit *et al.*, 1997, Hanisak, 1990). Thus the C:N ratio increases with increasing nitrogen limitation.

The C:N ratio of the thalli was determined at each sampling point for the nitrogen enriched and nitrogen limited samples (Fig 2.5). A dramatic increase in the C:N ratio was observed over the experimental period in the case of the nitrogen limited thalli suggesting that polysaccharide synthesis was indeed occurring. Although a slight increase in the C:N ratio was noted in the nitrogen enriched thalli, the C:N ratio remained relatively constant overall. Thus the slight increase in the C:N ratio, the decrease in nitrogen content and the slight discolouration of the nitrogen enriched thalli from red-brown to brown, suggests that the available nitrogen supply within the Erlenmeyer flasks may have been sub-optimal. Fresh PES media was supplied every three days and may need to be replaced

University of Cape Town

more regularly. In addition to this, diatom growth occurred in a few of the flasks and may have contributed to the slight loss in nitrogen content in the nitrogen enriched samples.

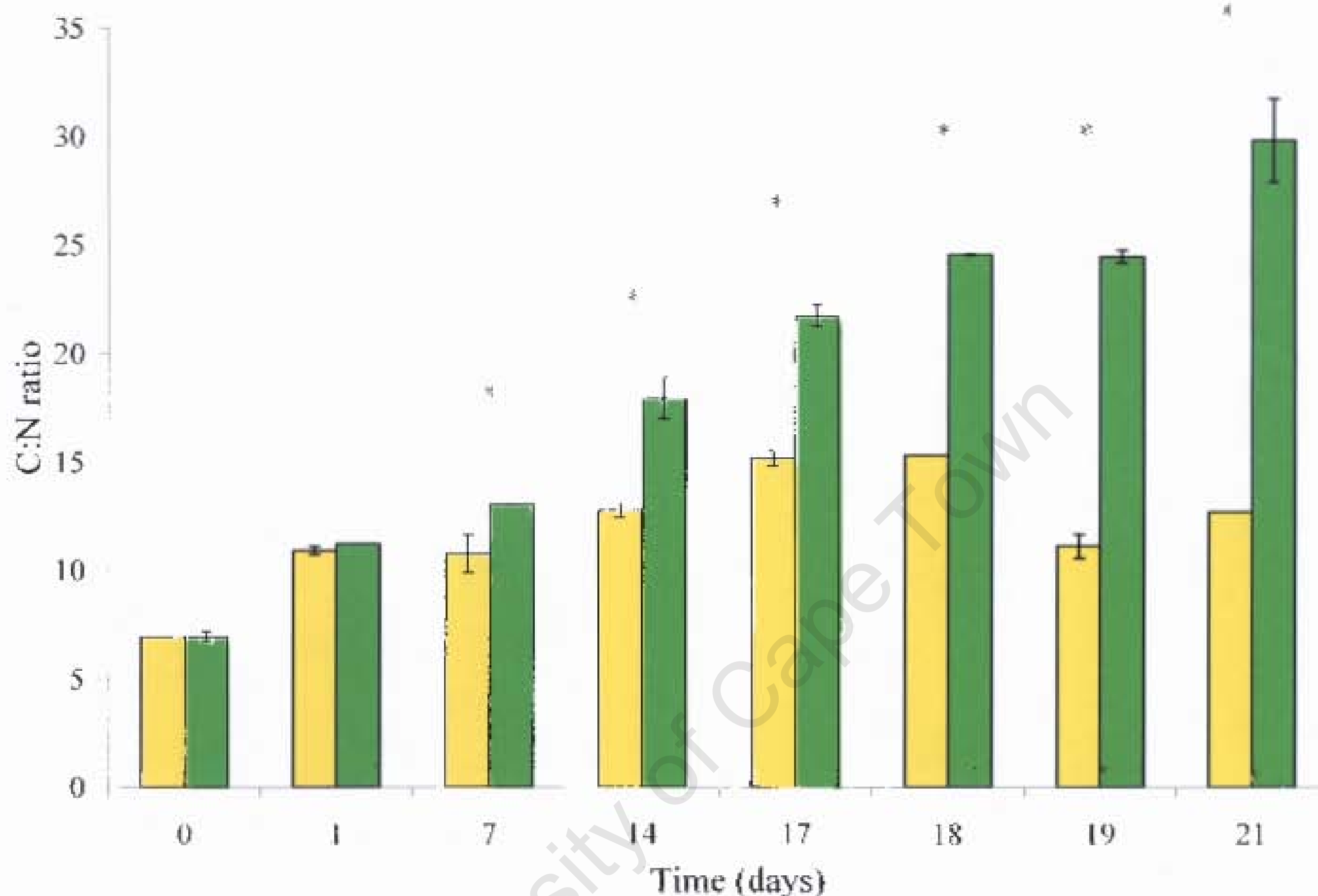


Fig 2.5 C:N ratio of nitrogen limited (■) and nitrogen enriched (□) *G. gracilis* samples over the 21 day sampling period. Bars represent the average of two experiments and the vertical lines indicate standard error. * indicates a significant difference between the nitrogen enriched and nitrogen limited thallus samples on the day of sampling, determined using a Holm-Sidak multiple comparison procedure, ($P < 0.05$).

The change in the nitrogen content of the nitrogen enriched thalli over the entire experimental period indicated that relative to day 0, *G. gracilis* had initiated an adaptive

University of Cape Town

response in order to maintain optimal growth and viability under the experimental conditions applied. Thus, in terms of collecting a reference sample for microarray analysis, it was decided that nitrogen enriched thalli exposed to the same experimental conditions as the nitrogen limited thalli would serve as a superior reference to the *G. gracilis* thalli from day 0 that had not been subjected to the experimental conditions. In this way both the test (nitrogen limited) and reference (nitrogen enriched) thallus samples are exposed to the same experimental conditions and any differences detected in gene expression are more likely to be due to the differences in nitrogen availability and content of the two thallus types.

Overall, *G. gracilis* thalli were sufficiently nitrogen limited after 17 days exposure to the nitrogen limited media. After 21 days exposure to the nitrogen limited media, the thalli showed signs of fragmentation and were not considered optimal for use in microarray analysis. It was thus decided that test and reference samples for microarray analysis would be obtained from day 18 nitrogen limited and nitrogen enriched samples which corresponded to results obtained in earlier experiments (Gebrekiros, 2003).

University of Cape Town

CHAPTER 3

OPTIMISATION OF MICROARRAY PROTOCOLS: RNA ISOLATION, cDNA SYNTHESIS, LABELLING AND HYBRIDISATION

3.1 INTRODUCTION 51

3.2 MATERIALS AND METHODS 53

3.2.1 Sample collection and preparation 53
3.2.2 Glassware, plasticware and reagent preparation 53
3.2.3 Quantitative and qualitative analysis of isolated RNA 54
3.2.4 RNA extraction protocols tested 54
3.2.4.1 Guanidium thiocyanate-phenol-chloroform based RNA isolation (adapted from Chomczynski and Sacchi (1987)) 54
3.2.4.2 Pine tree RNA isolation protocol (Azvedo et al., 2003) 56
3.2.4.3 TRIzol™ based RNA isolation 58
3.2.5 mRNA isolation from TRIzol™ extracted total RNA 59
3.2.6 Aminoallyl-cDNA synthesis and labelling 60
3.2.7 18S rRNA gene amplification 61
3.2.8 RNA amplification, cDNA synthesis and aminoallyl labelling using Ambion's MessageAMP™ II aRNA kit 61
3.2.9 Hybridisation and scanning 62

3.3 RESULTS AND DISCUSSION 62

3.3.1 Optimisation of microarray protocols using freshly obtained, healthy *G. gracilis* thalli 62
3.3.2 Optimization of microarray protocols using nitrogen limited and nitrogen enriched *G. gracilis* thalli 71
3.3.3 Optimization of the Cy-dye labelling reaction for TRIzol™ extracted RNA... 75

University of Cape Town

CHAPTER 3

OPTIMISATION OF MICROARRAY PROTOCOLS: RNA ISOLATION, cDNA SYNTHESIS, LABELLING AND HYBRIDISATION

3.1 INTRODUCTION

The success of a microarray experiment is dependent on the reproducibility of the results obtained following repetitive experimentation (Naderi *et al.*, 2004). Technical variations in the administration of experimental procedures such as RNA isolation, labelling and/or hybridisation, may result in variability of the gene expression detected for a specific target (Simon and Dobbin, 2003; Simon *et al.*, 2003). Additionally, inconsistency can be introduced through the biological variability inherent in the independent biological samples collected (Kerr and Churchill, 2001). Although biological variability may be alleviated in later data normalization processes, prevention of technical variation requires optimization of the various reagents, protocols and procedures used to perform the microarray experiment (Simon *et al.*, 2003). Thus, once aspects of the experimental design have been addressed, optimization of the technical procedures used for microarray analysis is required.

The quality of the RNA obtained for use in a microarray experiment plays a pivotal role in reducing the vulnerability of the technology to systematic variations (Badiee *et al.*, 2003). High quality RNA, as for various other molecular techniques such as the construction of cDNA libraries or RT-PCR, aids in ensuring the success of ensuing

University of Cape Town

cDNA synthesis, labelling and hybridization procedures for the microarray experiment. In addition, many labelling and hybridisation procedures require large amounts of high quality RNA in order for hybridisation to be viable (Naderi *et al.*, 2004).

Extraction of RNA from plant and seaweed tissues in general is particularly complicated due to the presence of polysaccharide and polyphenolic compounds released upon cell-disruption (Chan *et al.*, 2004). These compounds are reported to form tight complexes with isolated RNA, resulting in the formation of gelatinous pellets or viscous solutions (Sharma *et al.*, 2003). In addition, it has been suggested that RNA isolation from seaweed tissue can sometimes be strain and species specific (Chan *et al.*, 2004), and that a single RNA extraction procedure suitable for all plants does not exist (Sharma *et al.*, 2003). Thus, several RNA isolation procedures, previously described for the successful isolation of RNA from plant or seaweed tissues rich in polysaccharides, were attempted in order to determine an effective extraction protocol for the *G. gracilis* strain used in this study.

University of Cape Town

Once an effective RNA extraction protocol was established for both nitrogen enriched and nitrogen limited *G. gracilis* tissue, cDNA synthesis and labelling procedures were evaluated. All labelling reactions performed at the CAPAR facility, U.C.T. utilize an indirect means of incorporating the Cy-dye into the cDNA probe. During cDNA synthesis a unique amino-allyl-dUTP is incorporated to which the Cy-dye is able to associate. Thus, all labelling reactions were optimized using this indirect labelling procedure. Finally, the hybridization procedure was optimised by conducting self against self

hybridizations. High intensity hybridizations with low background signals were considered to be optimal.

3.2 MATERIALS AND METHODS

3.2.1 Sample collection and preparation

G. gracilis thalli (section 2.2.1) were collected from Jacob's Bay abalone farm, Cape Town, South Africa. Deionized water was used to remove sediment and visible epiphytes from the collected plants, and thalli were placed in a 20 L glass tank supplied with seawater. Water movement was achieved by supplying compressed air to the tank through a plastic airline. Temperatures were maintained at 15-17 °C, and illumination was provided by cool white fluorescent tubes at 100 µmol photons m^{-2} s^{-1} with a 16:8 light-dark cycle. Thalli maintained as described above were used initially for the determination of a successful RNA isolation procedure. Once a RNA isolation procedure proved successful, day 18 nitrogen limited and nitrogen enriched thalli (section 2.2.2) were used as a source of *G. gracilis* RNA. All RNA was stored at -70 °C.

University of Cape Town

3.2.2 Glassware, plasticware and reagent preparation

All solutions were treated with 0.1 % Diethylpyrocarbonate (DEPC) water (Appendix A.5.1) and autoclaved. Glassware was treated with chloroform, followed by 100 % ethanol and finally DEPC-water to remove RNases. Glass- and plasticware were additionally covered in foil and autoclaved at 121 °C for 40 min prior to use.

3.2.3 Quantitative and qualitative analysis of isolated RNA

Quantitative analysis of RNA was performed using a Nanodrop spectrophotometer. Absorbance readings were measured at 260 and 280 nm, with one unit of absorption at 260 nm representing 40 μg mL^{-1} of RNA. Ratio measurements at wavelengths 230, 260 and 280 nm were used to determine the purity of the isolated RNA. $A_{260/280}$ and $A_{260/230}$ values of between 1.7 – 2 were considered to indicate reduced protein contamination and low polysaccharide levels in the isolated RNA respectively, and were considered optimal. RNA integrity was determined by 1.2 % formaldehyde-agarose gel electrophoresis. Formaldehyde-agarose gel electrophoresis was performed in 1 X MOPS running buffer (Appendix A.2.4) at 70V as described by Sambrook *et al.* (1989). RNA was subsequently visualised on a 254 nm UV-transilluminator.

University of Cape Town

3.2.4 RNA extraction protocols tested

3.2.4.1 Guanidium Thiocyanate-phenol-chloroform based RNA Isolation (adapted from Chomczynski and Sacchi (1987))

With the aid of a mortar, pestle and liquid nitrogen, 2 g of *G. gracilis* thalli was ground to a fine powder and dispensed into ten 2 ml microfuge tubes (Eppendorf) such that each tube contained approximately 0.2 g of frozen, ground tissue. Lysis solution (Appendix A.5.3), hereafter referred to as Solution D, was aliquoted (1 ml) into each of the microfuge tubes. The samples were homogenized by vortexing for 2 min and returned to ice for a further 2 min before the sequential addition of each RNA extraction solution (below). Thus, 50 μl of 2 M sodium acetate, pH 4, 0.5 ml water-saturated phenol, pH 4, and 0.1 ml chloroform-isoamyl alcohol (49:1 [v/v]) were added to each of the microfuge

tubes. The final suspension was vigorously shaken for 10 s and was left to incubate on ice for a further 15 min.

Following incubation, samples were centrifuged at 10,000 *g* for 20 min at 4 °C. Three phases were obtained after centrifugation, with DNA and proteins present in the inter-phase and phenol phase, while RNA was present in the top aqueous phase. The top aqueous phase was carefully removed, and transferred to ten fresh 2 ml microfuge tubes. Isopropanol (1 volume) was added to each of the microfuge tubes, which were placed at -20 °C and left overnight to allow RNA precipitation to occur.

Sedimentation of the RNA pellet was performed by centrifugation at 10,000 *g* for 20 min at 4 °C. The RNA pellets so obtained were once again resuspended in 0.3 ml of solution D. The RNA suspensions were pooled and dispensed into fresh 1.5 ml microfuge tubes. The RNA was then precipitated at –20 °C for 1 hr with the addition of 1 volume of isopropanol. The microfuge tubes were subsequently centrifuged at 10,000 *g* for 10 min at 4 °C to obtain the precipitated RNA pellets.

The RNA pellets were then washed in 1 ml of 75 % ethanol. Washing was performed by centrifugation at 12, 000 *g* at 4 °C for 5 min, before reversing the pellet position and repeating centrifugation for a further 5 min. Residual ethanol was removed, and pellets were air dried for 15 min in a fumehood. Pellets were then resuspended in 30 µl of DEPC-treated water (Appendix A.5.1). RNA quantitation was performed at this point with the aid of a Nanodrop spectrophotometer.

University of Cape Town

An additional wash step was added in order to eliminate any contaminants that had precipitated together with the RNA. Thus, DEPC-water was added to the resuspended RNA, adjusting the volume to 0.1 ml, and 50 µl 8 M LiCl was added. The solution was incubated overnight at -70 °C to allow for RNA precipitation.

RNA pellets were obtained by centrifugation at 10,000 *g* at 4 °C for 10 min. The pellets were washed with 80 % ethanol and centrifuged at 10,000 *g* for 10 min. The pellets were left to air dry for 5 min before being resuspended in 20 µl of DEPC water. Any insoluble material was removed through centrifugation (1 min), and subsequent transfer of the aqueous solution to fresh microfuge tubes. At this point, the isolated RNA was once again quantitated using the Nanodrop spectrophotometer.

University of Cape Town

3.2.4.2 Pine tree RNA isolation protocol (Azvedo *et al.*, 2003)

G. gracilis thallal tissue (1.3 g) was ground to a fine powder using liquid nitrogen, a mortar and pestle. Ground tissue was added to two 25 ml centrifuge tubes (Beckman) containing 15 ml of extraction buffer (Appendix A.5.4) which had been heated in a 42 °C waterbath for 10 min. The ground tissue mixture was vigorously blended for 5 min by vortexing. The centrifuge tubes were then returned to the 42 °C waterbath, and incubated for 90 min to promote cell lysis.

Following incubation, 15 ml chloroform-isoamyl alcohol (24:1 [v/v]) was added to each centrifuge tube in order to extract the RNA. The mixture was thoroughly blended by vortexing, and subsequently centrifuged at 15,000 *g* for 15 min at 4 °C. The top aqueous

phase containing the RNA was carefully removed and transferred to clean centrifuge tubes. The extraction procedure was repeated with the addition of 1 volume of chloroform-isoamyl alcohol to the RNA solution. Centrifugation was repeated as before at 15,000 *g* for 15 min at 4 °C, and the top aqueous phase was once again recovered and transferred to fresh centrifuge tubes. Precipitation of the RNA was subsequently achieved through the addition of ¼ volume of 10 M LiCl (Appendix A.5.5), followed by incubation at 4 °C overnight.

The precipitated RNA was subsequently collected by centrifugation at 15,000 *g* for 30 min at 4 °C. The RNA pellets were washed in 2 M LiCl, and centrifugation was repeated as before. Following this the RNA pellet was washed in 2 ml of cooled 75 % ethanol and was centrifuged at 10,000 *g* for 10 min. The ethanol was decanted and the RNA pellets were left for 15 min in a fumehood to air dry. The dried pellets were resuspended in 200 µl DEPC-water.

University of Cape Town

In order to remove contaminating polysaccharides, the resuspended RNA was incubated at 65 °C for 10 min followed by immediate transfer to 4 °C for 10 min. Centrifugation at 15,000 *g* for 15 min at 4 °C was performed and the RNA suspension so obtained was gently recovered and transferred to 1.5 ml microfuge tubes. Any additional insoluble material was removed by an additional centrifugation step at 10,000 *g* for 5 min, and the RNA suspension was once again transferred to fresh 1.5 ml microfuge tubes.

3.2.4.3 TRIzol™ based RNA isolation

RNA was isolated according to the protocol described in the TRIzol™ (Life Technologies) catalogue. Briefly, *G. gracilis* tissue (0.6 g) was ground to a very fine powder using liquid nitrogen, and a mortar and pestle. Ground tissue was dispensed into microfuge tubes containing 0.75 ml TRIzol™ reagent. Each sample was vortexed for 5 min to promote the action of the TRIzol™ lysis solution and then incubated at room temperature for a further 5 min.

Following incubation, 0.2 ml chloroform was added to each sample in order retrieve the RNA through phase separation. Samples were incubated at room temperature for 3 min followed by centrifugation for 15 min at 12,000 *g* at 4°C. The upper aqueous phase was removed and the RNA precipitated with 0.5 ml isopropanol. Samples were incubated for 10 min at room temperature followed by centrifugation at 12,000 *g* for 10 min at 4 °C. The supernatant was removed and the pellets were washed in 1.5 ml chilled 75% ethanol by centrifugation at 6,000 *g* for 10 min, reversing the position of the pellet after 5 min. RNA pellets were then air-dried for 5 min and re-suspended in 10 µl DEPC water.

University of Cape Town

Re-suspension of the RNA pellets was aided by heating each sample for 5 min at 55 °C. At this point samples were pooled into a fresh microfuge tube, and any insoluble material was discarded. The pooled RNA was centrifuged for an additional 5 min and once again the RNA solution was transferred to a fresh microfuge tube with any remaining insoluble material discarded. This process was repeated until no further insoluble material was observed following centrifugation.

3.2.5 mRNA isolation from TRIzol™ extracted total RNA

Isolation of mRNA from total RNA was achieved using the Poly A Tract® mRNA Isolation System (Promega). This system makes use of a biotinylated-Oligo dT probe which binds to the mRNA, which in turn binds to streptavidin-paramagnetic particles (SA-PMPs). Briefly, total RNA (25 μg) was brought to a final volume of 500 μl in RNase-free water and denatured at 65 °C for 10 min. Biotinylated-Oligo (dT) probe (1 μg) and 13 μl of a 20X SSC solution were added to the denatured RNA, mixed and incubated at room temperature until the mixture had completely cooled. The mix was then added to a microfuge tube containing 100 μl 0.5X SSC soaked SA-PMP beads and left to incubate at room temperature for 10 min with gentle mixing through inversion every 1-2 min.

Following incubation, the SA-PMP beads were captured using a magnetic stand (Promega) and the supernatant carefully removed and discarded. The SA-PMP beads were then washed a further four times with 0.3 ml of a 0.1X SSC solution. In each case, the beads were gently, but thoroughly, resuspended and subsequently captured using the magnetic stand.

mRNA was finally eluted with the addition of 100 μl RNase-free water. Elution was repeated with a further 150 μl RNase-free water, and the mRNA eluate pooled. The eluted mRNA was vacuum dried to approximately 5 μl, and used directly in a cDNA synthesis reaction.

University of Cape Town

3.2.6 Aminoallyl-cDNA synthesis and labelling

cDNA was synthesised using the ImProm-II™ Reverse Transcription System (Promega). Synthesis was performed as described in the technical manual supplied, however a 5-(3-aminoallyl)-2'deoxyuridine-5'-triphosphate (AA-dUTP) (Sigma) was added in addition to the normal dNTP mixture (Appendix C) at a 2:3 ratio. Following cDNA synthesis, any remaining RNA was hydrolyzed with the addition of 10 µl 1M NaOH and 10 µl 0.5M EDTA. The mixture was incubated at 65 °C for 15 min after which the solution was neutralized with the addition of 10 µl 1M HCl. Unincorporated aa-dUTP and free amines were removed with the use of the RNeasy mini elute kit (Quiagen), and a final volume of 60 µl purified cDNA was collected. The cDNA was vacuum dried to 30 µl and quantified using a Nanodrop spectrophotometer set for single-stranded DNA quantification.

Following quantification, the sample was vacuum dried to completion and the dry aminoallyl-cDNA pellet resuspended in 4.5 µl 0.1M Na_2CO_3 (pH 9) solution (Appendix A.6.1). Cy3 or Cy5 monoreactive dye (Amersham Pharmacia) was added to the cDNA mixture which was wrapped in foil to prevent photobleaching, and left to incubate in the dark at room temperature for 1 hr. In order to enhance labelling, samples were gently mixed every 15 min. Uncoupled dyes were subsequently removed with the RNeasy mini elute kit (Quiagen) and the labelled cDNA collected for use in hybridisation to microarray slides.

University of Cape Town

3.2.7 18S rRNA gene amplification

18S rRNA gene amplification was performed using the 18S rRNA gene primers R18F and E18R (Appendix B.1.2), and 40 ng of synthesized cDNA. A PCR reaction mix (50 µl final volume) was prepared as per Table 3.1, and amplification was performed as described in Appendix B.2.2. The product obtained was visualized on a 1.5 % agarose gel.

Table 3.1 Composition of the reaction mix used for 18S rRNA gene amplification

Reaction Constituents	Volume	Final concentration
Taq polymerase buffer	5 µl	1X
dNTP	1 µl	0.2 mM (of each)
$MgCl_2$	2 µl	2.5 mM
R18F Primer (Appendix B.1.2)	5 µl	50 pmols
E18R Primer (Appendix B.1.2)	3 µl	50 pmols
Supertherm Taq polymerase (Life Technologies)	1 µl	5 U

University of Cape Town

3.2.8 RNA amplification, cDNA synthesis and aminoallyl labelling using Ambion's MessageAMP™ II aRNA kit

RNA amplification, cDNA synthesis and aminoallyl labelling using the MessageAMP™ II aRNA kit (Ambion) was performed according to the manufacturer's instructions (Ambion kit manual).

3.2.9 Hybridisation and scanning

All self-self hybridisations were carried out as described in Chapter 4 (section 4.2.5). Scanning of the hybridised slides was performed directly in order to obtain maximal readings from each slide. Low background contamination and a Cy3, Cy5 overlaid colour image of primarily yellow was considered to indicate optimal hybridization conditions as described by Yu *et al.* (2002).

3.3 RESULTS AND DISCUSSION

3.3.1 Optimisation of microarray protocols using freshly obtained, healthy *G. gracilis* thalli

University of Cape Town

Microarrays are susceptible to numerous variables which can be introduced into an experiment on a number of levels. Biological differences exist between specimens cultured and collected under the same experimental conditions, as well as within samples taken from the same individual. Additionally, differences can exist between RNA samples isolated from the same specimen, between arrays using the same RNA sample (labelling efficiency), and even between replicate spots on the same array (uneven hybridisation) (Simon *et al.*, 2003). Optimisation of each of the technical processes, namely RNA extraction, cDNA synthesis, labelling, hybridization and scanning, were consequently performed in order to reduce sources of potential variability and thereby ensure that the microarray analysis would be reproducible and thus successful.

As previously described (section 3.2.1), all preliminary RNA optimisation procedures were performed using fresh healthy *G. gracilis* thalli as the starting material. Once RNA could be successfully extracted from *G. gracilis*, RNA extraction from the experimental nitrogen limited and nitrogen enriched thalli would be carried out.

In general, RNA extraction from seaweed tissue is particularly difficult due to the high levels of polysaccharide and polyphenolic compounds present in the tissue. Initially, RNA isolations were performed using TRizol™ reagent according to the manufacturer's specifications. Large gelatinous pellets were obtained upon completion of the procedure, suggesting that a large quantity of polysaccharides from the *G. gracilis* tissue had co-precipitated with the RNA during the isolation procedure (Sharma *et al.*, 2003). It was for this reason that any aqueous solution surrounding the gelatinous pellet was removed and centrifuged for an additional period so as to obtain RNA free of contaminating polysaccharides (section 3.2.4.3).

University of Cape Town

Despite the addition of the aforementioned auxiliary recovery step, the isolated RNA had an $A_{260/230}$ ratio that was significantly less than 2.0, indicating that polysaccharide contamination was still prevalent (Table 3.2). Contrary to this, an $A_{260/280}$ value of between 1.7 and 2 was consistently obtained for all TRizol™ isolations suggesting that the RNA was free of contaminating proteins (Table 3.2). RNA yields, although highly variable, were satisfactory, with up to 90 µg RNA being obtained with this RNA isolation method (Table 3.2).

Table 3.2 *G. gracilis* RNA quality, quantity and problems encountered with each of the RNA isolation procedures attempted

Method	$A_{260/280}$	$A_{260/230}$	RNA yield (µg)	µg RNA used to synthesize cDNA	cDNA synthesized (ng)	Problems Encountered
TRIzol™	1.7 – 2	< 1	≤ 90	50	500	Gelatinous Pellet Low $A_{260/230}$ possible polysaccharide contamination Large ammount of RNA required for cDNA synthesis Labelling – unsuccessful
Chomczynski *et al.* (1987)						
Without LiCl wash	1.5 – 2	< 1	40	25	500	Low RNA yield following LiCl wash
With LiCl wash	1.7 – 2	1.5 – 2	10			Labelling – unsuccessful
Azvedo *et al.* (2003)	2	2	186	25	500 – 700	Isolation from control samples –successful Labelling of control samples - successful Isolation from experimental nitrogen limited and nitro enriched samples was unsuccessful

University of Cape Town

RNA integrity was determined by formaldehyde agarose gel electrophoresis (Fig 3.1 A). Distinct 28S and 18S rRNA bands were obtained suggesting that the RNA was intact and had not degraded. Thus although the $A_{260/230}$ ratio was very low, both the $A_{260/280}$ ratio and the RNA integrity itself were thought to be of suitable quality for subsequent use in cDNA synthesis and labelling reactions.

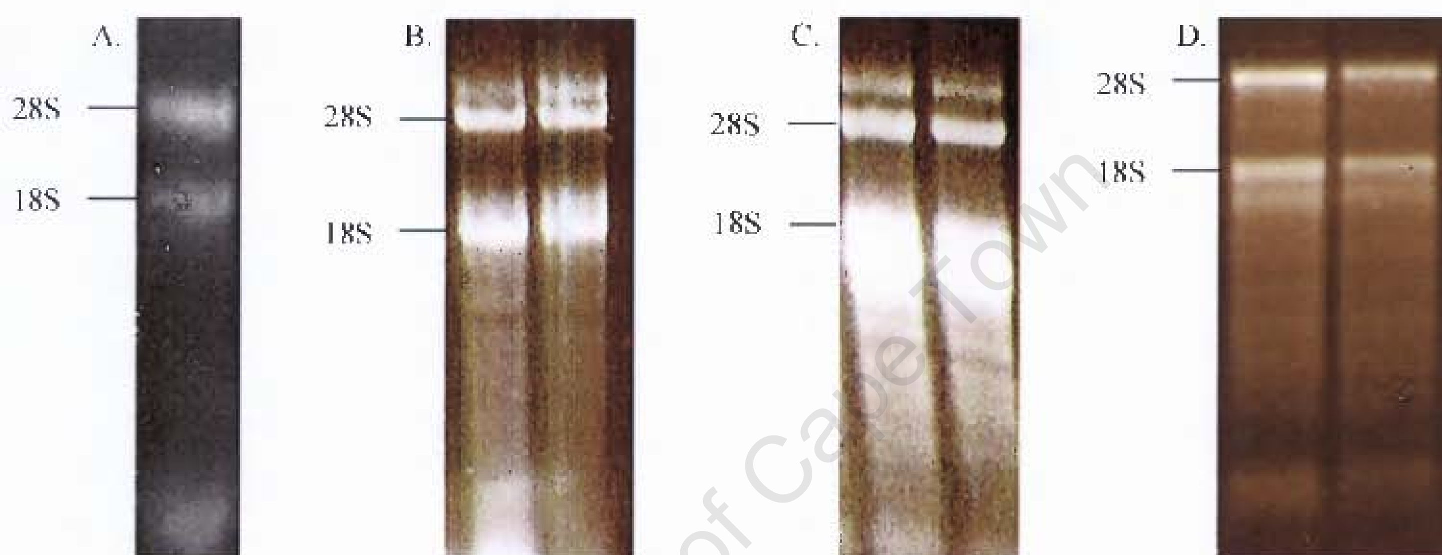


Fig 3.1. Investigation of RNA integrity isolated from healthy *G. gracilis* tissue using various extraction procedures. A) TRIzol™ isolated RNA. B) RNA isolated according to the method described by Chomczynski, *et al.*, (1987) before the addition of a LiCl wash step. C) RNA isolated according to the method described by Chomczynski, *et al.*, (1987) after the addition of a LiCl wash step. D) Pine Tree isolated RNA.

University of Cape Town

cDNA synthesis and labelling was attempted using the TRIzol™ isolated RNA. A minimum of 500 ng cDNA is required in order to perform the indirect labelling procedure according to the protocol utilized by CAPAR (section 3.2.6). Thus in order to achieve a sufficient quantity of cDNA for labelling, a starting concentration of 50 μg TRizol™ extracted RNA was required. Synthesised cDNA was purified and quantitated using the Nanodrop spectrophotometer with yields of 500 ng recorded (Table 3.2). However, all attempts to label the TRIzol™ derived cDNA, with either of the Cy flourecent dyes, was unsuccessful.

In general, RNeasy column purification results in the exclusion of short cDNA products, and thus limited cDNA recovery would occur with low quality RNA samples (Naderi *et al.*, 2004). In the current investigation, cDNA was column purified both before and after labelling, and in each case cDNA was retrieved. This suggests that the original starting material (TRIzol™ isolated RNA) was of suitable quality, and that the cDNA products synthesized from this RNA were of a suitable length for use in labelling. As a result, it was thought that although cDNA was synthesized, traces of contaminating proteins, DNA, other cellular material or even various reagents such as phenol, ethanol, or salts from the RNA extraction procedure itself may have remained associated with the nucleic acids and consequently prevented efficient labelling. The low $A_{260/230}$ readings for the TRIzol™ isolated RNA recorded following Nanodrop spectrophotometry support this hypothesis, introducing the possibility that potential polysaccharide contamination was hindering the labelling procedure.

University of Cape Town

Several protocols for RNA isolation from tissues rich in polysaccharides and polyphenolics were thus sought and attempted (Guena *et al*., 1998; Zeng *et al*., 2002; Sharma *et al*., 2003; Chomczynski and Sacchi, 1987; Azvedo *et al*., 2003). Although many of the procedures tried had been reported to work on other plant or seaweed tissues, the majority failed to retrieve RNA of sufficient quality or quantity from *G. gracilis*. Successful RNA isolation from seaweed tissues, however, has been reported to be strain and species specific (Chan *et al*., 2004), thus requiring optimisation. Of all the procedures tested, RNA of both high quality and quantity was only extracted using the protocols described by Chomczynski and Sacchi (1987), and Azvedo *et al*. (2003) (Table 3.2).

The Chomczynski and Sacchi (1987) protocol utilizes guanidinium thiocyanate and chloride in the extraction solution to effectively denature proteins and inhibit ribonucleases. In general guanidinium based extraction procedures have become the method of choice for RNA purification (Chomczynski and Sacchi, 1987), with the TRIzol™ reagent itself described by the manufacturers as 'an improvement to the single step RNA isolation method developed by Chomczynski and Sacchi'.

Isolation of RNA according to the Chomczynski and Sacchi (1986) protocol was found to be promising, with yields of up to 40 μg RNA obtained (Table 3.2). Nanodrop spectrophotometer readings of 1.5-2 for the $A_{260/280}$ ratios revealed that the isolated RNA was relatively free of contaminating proteins, while $A_{260/230}$ ratios of well below 1 indicated that possible polysaccharide contaminants may have been present (Table 3.2). Additionally, although intact 28S and 18S rRNA bands were visualized upon

University of Cape Town

formaldehyde gel electrophoresis, slight smearing of the RNA bands was apparent, suggesting that the isolated RNA may have contained possible contaminants (Fig 3.1 B). A slight adjustment to the original protocol with the inclusion of an additional LiCl precipitation step was therefore attempted.

Following the addition of the LiCl step, RNA yields decreased significantly from an average of 40 μg to only 10 μg (Table 3.2). However, RNA purity increased dramatically with the $A_{260/230}$ ratio increasing from well below 1 to above 1 and in some cases reaching a value of 2 (Table 3.2). Additionally, the clarity of the 28S and 18S rRNA bands improved (Fig 3.1 C). Thus the inclusion of the additional LiCl precipitation step improved both the purity and the quality of the purified RNA. The significant decrease in total RNA yield is thus explainable by virtue of the fact that contaminants present in the original RNA extracts would have contributed to the overall spectrophotometric quantification of the RNA, resulting in an over-estimate of the actual RNA content. Thus, the addition of the LiCl precipitation step not only aided in the isolation of 'cleaner' RNA, but also allowed for accurate quantification of the purified RNA.

The low yields of RNA obtained following the additional LiCl precipitation step meant that RNA extracts needed to be pooled in order to be sufficient for cDNA synthesis and subsequent labelling reactions. Starting concentrations of 25 μg RNA were used to synthesize the required 500 ng cDNA with greater yields occasionally obtained (Table 3.2). However, even though cDNA synthesis was successful, Cy3 and Cy5 labelling of the synthesized cDNA was unsuccessful.

University of Cape Town

The high purity of the RNA used for cDNA synthesis suggested that it was unlikely that contaminating polysaccharides were inhibiting the labelling reaction. In addition, the recovery of large amounts of cDNA both before and after labelling (data not shown) suggested that although labelling was not successful, the modified dUTP had been incorporated during the cDNA synthesis reaction. Thus, coupling of the modified amino-allyl group with the Cy-dye was in some way hindered. The common reagents used in the extraction solutions for the Chomczynski and Sacchi (1987) and TRIzol™ protocols were investigated as possible causative agents preventing the successful labelling of the synthesized cDNA. The hypothesis was addressed and an alternative RNA extraction protocol, devoid of guanidinium thiocynate and phenol, was sought in order to determine whether these reagents may have inhibited efficient labelling.

University of Cape Town

A hexacetyltrimethyl ammonium bromide (CTAB) based extraction procedure described by Azvedo *et al.* (2003), referred to as the 'Pine Tree' protocol, was attempted. In addition to the cell disrupting activity of CTAB, this protocol utilizes Proteinase K and polyvinylpolypyrrolidone (PVPP) to aid in the removal of contaminating proteins and phenolic compounds, respectively.

RNA extracted according to this protocol was found to be of unrivalled quality, quantity and integrity in comparison to any of the protocols previously employed. Nanodrop quantification revealed that from a single RNA isolation, a total yield of 186 μg of RNA could be achieved (Table 3.2). Overall, varying yields were obtained from individual isolations; however, in each case yields were substantial.

RNA quality was assessed by formaldehyde-agarose gel electrophoresis. The presence of clear distinct 28S and 18S rRNA bands indicated that the RNA was intact and of high quality (Fig 3.1 D). In addition, consistent $A_{260/280}$ and $A_{260/230}$ ratios of greater than or equal to 2 indicated that the RNA was free of any contaminating compounds (Table 3.2).

cDNA conversion and labelling of this RNA were thus attempted. cDNA yields of between 500 – 700 ng were typically obtained. As with the RNA extracted according to the Chomczynski and Sacchi (1987) protocol, only 25 µg of RNA was required to produce over 500 ng cDNA. Thus, 500 ng of this synthesized cDNA was used in a labelling reaction with the Cy5 flourescent dye label.

Cy5 fluorescent dye is known to quench proximal dye molecules and therefore has a lower labelling and fluorescence signal intensity to that of the Cy3-dye (t Hoen *et al.*, 2003). As a result Cy5 was chosen for optimizing the labelling procedure. The labelling reaction was successful, with the frequency of coupling of the Cy5-dye to the aminoallyl dUTP in the synthesized cDNA occurring at 8.58 nucleotides/dye molecule. The Institute of Genomic Research (TIGR) protocol (Hegde *et al.*, 2000) employed by CAPAR, states that a ratio of less than 50 nucleotides/dye molecule is optimal for hybridization, however no definition has been given as to how this value was experimentally determined (Guo *et al.*, 2004). Although the frequency of incorporation of 8.58 nucleotides/dye molecule was considered acceptable, it was decided that two cDNA samples (25 µg RNA per sample) would be combined before labelling so as to obtain a greater signal intensity following hybridization.

University of Cape Town

In order to determine whether the hybridization protocol was optimal for microarray analysis or whether modifications were required, a trial self-against-self hybridization was performed. RNA was isolated according to the 'Pine Tree' protocol, divided into 2 aliquots (25 µg each), converted to cDNA and labelled with Cy3 or Cy5 dyes respectively. Since the RNA for both dye-labelled samples was identical, optimal hybridization should ideally produce similar intensities in both channels for every spot. Thus following hybridization and scanning, photomultiplier tube settings (PMTs) for each channel were adjusted, and a Cy3, Cy5 overlaid false colour image of primarily yellow was produced (data not shown).

Overall, low background levels were detected, and the first hybridisation was considered successful. Consequently, it was felt that protocols for RNA isolation, cDNA synthesis, labelling and hybridisation were optimal and that hybridization with experimental samples rather than freshly obtained, healthy *G. gracilis* thalli could be implemented.

University of Cape Town

3.3.2 Optimization of microarray protocols using nitrogen limited and nitrogen enriched *G. gracilis* thalli.

RNA isolation was performed according to the 'Pine Tree' protocol using frozen day 18 nitrogen enriched and nitrogen limited samples collected as per Chapter 2 (section 2.2.2). Unexpectedly, RNA extraction from both the nitrogen limited and nitrogen enriched samples was unsuccessful, with little or no RNA obtained from either sample.

As previously described (section 3.2.1), technical optimization was performed on freshly obtained healthy *G. gracilis* thalli. Thus, the use of frozen rather than fresh thallal tissue was considered as a possible source of variation from the original optimization procedure. Storage and freezing of the thallal samples may have allowed for the degrading action of RNases, resulting in the failure to isolate RNA. In order to address this, fresh day 18 nitrogen limited and nitrogen enriched samples were collected (section 2.2.2) for direct use in RNA extraction. In addition to the experimental samples collected, control samples of fresh *G. gracilis* thalli were collected simultaneously and used for RNA extraction.

RNA isolation from the control samples was successful. However, little or no RNA was obtained from the nitrogen enriched and nitrogen limited samples. Possible sources of variation between the control and experimental samples were thus assessed. During culture of the nitrogen limited and nitrogen enriched samples, diatom build-up and contamination within the Erlenmeyer flasks was observed. This was apparent regardless of the regular thallal rinsing regime followed each time the media was replaced (section 2.2.2). A new nitrogen deprivation experiment was established according to the procedure specified in Chapter 2 (section 2.2.2), except that germanium oxide (GeO_2) (w/v 4.47 mg/l) was included in both the nitrogen enriched and nitrogen limited media.

University of Cape Town

Germanium oxide (GeO_2) is a known control agent of diatom growth that is commonly used in Rhodophyta culture (Markham and Hagmeier, 1982). It acts by specifically inhibiting cell division in diatoms by blocking cell wall formation (Markham and Hagmeier, 1982). Use of this control agent in the nitrogen deprivation experiments was

not originally considered due to the unknown effect of GeO_2 on *G. gracilis* gene induction and repression. Thus in order to eliminate any possible inducers of a genetic response other than those resulting from nitrogen limitation, GeO_2 was not initially included in the culture media.

The inclusion of GeO_2 in the culture media, although substantially reducing diatom growth, did not result in the successful isolation of RNA from the experimental *G. gracilis* samples. Thus diatom growth was not responsible for the different efficiencies in RNA isolation between control and experimental samples.

It has been well documented that as total thallus nitrogen content decreases, polysaccharide synthesis is enhanced and consequently the polysaccharide content within the cell wall is increased (Smit *et al*., 1996). Analagous results were obtained for *G. gracilis* exposed to nitrogen limited media over the 21 day sampling period in the current investigation (section 2.3.1). The C:N ratios of both the nitrogen limited and nitrogen enriched thalli increased over the experimental period indicating that internal nitrogen was being utilized by the seaweed and polysaccharides were being synthesized. The increase in the C:N ratio of nitrogen limited thalli was far greater than that of the nitrogen enriched thalli. Nevertheless, thalli exposed to the nitrogen enriched and nitrogen limited conditions had an increased carbon content relative to fresh healthy seaweed that had not been exposed to the experimental conditions. Thus a possible explanation for the inability to isolate RNA from nitrogen limited and nitrogen enriched experimental samples is that perhaps the increased level of polysaccharides in the thallus

University of Cape Town

tissue may interfere with the action of the extraction buffer, preventing efficient and successful RNA isolation.

The efficiency of the extraction buffer has been shown by Azvedo *et al.* (2003) to be sensitive to the amount of starting material used, with the integrity of the RNA decreasing as the amount of starting material increases. Thus, although the starting material was kept constant, it is possible that changes in the polysaccharide content of the tissue may have negatively affected the efficiency of the extraction buffer.

In addition to the increased polysaccharide content, the level of ribonucleases within plant tissues have been found to increase during senescence, wounding, and pathogen attack (Shimizua *et al.*, 2001; LeBrasseur *et al.*, 2002; Green, 1994; Hugot *et* al., 2002). Thus, ribonuclease levels may be elevated in nitrogen limited and nitrogen enriched experimental samples relative to the fresh healthy thallus samples used in the optimization procedures.

In order to determine whether RNA isolation from the nitrogen limited and nitrogen enriched experimental samples was at all possible, RNA isolation according to the Chomczynski and Sacchi (1987) and TRIzol™ protocols was attempted. RNA was successfully isolated from both nitrogen limited and nitrogen enriched samples when using these protocols. However, the yield of RNA obtained from the Chomczynski and Sacchi (1987) procedure was very low. Acceptable and consistent RNA yields were obtained for both the nitrogen enriched and nitrogen limited samples following the

University of Cape Town

TRIzol™ isolation protocol. Thus, although the RNA purity as determined by the $A_{260/230}$ value for TRIzol™ extracted RNA had previously been very low, and although previous attempts to label RNA extracted according to this protocol had failed, the ability to isolate acceptable yields of RNA from both nitrogen enriched as well as nitrogen limited samples meant that this procedure could be used to obtain samples for microarray analysis. Thus, emphasis was placed on optimizing the labelling procedure for TRIzol™ extracted *G. gracilis* RNA.

3.3.3 Optimization of the Cy-dye labelling reaction for TRIzol™ extracted RNA

Failure to label cDNA generated from TRIzol™ extracted RNA was believed to be due to interfering contaminants such as proteins, DNA, other cellular material or phenol, ethanol and salts associated with the RNA isolation procedure. In addition, another potential source inhibiting the labelling reaction may be transcript size, as short cDNA transcripts would incorporate less of the modified amino-allyl dUTP base.

University of Cape Town

In order to asses the effects of RNA purity on labelling, and to determine whether transcripts of substantial length were synthesized during the cDNA synthesis reaction, various purification protocols together with 18S rRNA gene amplification were performed. TRIzol™ extracted RNA was purified using either RNeasy mini-elute column purification (Quiagen), mRNA isolation (Promega), DNase treatment (Promega) (Appendix C.1.5) or a combination of column purification followed by mRNA isolation. Improvements in all purified RNA samples were observed, with $A_{260/230}$ ratios and cDNA

yields increasing (data not shown). The improved $A_{260/230}$ ratios suggested that the cDNA was of a superior quality to that used in earlier labelling reactions.

The 18S rRNA gene was amplified from the synthesized cDNA using the R18F and E18R primers (Appendix B.1.2). An expected 1.7 kb fragment was obtained following PCR (Fig 3.2). Thus cDNA transcripts of a relatively large size were successfully synthesised from the TRIzol™ extracted RNA indicating that the modified dUTP is likely to have been incorporated during cDNA synthesis and should be available for Cy-dye labelling.

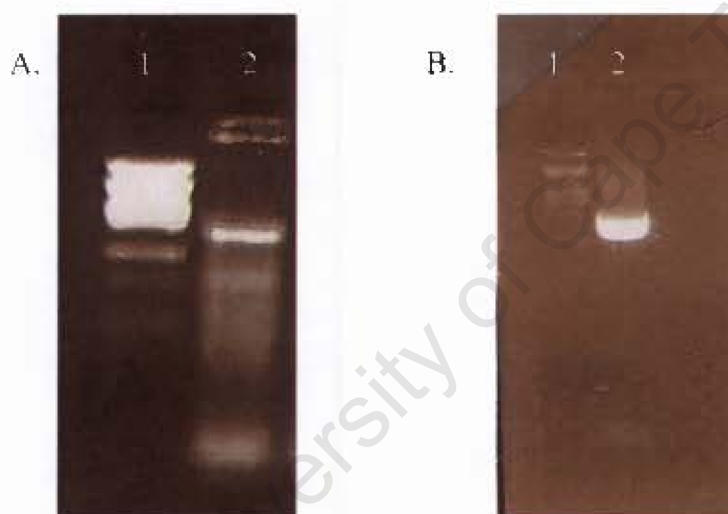


Fig 3.2 Formaldehyde-agarose gel electrophoresis of the 18S rRNA gene amplified from cDNA synthesized from: A) mRNA extracted from total TRIzol™ isolated *G. gracilis* RNA, and B) RNeasy mini-elute column purified TRIzol™ extracted *G. gracilis* RNA. Lane 1, Lambda DNA molecular weight marker; Lane 2, 1.7 kb 18S rRNA gene amplification product.

University of Cape Town

Although 18S rRNA gene amplification proved that cDNA transcripts of substantial length were synthesized, it additionally allowed for visual assessment of the cDNA quality. When 18S rRNA gene amplification was performed using cDNA derived from mRNA, the desired 1.7 kb fragment was obtained, but smearing and the presence of residue within the loading wells was visible following formaldehyde-agarose gel electrophoresis (Fig 3.2A). This suggested that contaminants were present that may prevent efficient labelling. 18S rRNA gene amplification of cDNA derived from RNeasy mini-elute column purified TRIzol™ isolated RNA however produced a very clean, distinct 1.7 kb fragment when viewed under UV transillumination (Fig 3.2.B). As a result, TRIzol™ isolated *G. gracilis* RNA purified in this way was believed to be free of contaminants.

Labelling of the purified RNA however was unsuccessful. As a result, an alternative approach for labelling the TRIzol™ isolated *G. gracilis* RNA was sought. Ambion's MessageAMP™ II aRNA kit was purchased and used together with the TRIzol™ isolated *G. gracilis* RNA. It was believed that perhaps amplification of the mRNA present in the total RNA extract would allow for the extraction of a 'cleaner' RNA product for labelling.

Ambion's MessageAmp™ II aRNA kit was thus used together with 1 μg of TRIzol™ isolated RNA from each of the experimental *G. gracilis* samples. In this case, labelling of the amplified RNA (aRNA) proved successful (section 4.3.3) with the retrieval of labelled aRNA from both the nitrogen limited and nitrogen enriched *G. gracilis* thalli.

University of Cape Town

It was thus decided that microarray analysis would utilise RNA isolated from nitrogen limited and nitrogen enriched samples according to the TRIzol™ extraction protocol, and that the RNA so obtained would be amplified and labelled with the aid of the Ambion MessageAMP™ II aRNA kit. The labelled product was of suitable quantity and quality for hybridisation and thus optimization was thought to be complete. Further optimisation of the RNA isolation protocol may be beneficial in the future in order to ensure consistent yields of RNA and cDNA synthesis, but owing to time limitations the aforementioned protocol was believed to be optimal for subsequent procedures.

University of Cape Town

CHAPTER 4

MICROARRAY HYBRIDISATION

4.1 INTRODUCTION 80
4.2 MATERIALS AND METHODS 82

4.2.1 Array construction 82
4.2.1.1 Origin of targets used for microarray 82
4.2.1.2 Transformation into E. coli DH5α 83
4.2.1.3 PCR amplification of cDNA inserts 84
4.2.1.4 Purification of PCR products 85
4.2.1.5 Array design and synthesis 87
4.2.2 Sample acquisition 88
4.2.3 TRIzol™ (Invitrogen Life Technologies) RNA isolation 89
4.2.4 RNA amplification, cDNA synthesis and aminoallyl labelling of each set of biological repeats 89
4.2.5 Hybridisation 91
4.2.5.1 Probe hybridisation mixture 91
4.2.5.2 Preparation of cover-slips for microarray hybridisation 91
4.2.5.3 Prehybridisation of microarray slides 91
4.2.5.4 Probe hybridisation 92
4.2.6 Scanning of hybridised microarray slides 93
4.2.7 Processing of scanned microarray images 94
4.2.7.1 Design and generation of a GenePix array list file (GAL) 94
4.2.7.2 Extraction of foreground intensity pixels 95
4.2.7.3 Background correction 95
4.2.7.4 Automated and visual quality control 96
4.2.8 Statistical analysis 96

4.3 RESULTS AND DISCUSSION 97

4.3.1 Synthesis of target cDNA and construction of nitrogen limited *G. gracilis* array 97
4.3.2 Sample preparation and RNA Isolation 100
4.4.3 RNA amplification, cDNA Synthesis, labelling and hybridisation 102
4.4.4 Image Analysis 106
4.4.5 Normalisation and statistical analysis 110

University of Cape Town

CHAPTER 4

MICROARRAY HYBRIDISATION

4.1 INTRODUCTION

cDNA microarrays provide a means for rapid, parallel investigation of the expression of thousands of genes in a single assay. Thus, to manage the unprecedented volume of data obtained, microarray experiments require a vast deal of planning. In order to plan or design the microarray experiment however, the overall objective must be defined. The objective of the current microarray investigation was to identify genes that are differentially expressed in *G. gracilis* in response to nitrogen limitation.

Once the objective of a microarray experiment has been stipulated, various aspects of the experimental design can be addressed. Attentive design aids in reducing the various systematic biases that occur when performing a microarray experiment, while making sure that optimal use is made of the resources available (Kerr and Churchill, 2001; Simon and Dobbin, 2003). Design issues therefore address aspects such as which samples are to be used for hybridisation, how these samples should be labelled, on how many arrays the samples should be hybridised, as well as the design of the DNA microarray itself.

As commercial *G. gracilis* microarrays are unavailable, it was necessary to design a microarray for *G. gracilis*. As a result, cDNA fragments from a previously established

University of Cape Town

nitrogen limited *G. gracilis* cDNA library (Gebrekiros, 2003) were amplified and used for printing the *G. gracilis* microarray.

Due to the high costs involved in performing a microarray experiment, it was decided that only three hybridisations would be performed. RNA from three nitrogen limited and three nitrogen enriched *G. gracilis* thalli was extracted and converted to cDNA using Ambion's MessageAMP™ II aRNA kit. The nitrogen limited and nitrogen enriched samples were subsequently paired for use in microarray hybridisation. Each paired group was referred to as a biological repeat.

In order to account for any dye-bias that may result due to the inherent characteristics of the cyanine (CY) dye label, a 'dye-swop' replicate was included. As a result, the nitrogen enriched samples from biological repeat 1 and 2 were labelled with the red Cy5 dye, while the nitrogen limited samples were labelled with the green Cy3 dye. Biological repeat 3 served as the 'dye-swop' replicate, with the nitrogen enriched sample labelled with the green Cy3 dye and the nitrogen limited sample with the red Cy5 dye.

Once all aspects of the microarray design were addressed, the biological samples were hybridised to the *G. gracilis* microarrays. The raw intensity data obtained from the scanned images of the hybridised slides was graphically represented in various diagnostic plots and the relationship between dye-bias and intensity was determined.

University of Cape Town

The raw data was then normalised using "Robust-spline" normalisation, which is an empirical Bayes compromise between "Print-tip loess" and "Global loess" normalisation. "Robust-spline" normalisation uses five parameter regression splines instead of the loess curves to correct the measured level of expression (M-values) for both sub-array spatial variation and for intensity-based trends (Smyth and Speed, 2003). The variation in dye intensities (A-values) between arrays was then corrected using "Aquantile" normalisation. "Aquantile" normalisation adjusts the A-values of all the arrays so that they have the same empirical distribution, while leaving the M-values unchanged.

An empirical Bayes *t*-test was used to compare the expression levels between the samples. A list of the top 100 differentially expressed genes was obtained, with differential expression having statistical merit if the calculated *P*-value was less than 0.05 after Bonferonni adjustments were made and/or if a two-fold induction (M = 1) or one-half repression (M = -1) in the measured level of expression (M-values) was observed. As a result, those genes found to be differentially expressed in *G. gracilis* in response to nitrogen limitation were identified and included in further analyses.

University of Cape Town

4.2 MATERIALS AND METHODS

4.2.1 Array construction

4.2.1.1 Origin of targets used for microarray

cDNA clones containing cDNA fragments from nitrogen limited or diseased *G. gracilis* were used as the targets for microarray. These cDNA clones were obtained from previously established un-normalised cDNA libraries (Gebrekiros, 2003; Iyer,

unpublished), constructed using the Creator™ SMART™ cDNA Library Construction kit (CLONTECH Laboratories, Inc.). Briefly, synthesised cDNA was purified and digested with *Sfi*I restriction endonuclease. The restriction digests were electrophoresed through a 1.1% agarose gel, and fragments of 1.2 – 2.8 kb were excised, purified and ligated into *Sfi*I-digested, dephosphorylated pDNR-LIB vectors (CLONTECH) (Appendix D). The ligated products were subsequently stored as DNA at - 70°C.

The cDNA fragments derived from diseased *G. gracilis* were transformed (4.2.1.2), PCR amplified (4.2.1.3) and purified (4.2.1.4) by a fellow colleague, Christopher Ealand, as part of his studies. cDNA fragments derived from nitrogen limited and diseased *G. gracilis* were amalgamated onto a single microarray slide in order to increase the number of *G. gracilis* cDNA targets on the array. In addition, amalgamation of these cDNA targets on the array enhances the possibility of detecting genes that are differentially expressed in both nitrogen limited and diseased *G. gracilis*.

University of Cape Town

4.2.1.2 Transformation into E. coli DH5α

As the previously established nitrogen limited *G. gracilis* cDNA library was stored as DNA, transformation of the ligated products into *Escherichia coli* cells was required. pDNR-LIB vectors (Appendix D) containing 1.2-2.8 kb fragment inserts were therefore transformed into *E. coli* DH5α cells made competent with $CaCl_2$ using the method described by Draper *et al.* (1988) (Appendix C.1.2). Transformed colonies were randomly picked with sterile toothpicks and inoculated onto fresh Luria agar (LA) (Appendix A.1.6) plates supplemented with 30 μg/ml chloramphenicol (Cm) (Appendix

A.4.1). Each plate contained a grid in which 30 individual colonies could be stabbed and labelled. A total of 1 200 colonies were transferred to these plates, which were incubated overnight at 37 °C and maintained for a fortnight at 4°C before transfer to fresh LA plates supplemented with 30 µg/ml Cm.

In addition, 50 % glycerol stocks of each of the 1,200 colonies were prepared. Transformed colonies were inoculated into sterile 96 well plates (NUNC™) containing 100 µl Luria broth (LB) (Appendix A.1.5) supplemented with 30 µg/ml Cm, and were incubated overnight on a shaker at 37 °C. A duplicate 96 well plate was inoculated with 50 µl of the overnight cultured colonies. Glycerol stocks were prepared by adding a further 50 µl of a 50 % glycerol solution to each 96 well plate. These plates were then stored at -70 °C until required.

University of Cape Town

4.2.1.3 PCR amplification of cDNA inserts

Transformed colonies were picked from LA plates supplemented with 30 µg/ml Cm (4.2.1.2) and inoculated into sterile 96 well round-bottomed plates (NUNC™) containing 100 µl of LB supplemented with 30 µg/ml Cm. A total of 95 wells were inoculated with transformants, while the remaining well was left un-inoculated as a PCR control. The cultures were incubated on a shaker (100 rpm) at 37°C. The O/N cultures were then centrifuged at 1 500 rpm before 5 µl of each culture was transferred into sterile 96 well PCR plates (COSTAR). The PCR plates were then sealed with a plastic cover and heated at 96 °C for 15 min. A PCR master mix (final volume 100 µl) (Table 4.1) was concurrently prepared, and pre-heated at 96°C for 15 minutes, before 98 µl of this master

mix was added to the 5 µl of freshly cultured transformants. The PCR plates were centrifuged at 1 500 rpm for 5 min and placed into a PCR machine (TECHE GENIUS) preheated at 96 °C. The cDNA fragments were subsequently amplified using the cycling conditions stipulated in Appendix B.2.1. The amplified cDNA fragments were then visualized on a 1 % agarose (TAE) gel following electrophoresis at 60 V for 45 min.

Table 4.1 Composition of the reaction mix used for colony PCR.

Reaction Constituents	Volume	Final concentration
Primers (Appendix B.1.1)	3 µl	0.3 µM
dNTP	0.4 µl	40 µM
$MgCl_2$	12 µl	3.0 mM
Taq Polymerase buffer	10 µl	1 X
Southern Taq polymerase	0.4 µl	2 U

University of Cape Town

4.2.1.4 Purification of PCR products

In order to remove unincorporated amino acids, primers, enzymes and salts, cDNA fragments were transferred to 96 well Millipore plates (Millipore, MultiScreen™ PCR). The Millipore plates were placed on a vacuum apparatus (Millipore, Vacuum/Pressure Pump) and a vacuum applied. All unwanted dNTPs and reagents were filtered and the amplified cDNA fragments were transferred to the Millipore membrane. The vacuum was released and the Millipore membrane at the bottom of each well of the Millipore plate

was washed with 100 μl of sterile Milli Q water. The water was removed by vacuum and the wash procedure repeated in order to remove all remaining unbound dNTPs and DNA from each of the membranes.

The amplified cDNA fragments bound to the membrane were subsequently re-suspended in 50 μl of 50 % DMSO and left on a shaker at 300 rpm for 2 hrs to allow for complete re-suspension. Following re-suspension, the 50 μl of purified PCR product was divided such that 20 μl was transferred to a 384 well printing plate (GENETIX) destined for use in printing the microarray, while the remaining 30 μl was placed in a second 384 well plate to be used for quantification and electrophoresis of purified products. All products were stored at -20 °C.

The purified products were visualized as in 4.2.1.3. Quantification of each product was performed using the Gene Tools software (version 3.06). Gene Tools applies an algorithm to each product on the gel based on the spectrophotometric quantification of at least 3 PCR products included on the gel. Thus a few PCR clones were quantitated on a Nanodrop spectrophotometer and used to determine the concentrations of the remaining PCR clones. A data file of all 1044 purified PCR products was established detailing the identification number, PCR status and concentration of each PCR product. In addition, the position of each product in each of the 96 well culture, PCR and printing plates was recorded. The data was later used for the generation of a GAL file required for microarray image analysis (see section 4.2.7.1) and as a means of locating the differentially expressed clones for further analysis.

University of Cape Town

4.2.1.5 Array design and synthesis

All slides were printed by CAPAR, U.C.T. in a single print run using a Micro Grid Biorobotics printer. Pins were set up in a 4 x 4 arrangement, allowing 16 different inventory wells containing amplified cDNA fragments to be visited, loaded and printed in a single source visit. In total, 144 source visits were performed, with each array printed in duplicate on a slide, and each product printed in duplicate adjacent to each other within an array (Fig 4.1). In total, 2 * 1044 cDNA's from the *G. gracilis* Nitrogen limited library and 2 * 588 cDNA's from the *G. gracilis* diseased library were printed on an array. As a result, 3264 *G. gracilis* cDNA fragments were printed on an array, and thus 6528 on a slide. In addition to the amplified *G. gracilis* cDNA fragments, 1152 Lucidea Universal ScoreCard controls (Amersham) were printed (576 per array) to serve as normalisation controls on the slide. Thus, 7680 'target' spots were printed in total onto each slide. Following printing, each slide was UV cross-linked at 120 mJ to ensure immobilization of the cDNA 'targets' and was stored in a dark sealed container.

University of Cape Town

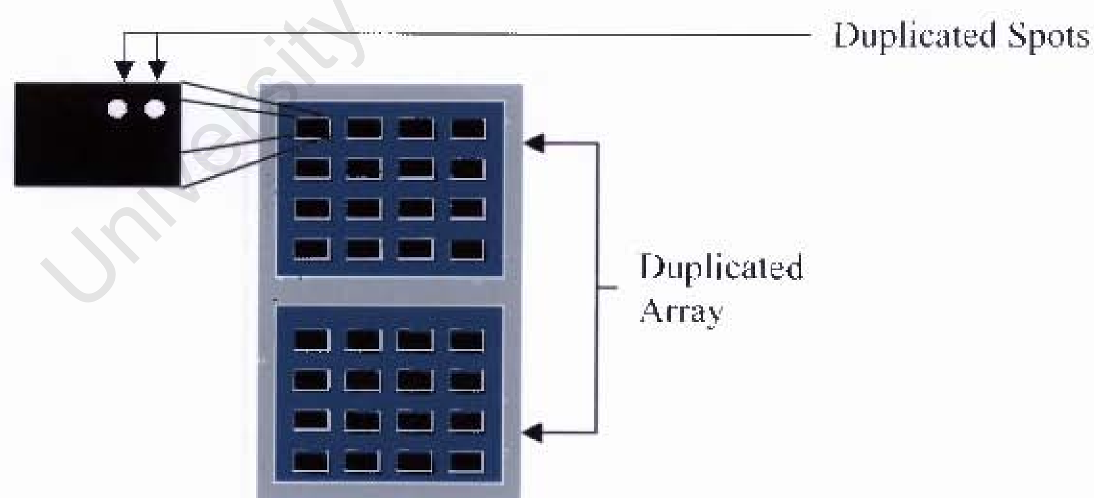


Fig 4.1 *G. gracilis* microarray slide layout showing duplication of the entire array on the slide as well as duplication of the cDNA target within an array.

All data pertaining to each cDNA target printed on the array was stored in an Excel file. In addition, to allow for easy location of the original cDNA clone used to synthesise each cDNA target on the array, each cDNA clone was assigned a unique identity (ID). The ID assigned to each cDNA clone was based on the test species, the cDNA library from which the 'target' spots were derived, and lastly the position of the clone in the 50 % glycerol stock 96 well culture plate. Thus an example of the unique ID used in the current investigation is: Gg_NL_05B02, where Gg represents *G. gracilis*, NL represents the nitrogen limited cDNA library from which the cDNA clones were derived, and 05B02 is the plate number and position on the 50 % glycerol plate where the clone containing the cDNA fragment can be found.

University of Cape Town

4.2.2 Sample acquisition

G. gracilis thalli weighing 2.0 ± 0.05 g fresh weight were transferred to six 1 L Erlenmeyer flasks containing 250 ml artificial seawater (ASW) (Appendix A.1.1) with a salinity of 3.6 %. Erlenmeyer flasks were divided such that three flasks contained control, nitrogen enriched thallus samples, while the remaining three flasks housed experimental, nitrogen limited thallus samples. The ASW in the control and experimental flasks was supplemented with either PES (1/3 strength) (Appendix A.1.4.1) or PES-N media (Appendix A.1.4.2), respectively. GEO_2 was added (4.47 mg/L) to both the experimental and control flasks to prevent diatom growth. Experimental conditions identical to those used in the nitrogen deprivation experiments (Chapter 2, section 2.2.2) were applied. All samples were removed after 18 days and flash frozen until required for RNA extraction.

In total, three sets of biological repeats were prepared for miroarray analysis. Each set of biological repeats consisted of both a nitrogen enriched and nitrogen limited sample from the above experiment.

4.2.3 TRIzol™ (Invitrogen Life Technologies) RNA isolation

RNA was isolated according to the protocol described in the TRIzol™ (Life Technologies) catalogue (section 3.2.4.3). Isolated RNA was purified using an RNeasy mini elute kit (Quiagen). RNA purity and integrity were evaluated by absorbance at 260 and 280 nm, and by denaturing formaldehyde agarose gel electrophoresis. For each of the subsequent RNA amplification reactions, 1 μg of the Total RNA was used.

University of Cape Town

4.2.4 RNA amplification, cDNA synthesis and aminoallyl labelling of each set of biological repeats

RNA amplification, cDNA synthesis and aminoallyl labelling of the *G. gracilis* RNA and the Lucidea Universal ScoreCard controls was performed using Ambion's Message AMP™ II aRNA kit as stipulated in the user manual. Ambion's online master mix volume calculator was used to determine the volume of reagents required in the preparation of master mixes for the various reactions (http://www.ambion.com/techlib/ap pend/mm_calcs/msgamp2_96_mm_calc.php).

Labelled aRNA was diluted 1:10 with nuclease-free water and quantified using a Nanodrop spectrophotometer set to read microarray samples. Promega's online labelled

aRNA calculator was used to determine the frequency of dye incorporation which is the number of Cy-labelled nucleotides incorporated per 1 000 nucleotides of cDNA (http://www.prontosystems.com/technical_support/calculator/index.asp).

These values were checked manually by first calculating the total picomoles (pmol) of aRNA synthesized using the following equation:

$$\text{Pmol nucleotides} = [OD_{260} \times \text{volume } (\mu l) \times 37 \text{ ng}/\mu l \times 1000 \text{ pg/ng}] / 324.5 \text{ pg/pmol}$$

Where: 1 OD_{260} = 37 ng/μl for cDNA and the average molecular weight of a dNTP = 324.5 pg/pmol.

The frequency of dye incorporation was then calculated using:

$$\text{pmol Cy3} = OD_{550} \times \text{volume } (\mu l)/0.15$$

$$\text{pmol Cy5} = OD_{650} \times \text{volume } (\mu l)/0.25$$

$$\text{nucleotides/dye ratio} = \text{pmol cDNA/pmol Cy dye}$$

Overall, a dye incorporation of greater than 200 pmol per sample and a ratio of less than 50 nucleotides/dye molecule is optimal for microarray hybridisations (Hedge *et al*., 2000).

University of Cape Town

4.2.5 Hybridisation

4.2.5.1 Probe hybridisation mixture

A probe hybridisation mixture (Appendix A.6.9) was prepared by mixing 1 μg of each labelled Cy3- and Cy5-cDNA together. The blocking reagents, mouse COT1-DNA (Life Technologies) and Poly(A)-DNA (Sigma), were added to the labelled cDNA mixture and the volume adjusted to 30 μl. An equal volume (30 μl) of hybridisation buffer (Appendix A.6.5) was added and the mixture left until required for hybridisation.

4.2.5.2 Preparation of cover-slips for microarray hybridisation

The cover-slips (22 mm X 60 mm) (Erie Scientific company) for use in hybridisation were submerged in 100 % acetone for 1 hr at room temperature on a LASEC bench rotational shaker. The cover-slips were then washed in 0.2 % SDS for 10 min followed by a further two washes for 10 min in MilliQ water. The cover-slips were then placed in a pre-heated 42 °C oven and left to dry.

University of Cape Town

4.2.5.3 Prehybridisation of microarray slides

Microarray slides were pre-hybridised in 60 μl of pre-hybridisation buffer (Appendix A.6.4) before use. A diamond marker was used to outline the edges of the array in order to indicate where the printed array commenced and terminated. The slides were then placed into a hybridisation chamber (Arraylt™) and a cover-slip was gently placed onto each slide. Pre-hybridisation buffer was injected underneath the cover-slip,

and dispersed over the entire array. The hybridisation chamber was sealed and incubated for 2 hrs in a pre-heated container of water in a 42 °C oven.

The slides were then washed and the cover-slip removed by gentle submersion in a series of five separate containers of MilliQ-Plus (Millipore) water. The slides were dipped four to five times before being transferred with forceps to the next container. The slides were then immersed in isopropanol for 1 s and promptly dried through centrifugation at 1000 rpm for 5 min. Slides were used immediately to ensure optimal hybridisation efficiency.

4.2.5.4 Probe hybridisation

The microarray hybridisation area was prepared by gently lowering a pre-cleaned glass cover-slip (section 4.2.5.2) over the array area of a pre-hybridised slide (section 4.2.5.3) and placing the slide into a hybridisation chamber. The probe cDNA in hybridisation mixture (section 4.2.5.1) was denatured at 95 °C for 3 min, snap-cooled on ice for 30 s, and briefly centrifuged. The cDNA probe (60 µl) was immediately injected under the one corner of the cover-slip, allowing the solution to wick along the length of the array. To maintain humidity inside the chamber, 10 µl of MilliQ water was added to each of the reservoir wells at either end of the hybridisation chamber. The chamber was then tightly sealed and placed into a container of water pre-heated to 42 °C. The water container housing the hybridisation chamber was covered in foil and placed in a 42 °C oven for 16-20 hrs.

University of Cape Town

Following hybridisation, the slides were removed from the chamber and washed sequentially with increasing stringency wash buffers as follows. First the cover-slips were removed by submerging the slide in a staining dish containing 42 °C pre-heated low stringency wash buffer (Appendix A.6.6). The staining dish was then covered in foil and placed on a rotational shaker (LASEC Lab Rotar) for 5 min. The slides were then removed and placed in a staining dish containing a medium stringency wash buffer (Appendix A.6.7), and were once again left for 5 min on the shaker. This process was repeated by placing the slides in a high stringency wash buffer (Appendix A.6.8), after which the slides were briefly dipped several times in a staining dish containing 100 % ethanol. The slides were immediately dried through centrifugation at 1000 rpm for 5 min and placed in a light-tight slide box until scanned. Scanning was performed directly in order to obtain maximal readings from each slide, after which, the slides were returned to the slide box and stored in the dark at room temperature.

University of Cape Town

4.2.6 Scanning of hybridised microarray slides

Hybridised slides were scanned with a GenePix 4000B scanner (Axon Instruments, Inc.) using appropriate gains on the photomultiplier tube (PMT) to obtain the highest intensity without saturation. The PMT for each channel (Cy3 and Cy5) was thus adjusted within the range of the emission wavelengths of the Cy3 and Cy5 flourescent dyes, which is 510 to 550 nm and 630 to 660 nm, respectively. GenePix Pro software (version 6) was utilized in tandem with the GenePix 4000B scanner in order to acquire and analyse the intensity of fluorescence at each pixel location on each microarray. A 16 bit tagged image

file (TIFF), was then generated in order to store the fluorescence intensities for each channel.

4.2.7 Processing of scanned microarray images

GenePix Pro software was used to obtain an accurate intensity measurement for each probe on an array. This was achieved through four main steps: A GenePix Array List (GAL) file was generated; foreground intensity pixels were extracted; background artefacts were corrected for, and problem targets were flagged and excluded from further analysis.

4.2.7.1 Design and generation of a GenePix array list file (GAL)

During the printing process a data file (section 4.2.1.4) containing all the original positions and unique identities of each probe was established. This file was used to create a GAL file required for automated array analysis. A GAL file consisting of an 8 X 4 grid, with a 15 X 16 block format within each grid, was thus generated according to the placement and spacing of probes on the array during the printing process. This GAL file was superimposed onto the pixels generated for each array image in a manner that isolated each target (spot) within a cell, giving an identity to the target within that cell. GenePix Pro refers to each cell surrounding a target as a *feature-indicator*.

University of Cape Town

4.2.7.2 Extraction of foreground intensity pixels

During printing, substantial variation occurs in the size and shape of the spots within and across arrays. As a result, GenePix Pro's proprietary spot-finding algorithm was used to align the features within the GAL file to their appropriate target areas on an array. The foreground intensities of each spot were then determined using an adaptive circle segmentation method that fits the best diameter for each spot.

4.2.7.3 Background correction

Background intensities due to artifacts such as non-specific binding of labelled probe, comets, dust, as well as auto-fluorescence from target DNA, were corrected for using the GenePix Pro local background correction algorithm. This algorithm works by creating a circle three times the diameter of the circle identifying the foreground region of a target (Yang *et al.*, 2001b). Pixels occurring within the larger outer circle, excluding a two pixel wide region around neighboring target circles, are considered to be background artifacts (Fig 4.2).

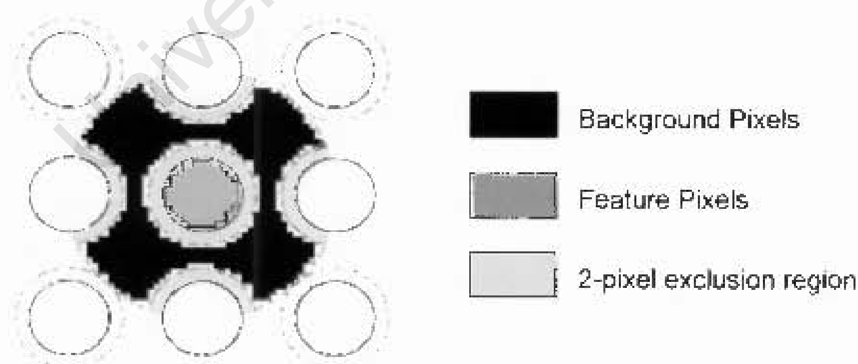


Fig 4.2 Diagram of the local background correction algorithm used by GenePix Pro software (Figure reproduced from GenePix Pro 6 user manual)

University of Cape Town

Thus, the foreground and background intensities measured for each dye within or surrounding each target were determined, and background correction was completed through subtraction of the computed background intensity values from the respective foreground intensity values.

4.2.7.4 Automated and visual quality control

GenePix Pro software automatically flags and excludes spots if the pixel level within the feature indicator of a spot is undetectable or saturated. These spots are marked by the software as "not found". Visually, any targets with high intensities due to anomalies such as dust or scratches on the slide, or targets with weak foreground intensities, were manually flagged as "Bad" and were excluded from further analysis. A results list was then generated, containing information for each target spot on the array.

University of Cape Town

4.2.8 Statistical analysis

Microarray data was analysed using the open source statistical programming environment R, making use of the command driven Limma bioconductor package. The Limma package can be obtained from the biocunductor project site http://www.bioconductor.org. As R is an open source statistical programming environment, the syntax of the commands used to analyse data are refined from time to time. The commands may therefore vary from those used in the current investigation (all relevant updates will be specified in the online Limma manual). All commands used in the current investigation are recorded as an Rscript in Appendix E.

'Targets' and 'SpotTypes' text files were generated to enable Limma to identify the origin of each target spot printed on the array. All intensity data for any undesirable target spot (4.2.7) was thus easily identified and removed. In addition, all intensity data for the Lucidea Universal ScoreCard control targets were removed. Robust-spline and Aquantile normalisation was performed to normalise dye-bias within and across the arrays, respectively.

Normalised gene expression data was summarized by a design matrix of intensity log-ratios $M = \log_2 R/G$, with m rows corresponding to the genes under analysis and $n = n_1 + n_2$ columns corresponding to n_1 control hybridisations and n_2 treatment hybridisations. A linear model was fitted to the data and an empirical Bayes statistical function was applied to compute moderated t and B-statistics. Bonferroni correction was applied to the data to account for the number of standard errors that occur when estimating log-fold changes and a table of the top 100 ranked genes was generated. Differential expression was considered significant if a two-fold induction ($M = 1$) or one-half repression ($M = -1$) in the measured level of expression was observed and if the calculated P-value was less than 0.05.

University of Cape Town

4.3 RESULTS AND DISCUSSION

4.3.1 Synthesis of target cDNA and construction of nitrogen limited *G. gracilis* array

The overall objective of this study was to use microarray technology to identify genes from *G. gracilis* that are differentially expressed under nitrogen limiting conditions. In

order to perform a microarray experiment genes from the organism under question must be printed onto microarray slides in order to act as targets for the labelled probes. In general, arrays exist for "model organisms" such as humans, *Arabidopsis* and mice, where the entire genome has been sequenced, and each gene within the genome has been identified. *G. gracilis* however is a "non-model organism" and thus little is known about its genetic makeup. Thus, the primary objective of this study was to create a microarray containing genes from *G. gracilis*.

Microarray slides were printed using target DNA from a previously established nitrogen limited *G. gracilis* cDNA library. cDNA clones (1044) from the nitrogen limited *G. gracilis* cDNA library were successfully amplified, purified and visualised on a 1 % agarose gel (Fig 4.3). A few cDNA clones containing multiple (Fig 4.3 row 1, lane 6) or no (Fig 4.3 row 2, lane 6) inserts were identified and the 'status' of each product together with its concentration was recorded in an Excel data file for future reference. In order to locate the original cDNA clone used in PCR, all products irrespective of their 'status' were printed as targets on the microarray slides.

University of Cape Town

In any microarray experiment comparisons of interest are made across samples but within genes (Kerr and Churchill, 2001). Therefore, the more times a specific gene is replicated within an array the more chance you have of determining whether a gene is differentially expressed and at the same time decreasing the chance of making a Type 1 error. Essentially, replicated spots act as technical repeats and allow the quality of hybridisation within an array to be assessed. Thus, slides were designed with each target duplicated

next to each other on an array, and each array duplicated on a slide to increase the strength of the statistical analysis for each gene. In total, 7680 target spots were printed on a single microarray slide. Of the 7680 target spots, 3264 were derived from *G. gracilis* amplified products, and 576 were Lucidea Universal ScoreCard controls.

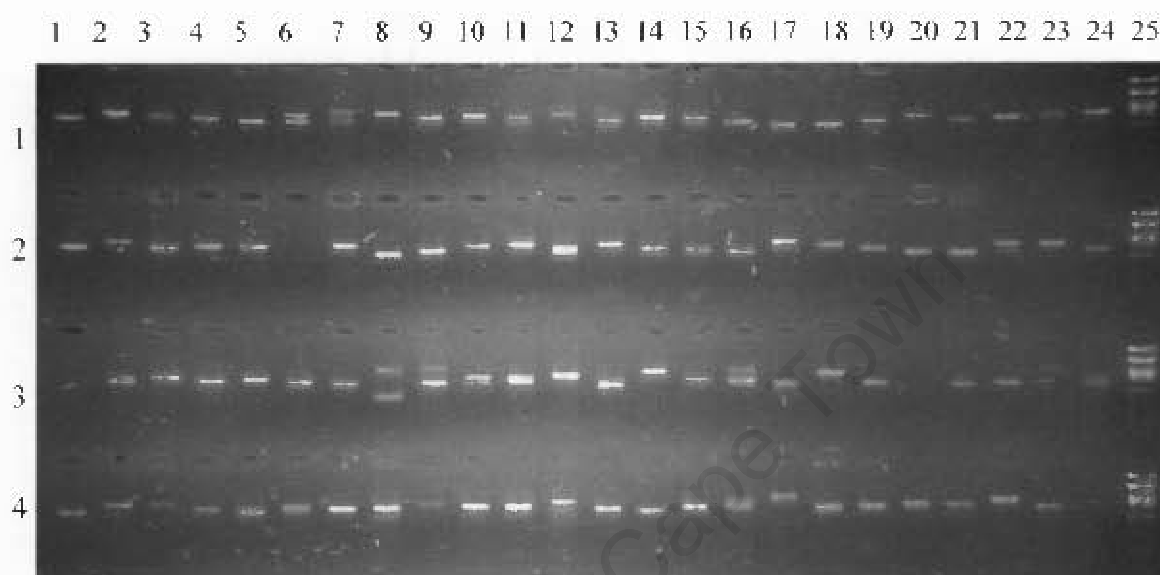


Fig 4.3 Purified PCR products from the nitrogen limited *G. gracilis* cDNA library. Rows 1-4, lanes 1-24, contain the purified PCR products from a single 96 well plate. Row 4, lane 24 contains the PCR control, while row 1-4, lane 25, was loaded with Lambda DNA molecular weight marker.

University of Cape Town

In order to store information pertaining to each target spot printed on the array, as well to locate each target on an array, a unique identity was assigned to each target spot and was stored in the web-based database, BASE. Additionally, following hybridisation, each unique ID was used by the GenePix Pro software to identify each target printed on the array.

4.3.2 Sample preparation and RNA Isolation

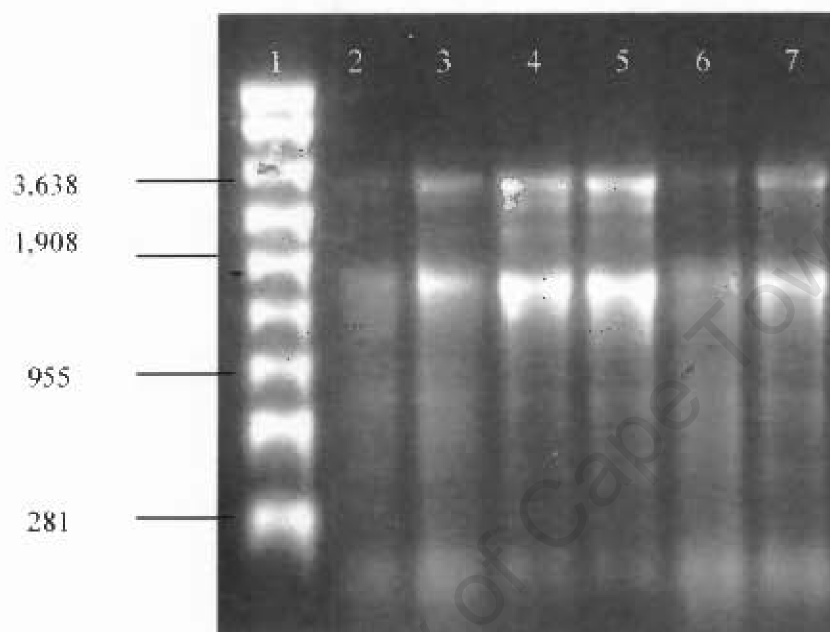
Microarray experiments need to be replicated in order to obtain statistically relevant information regarding gene expression in each sample. Replication can be performed on two levels, technical and biological. Technical replication accounts for variations that result from the administration of the experimental procedures such as RNA isolation, labelling and/or hybridisation. Replicating the microarray experiment with RNA isolated from the same specimens will provide information on the reproducibility of the array itself. Technical replication however, does not take into account biological variability and will not confidently answer the biological question at hand. Thus, replicates of biologically independent specimens are required in order to confirm whether a specific gene is differentially expressed under the experimental conditions. In addition, the greater the number of replicate arrays, the greater the statistical power of comparison of gene expression as well as the ability to detect errors.

University of Cape Town

Table 4.2 Purity of RNA from each sample according to the $A_{260/280}$ and $A_{260/230}$ ratios.

Biological Repeat	Sample	$A_{260/280}$	$A_{260/230}$
1	Nitrogen enriched	2.14	1.63
1	Nitrogen limited	2.15	0.59
2	Nitrogen enriched	2.05	0.38
2	Nitrogen limited	2.15	0.37
3	Nitrogen enriched	2.09	1.23
3	Nitrogen limited	2.08	0.68

Only three hybridisations were performed in the current investigation due to the high costs involved in performing a microarray experiment. Nitrogen limited and nitrogen enriched *G. gracilis* thalli were collected after 18 days exposure to their respective media (Chapter 2) and were paired for use in microarray hybridisation. Each pair so grouped was referred to as a biological repeat.



University of Cape Town

Fig 4.4 Denaturing formaldehyde agarose gel electrophoresis of each RNA sample. Lane 1. RNA marker (Promega), lanes 2 and 3, RNA from nitrogen enriched and nitrogen limited *G. gracilis* thalli comprising biological repeat 1. Lanes 4 and 5, RNA from nitrogen enriched and nitrogen limited *G. gracilis* thalli comprising biological repeat 2. Lanes 6 and 7 RNA from nitrogen enriched and nitrogen limited *G. gracilis* thalli comprising biological repeat 3.

RNA was extracted from each of the three biological repeat samples using the TRIzol™ reagent. RNA quality was determined according to the $A_{260/280}$ and $A_{260/230}$ ratios (Table 4.2), as well as the clarity of the 28S and 18S rRNA bands on a denaturing agarose formaldehyde gel following electrophoresis (Fig 4.4). The $A_{260/280}$ ratios lay within the optimal range of 1.7 – 2 for all samples, suggesting that the RNA was free of contaminating proteins. The $A_{260/230}$ ratios, however, were extremely low for all samples except the nitrogen enriched samples from biological repeat 1 and 3. Generally, low $A_{260/230}$ ratios are indicative of possible polysaccharide contamination. *G. gracilis* polysaccharide content is believed to increase in response to nitrogen limitation (Smit, 2002) and this may account for the low $A_{260/230}$ ratios obtained for the nitrogen limited samples.

RNA integrity, determined according to denaturing formaldehyde agarose gel electrophoresis, revealed that the RNA was essentially intact, despite slight smearing of the bands. Overall, the RNA was felt to be of suitable quality for use in the micoarray experiments.

University of Cape Town

4.4.3 RNA amplification, cDNA Synthesis, labelling and hybridisation

Ambion's MessageAMP™ II aRNA kit was used to amplify and label RNA isolated from each biological sample. Using 1 μg of HeLa total RNA as a control, Ambion have determined that an *in vitro* transcriptase (IVT) reaction of 14 hrs should optimally result in the synthesis of 176 μg amplified RNA (aRNA) (Ambion manual version 0404). After the 14 hr IVT reactions with all three biological samples, the maximal yield of aRNA

FOI was found to be less than or equal to 50 nucleotides/dye molecule and was thus optimal for use in microarray hybridisation.

Table 4.4 Labelled aRNA dye flourescence, Concentration and FOI for each biological replicate.

Biological Repeat	Fluorescent Label	OD_{550}	OD_{650}	Concentration ng/µl	FOI
1 Nitrogen enriched	Red	–	0.163	580.9	43.99
1 Nitrogen limited	Green	0.102	–	481.0	37.99
2 Nitrogen enriched	Red	–	0.224	621.6	46.77
2 Nitrogen limited	Green	0.133	–	658.6	43.69
3 Nitrogen enriched	Green	0.124	–	570.0	50.53
3 Nitrogen limited	Red	–	0.176	479.0	49.0

University of Cape Town

The labelled nitrogen enriched and nitrogen limited samples were paired and used as probes in hybridisation reactions to the target *G. gracilis* cDNA printed on the microarray slides. The hybridised slides were scanned and 16 bit TIFF images of each of the three microarray slides was obtained for each channel (Cy3 and Cy5).

4.4.4 Image Analysis

Each target on the microarray should have a quantity of labelled probe (labelled aRNA) bound to it that is proportional to the level of expression of the target gene. During printing however, substantial variation occurs in the size and shape of the spots within and across arrays. In addition, background intensities due to artifacts such as non-specific binding of labelled probe, comets, dust, as well as auto-fluorescence from target DNA, contribute to the total intensity measurement obtained for each target spot. As a result, a number of data cleaning steps or low-level analyses, referred to as image analysis, is required in order to account for any intensity measurement that is not specifically due to the hybridisation of probe to target.

A few targets spots were omitted from all further analysis on all three microarrays. On the whole, the majority of spots eliminated were spots that had a very weak signal, and thus, negative *R* and *G* values. Such spots are generally considered too faint to show acceptable evidence of differential gene expression, and consequently, the loss of information which results from omitting these spots from future analyses is usually not substantial (Simon *et al.*, 2003a). In total, the maximum percentage of spots excluded on an array was 22.4 %, which was considered acceptable as the GenePix software, using a "median" background, may often exclude 30 % or more negative values (Simon *et al.*, 2003a).

The lucidea control spots had a far higher intensity than that of the *G. gracilis* target spots. This suggested that during the IVT reaction, amplification of lucidea RNA was far

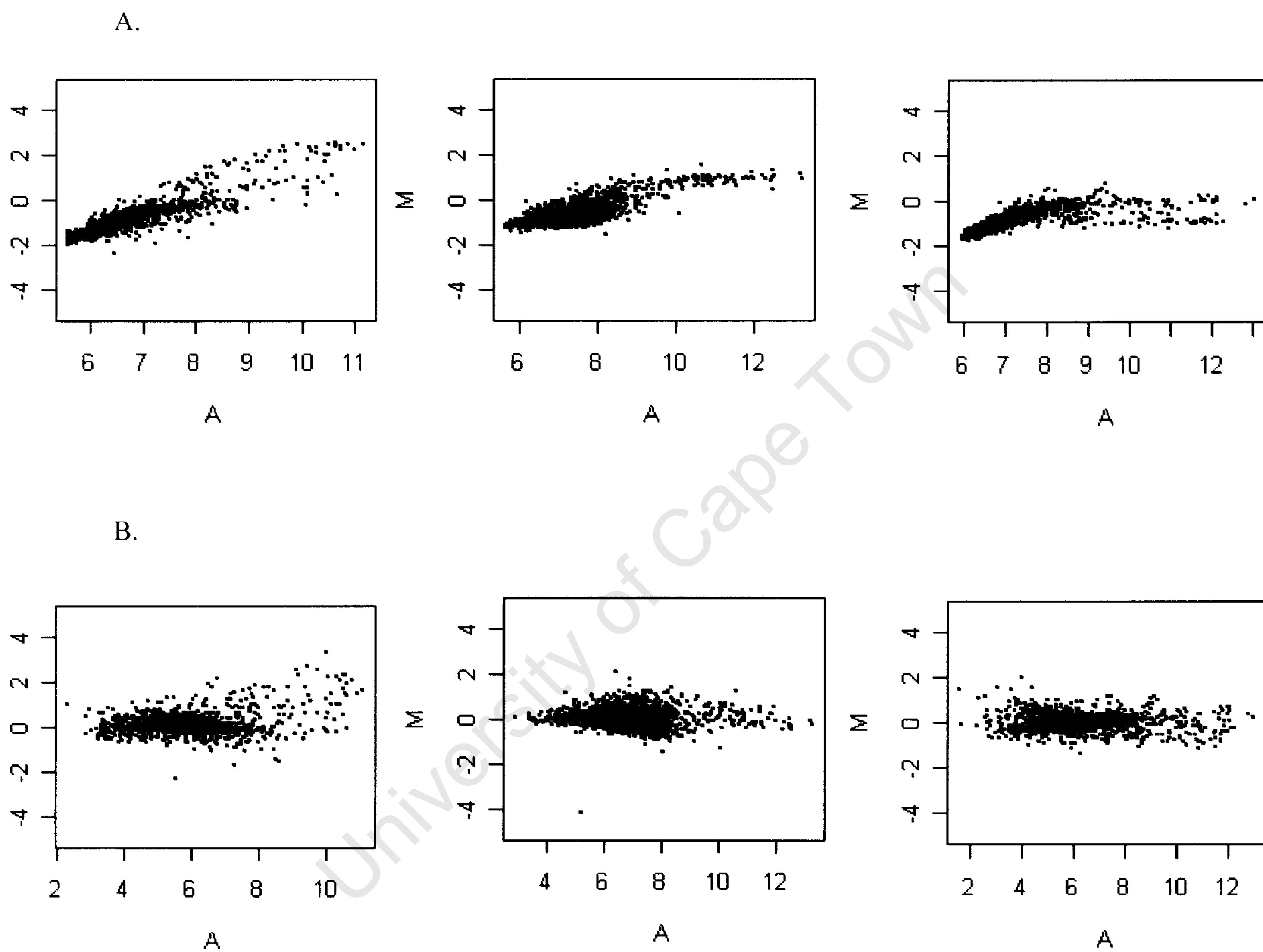
University of Cape Town

superior to that of the *G. gracilis* RNA from each of the samples. The difference in intensity between the *G. gracilis* genes and the lucidea control spots indicated that subsequent normalisation could not be based on the intensity values of the lucidea controls, and that an alternative normalisation procedure was required.

Total foreground and background fluorescence intensity measurements for each dye sample (Cy3 and Cy5) bound to each of the targets on the array was determined, and an output file containing the signal intensity information for each target spot on the array was generated. The Cy3/Cy5 expression ratios for each spot on the array were log_2 transformed to allow for a more informative display of gene expression intensity in terms of an MA plot.

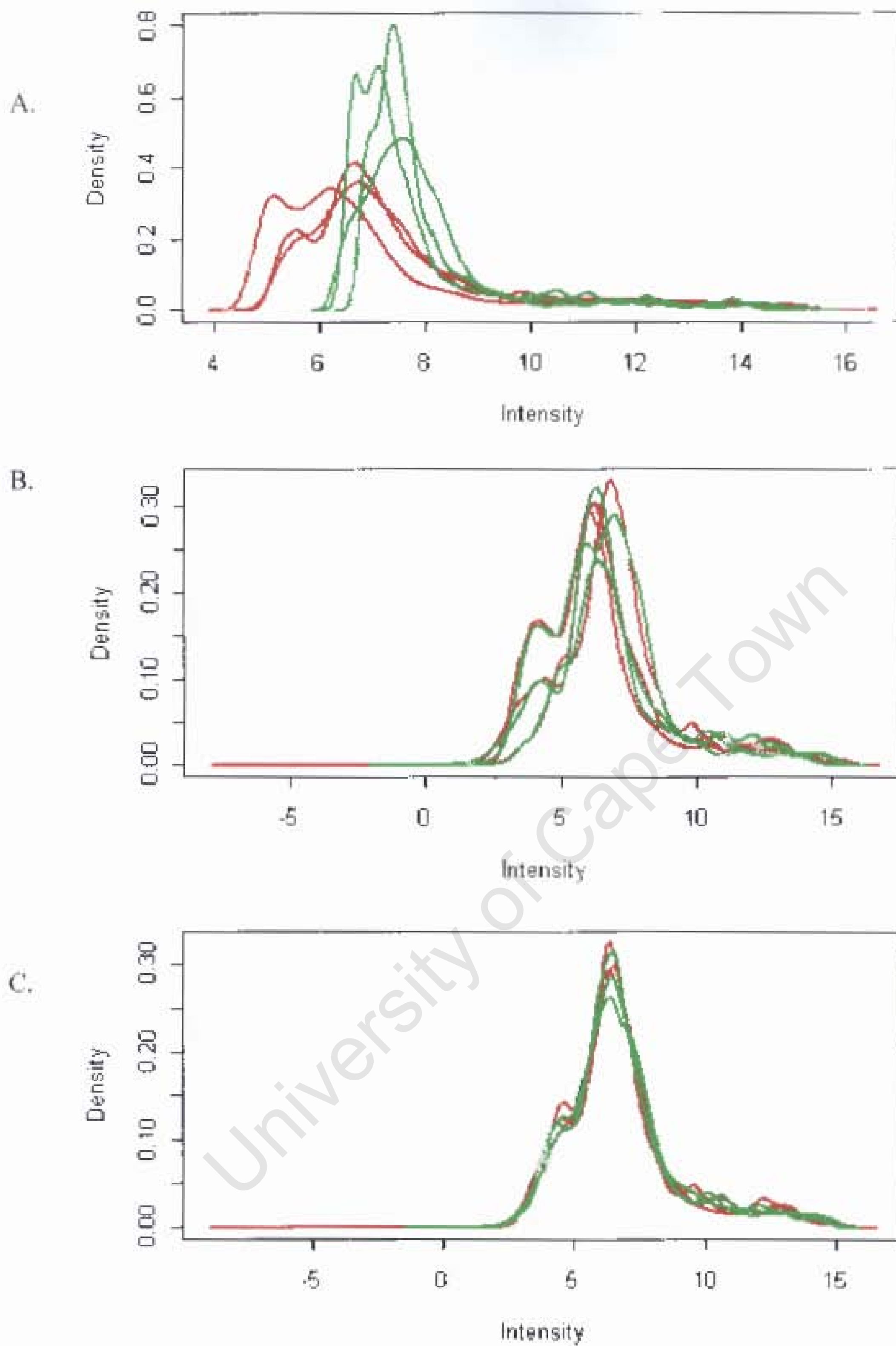
Generally, when two identical mRNA samples are labelled with different dyes it is rare to obtain equal fluorescence intensities for the dyes, and often the intensities for the green Cy3 dye are higher than that of the red Cy5 dye (Yang *et al.*, 2001). In order to determine the relationship between dye-bias and fluorescence intensity in the current investigation, the raw data for each array was graphically represented as an MA-plot (Fig 4.7A). The MA-plots for each array revealed that the M values (i.e. the log_2 Cy3/Cy5 ratio for each gene) tended to lie below zero at lower fluorescence intensities, demonstrating a dye-bias towards the Cy3 dye. This was confirmed by a density plot of the raw data from all three arrays (Fig 4.8A).

University of Cape Town



University of Cape Town

Fig 4.7 MA plots of genes on all three arrays. A, Raw data for arrays 1, 2 & 3. B, Robust-spline normalised arrays 1, 2 & 3.



University of Cape Town

Fig 4.8 Density plots of Cy3 and Cy5 fluorescence intensity data for all three microarrays. A, Raw data before normalization; B. After within-array Robust-spline normalization; C. After between-array Aquantile normalisation.

4.4.5 Normalisation and statistical analysis

Normalisation was performed to balance the fluorescence intensities of the Cy3 and Cy5 dyes, as well as to allow comparison of gene expression levels across the arrays. Robust-spline normalisation, which is an empirical Bayes compromise between print-tip loess and global loess normalisation, was used to normalize the intensity data within the arrays. Print-tip loess normalisation corrects the M-values for both sub-array spatial variation and for intensity-based trends. However, the procedure is unreliable for small arrays, and arrays where a large number of missing M values are detected ("bad spots") (Smyth, 2004), as is the case with the arrays (small) in the current investigation. Alternatively, data can be normalised using either a global loess or robust-spline normalisation algorithm. Since global loess normalisation does not take sub-array variation into account (Smyth and Speed, 2003), robust-spline normalisation was chosen as a means of normalizing the data within the arrays.

Following robust-spline normalisation, the MA plots showed that the dye fluorescence data had been successfully normalised within each array, with the data centred around zero on the Y-axis (log intensity axis) (Fig 4.7B). Similarly, inspection of the density plot generated after robust-spline normalisation clearly shows that the Cy3 and Cy5 dyes within the arrays were aligned, and thus the dye effect within each array had been corrected (Fig 4.8B). However, differences between the dye intensities across the arrays remained evident.

University of Cape Town

Aquantile normalisation was performed to reduce inter-array (between array) variability. A density plot of the aquantile normalised data showed that the Cy5 and Cy3 intensities for all three biological repeat samples overlapped (Fig 4.8C), indicating that the dye intensities of each array were balanced and thus no-one particular array would dominate in further statistical analyses.

For any organism, it is generally assumed that the majority of genes should have a constant level of expression under conditions of biological comparison, and therefore, only a small proportion of genes are expected to be differentially expressed (Yang *et al.*, 2001). As was expected, the majority of genes on the array were not differentially expressed, having M-values of around zero (Fig 4.7B). Although a few outlying genes with a two-fold variation in intensity were visible on all arrays, the identity of these genes cannot be determined on an MA-plot and, in fact, such spots may not represent the same gene across all arrays. Statistical analysis was thus performed in order to determine which genes, if any, were differentially expressed across all three arrays.

As two independent sample types, namely nitrogen limited and nitrogen enriched thalli, were used for the microarray experiment, it was appropriate to employ a standard *t*-test to compare the expression levels between the samples (Dopazo *et al.*, 2001). The limma package uses an empirical Bayes *t*-test to compare the intensity data. This method is believed to result in a more stable inference and improved power for experiments with a

University of Cape Town

small number of arrays (Smyth, 2004), as in the current investigation where only three arrays were performed.

A statistical *P* value is produced after *t*-test analysis which represents the probability that the observed difference in gene expression is due to random chance. In the current investigation, the null hypothesis set for expression analysis was that: no genes are differentially expressed under nitrogen limiting versus nitrogen enriched conditions. Alternatively, the null hypothesis could read: any difference in gene expression observed between the two samples is due to random chance and not differential expression of a gene. Thus, a very small *P* value indicates that the tested gene is likely to be differentially expressed, and the null hypothesis is rejected.

When many hypotheses are tested, as is the case in a microarray experiment, the probability that at least one Type I error is committed can increase sharply with the number of hypotheses (Dudoit *et al.*, 2002). Thus, *P*-values generated by the *t*-test analysis were adjusted to correct for any multiplicity errors using Bonferroni correction.

Following statistical analysis, a list of the top 100 ranked differentially expressed genes was generated (Table 4.5). The list shows the M and A values, *t* and B-statistics, and *P*-value for each of the top 100 ranked genes. The majority of the differentially expressed genes on the array were down-regulated in response to nitrogen limitation, as indicated by a negative M value.

University of Cape Town

Table 4.5 List of the top 100 differentially expressed genes, showing M, A, *t*, *P* and B statistics

Clone ID	cDNA clone number	M	A	*t*	*P*.Value	B
Gg_IB9_05B02	394	-1.79129	10.15276	-7.21343	$1.57E^{-07}$	14.07972
Gg_IB9_01B07	19	-1.32435	8.20327	-6.76102	$1.35E^{-06}$	12.09911
Gg_IB9_06E01	524	-1.4542	10.37871	-6.74947	$1.42E^{-06}$	12.04624
Gg_IB9_06B02*	489	-1.34914	9.614396	-6.61796	$2.67E^{-06}$	11.44614
Gg_IB9_01E10	58	-1.36018	9.666296	-6.5875	$3.08E^{-06}$	11.30783
Gg_IB9_01A09	9	-1.33848	10.89262	-6.47317	$5.29E^{-06}$	10.79081
Gg_IB9_04B04	301	-1.59411	10.2733	-6.48257	$5.20E^{-06}$	10.78006
Gg_IB9_03B05	207	-1.26123	10.98404	-6.07839	$3.32E^{-05}$	9.035053
Gg_IB9_01E06*	54	-1.20424	7.585129	-5.97712	$5.27E^{-05}$	8.592665
Gg_IB9_02E09*	152	-1.43754	5.562388	-5.95367	$5.99E^{-05}$	8.478717
Gg_IB9_06B01	488	-1.20413	10.92696	-5.90633	$7.27E^{-05}$	8.285518
Gg_IB9_05A02	382	-1.40098	8.664451	-5.90051	$7.62E^{-05}$	8.25224
Gg_NL_11B07*	968	1.632634	5.327394	5.799987	0.000118	7.827397
Gg_NL_07C11	604	-1.14631	6.981341	-5.79901	0.000118	7.823203
Gg_IB9_01A10	10	-1.13207	8.100718	-5.75072	0.000147	7.616568
Gg_IB9_01B05	17	-1.39421	9.118133	-5.74252	0.000155	7.584886
Gg_IB9_06F02	537	-1.23786	10.08851	-5.7262	0.000164	7.511961
Gg_IB9_05A01	381	-1.36144	8.117853	-5.70587	0.000183	7.431352
Gg_IB9_01A04*	4	-1.10058	7.623235	-5.70183	0.000182	7.40823
Gg_IB9_05H07*	471	1.163628	5.377472	5.666366	0.000214	7.257659
Gg_NL_11B09	970	-1.13481	9.650547	-5.65588	0.000224	7.213237
Gg_NL_08B09	685	-1.34074	8.817149	-5.59776	0.000295	6.981286
Gg_IB9_01B12	24	-1.05729	7.683456	-5.49514	0.000455	6.53761
Gg_IB9_03G09	271	-1.77904	11.71097	-5.34477	0.000902	5.823666
Gg_IB9_01A03	3	-1.69522	6.976093	-5.09296	0.002633	4.861064
Gg_IB9_01C02	26	-0.96235	7.6311	-5.04035	0.003201	4.684791
Gg_IB9_03C02*	216	-0.97553	8.00258	-4.98936	0.003962	4.482853
Gg_IB9_05F10	450	-1.25217	10.92723	-4.91259	0.005511	4.238788
Gg_NL_03H07	281	1.176665	7.395999	4.893108	0.005972	4.163835
Gg_NL_11D10*	995	-0.92574	5.121229	-4.83573	0.007476	3.881984
Gg_IB9_01A01	1	-0.91914	6.396531	-4.78284	0.009279	3.677794
Gg_IB9_05G12*	464	-1.12416	8.980202	-4.7536	0.010557	3.632463
Gg_IB9_05C01	405	-0.88978	7.632783	-4.65016	0.01586	3.171891
Gg_NL_07A12*	581	1.200957	2.753636	4.596435	0.019834	3.045437
Gg_IB9_01D02*	38	-0.87741	6.88722	-4.60107	0.019296	2.986992
Gg_IB9_05D02	418	0.871987	8.607954	4.545951	0.024014	2.780963
Gg_NL_02H02	181	0.879121	8.824442	4.543484	0.024249	2.77178
Gg_NL_11F11*	1020	-1.0615	6.652567	-4.50457	0.028511	2.708182
Gg_IB9_05D06	422	-0.85959	7.52329	-4.47103	0.032248	2.503534
Gg_IB9_03D02	228	-0.86968	10.79922	-4.18199	0.097805	1.462629
Gg_IB9_01D06	42	-0.7907	6.408516	-4.14718	0.111441	1.34055
Gg_NL_11C06	979	-0.79734	6.002865	-4.08025	0.142961	1.10782
Gg_IB9_03G10	272	-0.92935	6.40473	-3.90487	0.27264	0.621876
Gg_IB9_02H06	185	-0.91213	5.33638	-3.89433	0.283166	0.587093
Gg_IB9_03E07	245	-0.74133	8.362325	-3.8522	0.327597	0.335536
Gg_IB9_03E05	243	-0.89608	10.26166	-3.80717	0.386299	0.302222
Gg_IB9_03F03	253	-0.88927	7.813154	-3.7961	0.401717	0.266368
Gg_IB9_06C02	501	-0.72868	8.047721	-3.81739	0.370793	0.220541
Gg_IB9_06G07	554	-0.75464	9.907704	-3.77121	0.436529	0.069168
Gg_NL_04D02	323	-0.75776	6.910464	-3.74002	0.487032	-0.03226
Gg_IB9_05H02	466	-0.86561	7.505588	-3.68897	0.584366	-0.07646

University of Cape Town

Table 4.5 Continued

Gg_IB9_05G10	462	-0.75685	10.96751	-3.7185	0.525072	-0.10189
Gg_IB9_05E04	432	-0.72135	6.395511	-3.69331	0.573174	-0.18299
Gg_IB9_03E02	240	-0.70652	9.610885	-3.69197	0.575863	-0.18732
Gg_IB9_06G03	550	-0.71298	8.76566	-3.68013	0.59999	-0.22528
Gg_IB9_01D12	48	-0.70749	5.84307	-3.67201	0.617081	-0.25124
Gg_IB9_05G07	459	-0.71854	9.044228	-3.6511	0.663277	-0.31796
Gg_IB9_06A02	477	-0.71173	6.074909	-3.64807	0.670251	-0.32762
Gg_NL_03C03	217	-1.17636	3.856961	-3.53415	0.995602	-0.42455
Gg_NL_09A07	766	0.692444	7.958306	3.611578	0.759691	-0.44327
Gg_IB9_05D01	417	-0.8343	10.77241	-3.51431	1	-0.61931
Gg_IB9_06B03	490	-0.83139	8.152384	-3.51275	1	-0.62407
Gg_IB9_01C12	36	-0.67815	6.259575	-3.55256	0.928649	-0.62841
Gg_IB9_04C08	317	0.815603	7.540923	3.454118	1	-0.80166
Gg_IB9_04A03	288	-0.8038	8.553308	-3.4319	1	-0.86834
Gg_NL_01D05	41	0.81011	7.46814	3.412016	1	-0.92775
Gg_IB9_02F10	165	-0.68128	5.573618	-3.45119	1	-0.94085
Gg_NL_03F02	252	0.665054	6.648745	3.449005	1	-0.94751
Gg_NL_03D04	230	0.797662	7.181408	3.395752	1	-0.97613
Gg_IB9_06G12	559	-0.65125	7.589208	-3.40865	1	-1.06984
Gg_NL_03B09	211	-1.09908	6.078781	-3.30195	1	-1.09232
Gg_IB9_06G08	555	-0.65042	7.778628	-3.39624	1	-1.10723
Gg_IB9_02A07	102	-0.78017	10.29562	-3.31619	1	-1.21015
Gg_IB9_03D04	230	-0.69755	11.15064	-3.35844	1	-1.22047
Gg_IB9_03H08	282	-1.07853	8.455106	-3.24022	1	-1.26383
Gg_IB9_05E10	438	-0.65161	7.035702	-3.33909	1	-1.27805
Gg_IB9_01A02	2	-0.63194	7.517555	-3.29619	1	-1.40473
Gg_NL_09F03	822	-0.63747	9.473154	-3.2602	1	-1.50997
Gg_IB9_03F07	257	-1.04623	6.343998	-3.14318	1	-1.52818
Gg_IB9_05G04	456	-0.65343	10.86384	-3.25127	1	-1.53594
Gg_IB9_03G06	268	0.634349	10.05479	3.213803	1	-1.6443
Gg_IB9_05A03	383	-0.74479	8.344662	-3.16383	1	-1.64583
Gg_NL_11F12	1021	1.023048	5.213786	3.073542	1	-1.71388
Gg_IB9_01B01	13	-0.72304	5.962998	-3.08669	1	-1.86006
Gg_IB9_01D07	43	-0.60078	6.877393	-3.12608	1	-1.89394
Gg_IB9_03G03	265	-0.60061	8.655173	-3.11366	1	-1.92884
Gg_IB9_05H06	470	-0.59717	8.1435	-3.08696	1	-2.00344
Gg_NL_07G01	642	0.978534	8.633039	2.939806	1	-2.0609
Gg_IB9_06A06	481	-0.58151	7.793625	-3.03746	1	-2.14037
Gg_IB9_06C01	500	-0.61478	11.23986	-3.03275	1	-2.15331
Gg_IB9_02C08	127	-0.69604	5.981565	-2.96613	1	-2.1861
Gg_IB9_03G07	269	-0.69246	6.392096	-2.94372	1	-2.2455
Gg_IB9_06E03	526	-0.57893	6.819182	-2.99563	1	-2.25464
Gg_NL_04D10	331	-0.57727	7.593648	-2.99202	1	-2.26444
Gg_IB9_05G02	454	-0.68841	8.114263	-2.92773	1	-2.28767
Gg_NL_03G02	264	0.574325	7.29372	2.982458	1	-2.29034
Gg_NL_05E12	439	0.687671	8.152286	2.925774	1	-2.29282
Gg_IB9_01E12	60	-0.68407	5.338455	-2.91831	1	-2.31242
Gg_IB9_02D02	133	-0.6793	7.003526	-2.89678	1	-2.36873
Gg_NL_10G07	933	0.930853	6.628954	2.796559	1	-2.41829

* cDNA targets excluded from further analysis due to either low fluorescent intensity readings or poor starting quality (i.e. Doublet or smear)

University of Cape Town

Of the top 100 ranked genes, thirty nine were found to have *P*-values of < 0.05 and were considered to be significantly different and thus differentially expressed (Table 4.5). Most published microarray studies, however, have used a post-normalisation cut-off of a two-fold increase or decrease in gene expression level to define differential gene expression (Quackenbush, 2001). Thus, genes showing a two-fold induction (M = 1) or one-half repression (M = -1) of the measured level of expression (M-values), and a *P*-value below 0.05, were classified as statistically significant and thus differentially expressed.

Originally, a number of target spots with weak foreground intensities were flagged to be excluded from further analysis (section 4.2.7.4). As four target spots belonging to a specific cDNA clone were printed onto each microarray, one or more of these spots may have had a weak R and G foreground fluorescent intensity. Such spots are generally considered too faint to show acceptable evidence of differential gene expression. The remaining spots for the specific cDNA target showing acceptable R and G foreground intensities however were not excluded from further analysis. As a result, in cases where the majority of spots for a specific cDNA target were excluded due to weak intensity readings, R considered the remaining target spot(s) to be meaningful and retained the spot(s) for statistical analysis. Thus, the images of each of the genes determined to be differentially expressed were examined with GenePix Pro to check that the statistical results corresponded to the microarray image of each cDNA target spot on each of the arrays. Any genes where the majority of the target spots were flagged to be excluded from analysis were not considered to be statistically significant even though they were present in the results table (Table 4.5).

University of Cape Town

In addition, genes found to be up or down-regulated were compared to their visual images on GenePix Pro to confirm whether the regulation of all target spots for a specific cDNA target were similar in intensity and that no particular spot was dominating the statistical analysis. Up or down-regulated genes can be visually identified from the microarray images by the dye used to label the mRNA from each sample. The probe sample with the dominant mRNA species will bind preferentially to the array. As a result, if a gene is up-regulated in response to nitrogen limitation, a larger proportion of mRNA transcribed from that gene will be present in the original nitrogen limited mRNA sample relative to the nitrogen enriched sample. The Cy dye used to label the nitrogen limited mRNA sample will thus dominate the intensity readings. Conversely, if a gene is down-regulated in response to nitrogen limitation, a larger proportion of mRNA transcribed from that gene will be present in the original nitrogen enriched mRNA sample, and thus, the intensity readings for the dye used to label the nitrogen enriched samples will dominate. If both the nitrogen limited and nitrogen enriched mRNA binds equally to a target spot on the array, the target gene is considered to be constitutively expressed and will appear yellow on the array.

University of Cape Town

Lastly, the 'status' of each target spot found to be differentially expressed was confirmed to ensure that the target spot on the array did not originate from a PCR clone that had a doublet or smear. Thus, of the 39 genes found to be differentially expressed, twelve (489, 54, 152, 968, 4, 471, 216, 995, 464, 581, 38, 1020) were found to be unsuitable for further analysis, either because the intensity readings of the target spots were faint and therefore unsuitable for use as an indicator of differential gene expression or because

their original 'status' was poor (Table 4.5). In total, 27 of the 39 putatively differentially expressed genes were chosen for further characterisation.

University of Cape Town

CHAPTER 5

IDENTIFICATION OF DIFFERENTIALLY EXPRESSED cDNA TARGETS: SEQUENCING AND BIOINFORMATICS

5.1 INTRODUCTION 119

5.2 MATERIALS AND METHODS 121

5.2.1 Isolation and sequencing of plasmids containing differentially expressed *G. gracilis* cDNA fragments 121
5.2.2 Identification of differentially expressed cDNA framents 122

5.3 RESULTS 123

5.3.1 Identification of differentially expressed cDNA fragments 123

5.4 DISCUSSION 126

University of Cape Town

CHAPTER 5

IDENTIFICATION OF DIFFERENTIALLY EXPRESSED cDNA TARGETS: SEQUENCING AND BIOINFORMATICS

5.1 INTRODUCTION

Identification of a novel or unknown ORF allows for an understanding of the probable role of a gene within the organism from which the DNA sequence was derived. Furthermore, understanding the putative function of a gene may allow for the refinement of future studies related to that gene in order to investigate its regulation *in vivo*. DNA sequence similarity is considered to indicate conserved function and is thus a powerful means of discovering the biological function of novel sequences. Comparative sequence homology can therefore be used to identify the function of previously unidentified genes.

With the advent of internet in the 1990s, various databases filled with previously identified DNA sequences, were made freely available to researchers across the world. Today, using various search programs such as BLAST or FASTA, comparative gene studies can be performed against numerous databases. In addition, various search programmes can be used to identify protein signatures, such as divergent domains (e.g. Pfam), functional sites (e.g. PROSITE), or hierarchical definitions from superfamily down to subfamily levels (e.g. PRINTS). Alternatively, integrated documented resources such as InterProScan, can be used to capitalise on the strengths and diagnostic capacities of each of its member databases (Mulder *et al.*, 2005).

University of Cape Town

Essentially, search programmes such as BLAST or InterProScan, allow for the alignment of the unknown gene or protein sequence with gene or protein sequences from multiple databases (Korf *et al.*, 2003). An alignment is defined by Duret and Abdeddaim (2000) as a hypothetical model of all the mutations (insertions, substitutions, and deletions) that can possibly occur during sequence evolution. Overall, the best alignments represent those genes or proteins that are most likely to be evolutionarily similar.

In addition, search programs such as BLAST and FASTA are able to statistically evaluate the chances that an observed similarity occurred purely by chance, and in so doing accept or reject the hypothesis of homology (Duret and Abdeddaim, 2000). An expectation value (e-value) is generated that describes the probability due to chance that there is an alignment to any of the sequences in the database with a similarity greater than the similarity of the query to the sequence shown (Yona and Brenner, 2000). Thus, the lower the given e-value the less chance there is of the query matching any other sequence. Consequently, comparative homology searches allow putative identification of a novel gene if the calculated e-value is considered significant.

In the current investigation, cDNA fragments from the *G. gracilis* nitrogen limited library that were found to be differentially expressed following microarray analysis were sequenced and compared to sequences in various databases to allow for the putative identification of each cDNA fragment.

University of Cape Town

5.2 MATERIALS AND METHODS

5.2.1 Isolation and sequencing of plasmids containing differentially expressed *G. gracilis* cDNA fragments

Glycerol stocks (50%) of *E. coli* DH5-α transformants (section 4.2.1.2) containing recombinant plasmids harbouring differentially expressed cDNA fragments (Table 4.5) were steriley picked and inoculated into 5 ml Luria broth (Appendix A.1) containing 30 μg/ml chloramphenicol (Appendix A.4). The cultures were incubated overnight at 37 °C following which plasmids were extracted according to Sambrook *et al.* (1989) (Appendix C.1). The isolated plasmid DNA was resuspended in 50 μl sterile distilled water, quantified using a Nanodrop spectrophotometer and aliquoted into 5 μg/μl stocks. Additionally, DNA samples were electrophoresed on a 1.2 % agarose gel in order to verify plasmid integrity. The insert DNA was then amplified as in section (4.2.1.3) except that a specially designed LIBF (forward) primer (Appendix B.1.3) was used in place of the previously employed M13F primer. The amplified products were further purified using an E.Z.N.A® Cycle-Pure Kit (peQLab Biotechnologie GmbH) according to the manufacturer's instructions and stored as 20 ng/μl stocks.

Cycle sequencing was performed using the BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems). A 10 μl sequencing reaction was set-up as per Table 5.1, and samples were placed in a 2720 Thermal Cycler (Applied Biosystems). Cycle conditions were carried out as per Appendix B.2.3. Following this, samples were once again purified using the E.Z.N.A® Cycle-Pure Kit (peQLab Biotechnologie GmbH) according to the manufacturer's instructions.

University of Cape Town

Table 5.1 Reaction constituents for the BigDye Terminator v.3.1 Cycle Sequencing Kit

Reaction constituents	Stock	Volume	Final concentration
Primer (LibF or M13R) (Appendix B.2)	1.6 pmol/ul	1 ul	3.2 pMol
DNA	20 ng/ul	1 ul	20 ng/ul
Sequencing buffer	5X	2 ul	1X
Ready reaction mix	2.5X	2 ul	0.5X

The nucleotide sequence of the purified samples was read using a 3130 Genetic Analyzer (Applied Biosystems) employing 3130 Genetic Analyzer Data Collection Software version 3.0. The data was subsequently captured using DNA Sequencing Analysis Software version 5.2. The cDNA sequence data was edited to remove vector sequence using CHROMAS (version 2.01; Technelysium) and subsequently assessed in DNAMAN version 4.1.2.1 (Lynnon Biosoft).

University of Cape Town

5.2.2 Identification of differentially expressed cDNA framents

BLASTx and BLASTn homology searches were conducted against the National Center for Biotechnology Information database (NCBI) (nr). Sequences that returned BLAST matches with an E value of $1e^{-06}$ or less were considered significant, and the putative role and identity of the differentially expressed cDNA fragments were assigned. Any two or more sequences that exhibited homologous BLAST matches were assessed for sequence redundancy using CLUSTALx version 1.81. Those sequences that did not return any significant BLAST result were assessed for the presence of an open reading frame (ORF)

and/or conserved domains using the European Bioinformatics Institute InterProScan facility (EMBL-EBI) (http://www.ebi.ac.uk/InterProScan). Only ORFs longer than 50 codons (150 nucleotides) were accepted as potentially transcribed genes (Lee *et al.*, 1999c).

5.3 RESULTS

5.3.1 Identification of differentially expressed cDNA fragments

Twenty seven cDNA fragments found to be differentially expressed in response to nitrogen limitation (Table 4.5) were sequenced in order to determine their putative identity and biological role within *G. gracilis*. The majority of the 27 differentially expressed cDNA fragments were down-regulated in response to nitrogen limitation. As a result, a further 8 genes (766, 41, 252, 230, 642, 264, 439, and 933) from the top 100 table that were up-regulated in response to nitrogen were sequenced. Although the expression of these 8 cDNA fragments was not found to be statistically significant, the fact that they were ranked in the top 100 table and that the genes may have been significantly up-regulated at an earlier time point, was considered sufficient justification for sequencing the cDNA fragments.

University of Cape Town

The average length of the cDNA fragments inserted into the pDNR-LIB vector (Appendix D) used to create the *G. gracilis* nitrogen limited cDNA library, was between 1.2 and 2.5 kb. Following sequence analysis, the average readable sequence length recovered was approximately 500 bp.

Table 5.1 Putative identification and sequence homology of differentially expressed cDNA fragments from nitrogen limited *G. gracilis*.

cDNA clone	Regulation	Putative identity	Species	E-value	% Identity	Hypothetical Rol
19, 58, 9, 301, 207, 17, 537	Down	Chloroplast hypothetical protein	*Zea mays*	$1.00E^{-09}$	54 %	Cell structure
524, 381, 3	Down	Peptidase C14 caspase catalytic subunit p20 (WD-40 conserved domain)	*Anabaena variabilis* ATCC 29413	$2.00E^{-13}$	39 %	Defence
970	Down	High light inducible protein	*Trichodesmium erythraeum* IMS101	$4.00E^{-06}$	47 %	Defence
685	Down	Peptidase T2, asparaginase	*Strongylocentrotus purpuratus*	$4.00E^{-38}$	54 %	Metabolism
405	Down	Mitochondrion gene for large subunit ribosomal RNA	*Chondrus crispus*	$9.00E^{-25}$	80 %	Cell structure
252	Up	Cyclase dehydrase	*Anabaena variabilis* ATCC 29413	$1.00E^{-17}$	36 %	Metabolism
281	Up	Ribulose-phosphate 3-epimerase	*Arabidopsis thaliana*	$4.00E^{-62}$	73 %	Metabolism
642	Up	Aspartate aminotransferase	*Brevibacterium linens* BL2	$5.00E^{-50}$	48 %	Metabolism
766	Up	Gamma-glutamyl transferase	*Arabidopsis thaliana*	$6.00E^{-15}$	41 %	Metabolism
181	Up	Uridine/Purine nucleoside phosphorylase 1	*Strongylocentrotus purpuratus*	$7.00E^{-14}$	31 %	Metabolism
264	Up	Chromodomain helicase	*Aspergillus fumigatus* AF293	$8.00E^{-15}$	36 %	Protein synthesis
230	Up	Elongation factor 3 homolog	*Chlorella* virus CVK2	$3.00E^{-50}$	48 %	Protein synthesis
41	Up	Elongation factor 1α	*Monsiga brevicollis*	$9.00E^{-98}$	80 %	Protein synthesis

University of Cape Town

Homology searches using the BLASTx and BLASTn algorithms were conducted. Of the 35 sequences analysed, 6 unique sequences were identified that had no significant homology to any sequences in any of the databases searched. A further 8 sequences were found to match genes of unknown function and hypothetical proteins (i.e. at an E value < $1e^{-06}$) (data not shown). Among the remaining 21 sequences found to have significant matches within the NCBI database, a number of sequences were found to be redundant. Several sequences (19, 58, 9, 301, 207, 17 & 537) had homology to a chloroplast hypothetical protein from *Zea mays*, and 3 sequences (524, 381 & 3) were similar to a peptidase C14 caspase catalytic subunit p20 from *Anabaena variabilis* ATCC 29413 (Table 5.1). Multiple sequence alignment revealed that these sequences were indeed identical. Overall, *G. gracilis* cDNA sequences that matched known genes within the NCBI database (Table 5.1) were broadly classified according to function (Fig 5.1).

University of Cape Town

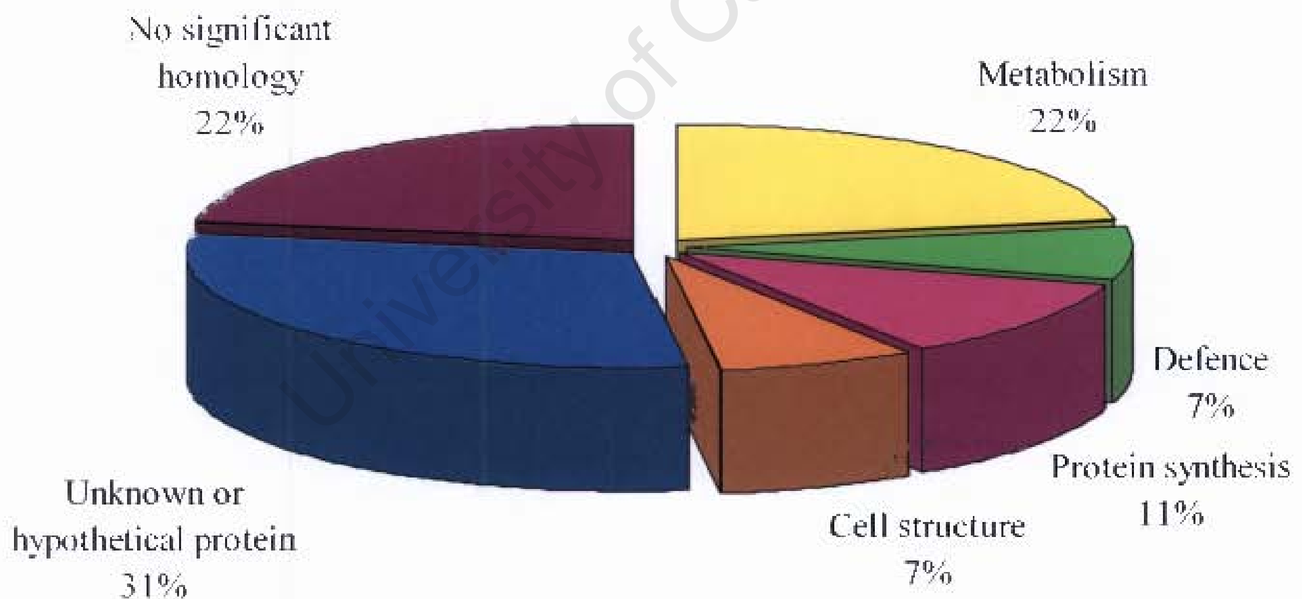


Fig 5.1 Pie diagram showing the distribution of the putatively identified genes from nitrogen limited *G. gracilis* within broad functional categories.

Briefly, six cDNA fragments were found to be related to genes involved in metabolism (281, 642, 766, 685, 181 and 252), three cDNA fragments were similar to genes involved in protein synthesis (264, 41 and 230), two cDNAs were similar to genes involved in cell structure (405 and {19, 58, 9, 301, 207, 17, 537}), and a further two *G. gracilis* cDNA clones were found to be related to genes involved in cell/organism defense (970 and {524, 381, & 3}).

5.4 DISCUSSION

Following microarray analysis a number of cDNA fragments from the *G. gracilis* nitrogen limited and diseased libraries were found to be differentially expressed (Table 4.5). It is important to state that the gene expression data obtained by microarray analysis is only a representation of gene expression at the point at which the *G. gracilis* samples were collected (day 18) and therefore does not reflect the overall genetic response of *G. gracilis* to nitrogen limitation. For example, the presence of short-lived transcripts of a particular cDNA fragment may give the impression that the gene is down-regulated in response to nitrogen limitation, whereas in truth it is not. In order to obtain greater clarity regarding gene regulation in response to nitrogen limitation in *G. gracilis*, time related microarray experiments should be performed to assess transcript abundance and expression over the entire experimental period.

University of Cape Town

Nevertheless, in order to determine the putative identity of each of the cDNA fragments found to be differentially expressed following microarray analysis, the cDNA fragments were sequenced and compared to sequences in the NCBI and EMBL-EBI databases. As

the nitrogen limited cDNA library was not normalised prior to printing the microarray slides, it was not surprising that a number of sequences on the array were found to be redundant. In particular, several cDNA fragments encoding a chloroplast hypothetical protein (19, 58, 9, 301, 207, 17, 537), and three encoding a putative peptidase C14 caspase catalytic subunit p20 (524, 381 and 3), were identified.

The clones encoding the putative chloroplast hypothetical protein and the peptidase C14 caspase catalytic subunit p20 protein were down-regulated in response to nitrogen limitation. The decrease in expression of the chloroplast hypothetical protein may be explainable by virtue of the fact that under conditions of nitrogen deprivation the overall growth rate of *G. gracilis* decreases and there is an increased focus on energy and homeostatic maintenance (Smit *et al.*, 1997). High photosynthetic rates are energy expensive and undesirable in conjunction with decreased growth. As a result, the number of chloroplast proteins that need to be synthesized may be less than that under optimal environmental conditions. Indeed, in expression studies of the marine coccolithophores, *Pleurochrysis carterae* and *Emiliania huxleyi*, genes mapping to multiple fucoxanthin chlorophyll *a/c* binding proteins (*fcp*) were found to be down-regulated in response to nitrogen deprivation (Dyhrman *et al.*, 2006). Down-regulation of various chloroplast and pigment proteins may therefore be a common transcriptional response to nitrogen starvation.

University of Cape Town

In addition, a decrease in chloroplast synthesis may account for the decreased expression of cDNA fragment 970, which was found to have homology to a high-light-inducible

protein from *Trichodesmium erythraeum*. High-light-inducible proteins (HLIPs) are polypeptides involved in protecting plant cells from high light intensity and have been found to be important in photoprotection during exposure to high irradiance (Salem and van Waasbergen, 2004). This is achieved through absorption of light and the transfer of the excitation energy to the photochemical reaction centres (Heddad and Adamska, 2000). The major photosystem (PS) II antenna for Rhodophyta is the phycobilisomes (Heddad and Adamska, 2000). Since the phycobilisome proteins are degraded in *G. gracilis* during nitrogen limitation, it may be possible that there is a decreased requirement for absorption and transfer of excitation energy by high-light-inducible proteins.

The reason for the decreased transcript abundance of the redundant putative peptidase C14 caspase catalytic subunit p20 clones is less obvious. Caspases are typically involved in apoptosis, which is programmed cell death. Although no plant caspases have yet been identified, there is strong evidence for their existence and involvement in plant apoptosis (Kuehn and Phillips, 2005). In particular, caspase-like proteases have been identified in *Arabidopsis* (Danon *et al*., 2004), *Pisum sativum* (Belenghi *et al*., 2004) and tobacco plants (Chichkova *et al*., 2004). Identification of the putative *G. gracilis* peptidase C14 caspase catalytic subunit p20 clones, suggests that caspase-like proteases may be present in marine algae. Recovery of the full nucleotide sequence of the putative peptidase C14 caspase catalytic subunit p20 gene using 3' and 5' Rapid Amplification of cDNA Ends (RACE) will therefore allow for further investigation of the presence and function of this gene product in *G. gracilis*.

University of Cape Town

A number of the remaining *G. gracilis* cDNA fragments were found to be homologous to various metabolic enzymes, the majority of which were up-regulated (181, 281, 252, 642, 766) in nitrogen limited *G. gracilis* (only cDNA fragment 685 was down-regulated). cDNA fragment 181 was found to have homology to a uridine or purine nucleoside phosphorylase I. Purine nucleoside phosphorylases catalyze the cleavage of guanine or inosine to their nucleotide bases and sugar-1-phosphate molecules (Bzowska *et al.*, 2001). Uridine phosphorylases act similarly, catalyzing the cleavage of uridine to uracil and ribose-1-phosphate (Cao and Pizzorno, 2004). The products from these enzymes could possibly be used by *G. gracilis* as either a carbon source or for the rescue of pyrimidine bases for nucleotide synthesis.

The amino acid sequence of cDNA fragment 281 includes the conserved domain of ribulose-phosphate 3-epimerase (RPEase). RPEase functions as a catalyst for the conversion of D-ribulose 5-phosphate to D-xylulose 5-phosphate in the Calvin reductive pentose phosphate cycle. In addition, it has been found to play a role in the oxidative pentose phosphate pathway, which supplies the cell with precursors for the biosynthesis of nucleic acids, amino acids, co-factors and cell wall constituents (Kopp *et al.*, 1999). In *Alcaligenes eutrophus*, RPEase activity was significantly enhanced under autotrophic growth conditions (Kusian *et al.*, 1992), while in *Sacromyces cerevisia*, cytosolic RPEase was found to be indispensable for protection against oxidative stress (Juhnke et al., 1996). The putative *G. gracilis* RPEase may therefore play a role in preserving the overall viability of *G. gracilis* under nitrogen-limiting conditions.

University of Cape Town

The identification of cDNA fragment 642 as a putative aspartate aminotransferase (AspAT) was particularly interesting as this enzyme is known to play an important role in both nitrogen and carbon metabolism (section 1.6). Aspartate aminotransferases catalyse the formation of 2-oxoglutarate and aspartate via reversible amino group transfer from glutamate to oxalacetate (Silvente *et al.*, 2003). Multiple aspartate aminotransferase isoenzymes have been identified in plants. Indeed, at least three isoenzymes of AspAT have been identified in several subcellular compartments in the corn plant *Panicum miliaceum* (Taniguchi *et al.*, 1995). Moreover, differential expression of the genes encoding each of these isoenzymes has been observed under different environmental stimuli such as light and nitrate availability (Taniguchi *et al.*, 1995). *P. miliaceum* plastid AspAT (pAspAT) transcript levels were found to gradually decrease during recovery from nitrogen limitation whereas mitochondrial AspAT (mAspAT) and cytosolic AspAT (cAspAT) transcripts were seen to increase (Taniguchi *et al.*, 1995). Although mAspAT transcript levels were found to be elevated, an equivalent increase in mAspAT activity was not observed. In contrast, enzyme activity of the cAspAT was found to correlate with cAspAT mRNA levels, and was consequently believed to be transcriptionaly regulated by nitrogen availability (Taniguchi *et al.*, 1995).

University of Cape Town

As gene transcripts can be regulated transcriptionally and post-transcriptionally, mRNA abundance may not always be an indication of enhanced or decreased protein activity. In the current investigation, the putative *G. gracilis* AspAT was found to be up-regulated in response to nitrogen limitation, corresponding to results obtained by Gebrekiros (2003) (MSc thesis, UCT). In order to classify the putative *G. gracilis* AspAT at the isoenzyme

level, recovery of the full length gene sequence using 3' and 5' RACE is required. Western blot investigation should also be performed in order to determine whether the observed increase in AspAT mRNA corresponds to an increase in the AspAT protein levels.

The deduced amino acid sequence of cDNA fragment 685 has homology to an asparaginase enzyme from the purple sea urchin, *Strongylocentrotus purpuratus*. Like the putative AspAT identified in this study, asparaginase proteins play a role in nitrogen metabolism as a major nitrogen transport and storage compound. In particular, L-asparaginase enzymes play a role in the mobilization of the amide nitrogen from asparagine for utilization in amino acid and protein synthesis in higher plants (Paul and Cooksey, 1981). In the current investigation, the putative asparaginase enzyme was shown by the microarray analysis to have a decreased expression after 18 days exposure to nitrogen limitation. In some organisms, such as *Saccharomyces cerevisae* and *Aspergillus nidulans*, nitrogen deprivation has been shown to cause an increase in L-asparaginase levels (Dunlop and Roon, 1975). However, in *Klebsiella aerogenes*, conditions which led to the depression of nitrogen (i.e. nitrogen deprivation) were shown to lead to a similar depression in L-asparaginase levels (Resnick and Magasanik, 1976).

University of Cape Town

L-asparaginase synthesis is activated by glutamine synthetase (GS), and is thus believed to be regulated in parallel with GS and other enzymes capable of supplying the cell with ammonia or glutamate (Resnick and Magasanik, 1976). Both GS and asparaginase are believed to be regulated by a common mechanism responding to intracellular nitrogen-

depletion (Paul and Cooksey, 1981). Indeed, both GS and asparaginase activity in a euryhaline *Chlamydomonas* species were found to be repressed by high levels of nitrogen (Paul and Cooksey, 1981). Down regulation of the *G. gracilis* asparaginase after 18 days exposure to nitrogen limiting conditions suggests that protein synthesis may be down-regulated at this point. In addition, as expression data from the current microarray investigation is representative of *G. gracilis* gene expression at day 18, the overall genetic response of *G. gracilis* to nitrogen limitation may not be correctly reflected. Thus, the putative asparaginase gene transcript levels may have been elevated in response to nitrogen limitation prior to day 18. Further characterization of the putative *G. gracilis* asparaginase at both the genetic and protein level, together with time related expression studies, should therefore be performed in order to assess *G. gracilis* asparaginase regulation.

cDNA fragment 252 was found to have homology to a cyclase dehydrase enzyme from the filamentous heterocyst-forming cyanobacterium, *Anabaena variabilis* ATCC 29413. Cyclase dehydrases are believed to be involved in the biosynthesis of polyketides (Ye *et al.*, 1994). Polyketides are categorized as lipids and are described as secondary metabolites that play a role in intercellular communication and defense (Hutchinson, 2003). Thus, the up-regulated putative cyclase dehydrase identified in the current investigation may be a specific response to nitrogen limiting stress conditions in *G. gracilis*. Alternatively, the intracellular lipid composition of the alga *Botryococcus braunii* was reported to change under conditions of nitrogen deficiency

University of Cape Town

(Zhila *et al.*, 2005). Thus, the putative cyclase dehydrase may be involved in aiding this process in *G. gracilis*.

cDNA fragment 766 codes for another interesting putative enzyme, namely gamma-glutamyltransferase. Gamma-glutamyltransferase is a glycoprotein that catalyzes the transpeptidation and hydrolysis of the gamma-glutamyl group of glutathione (GSH) and related compounds (Ikeda and Taniguchi, 2005). Storozhenko *et al.* (2002) reported that the GSH compounds are involved in a number of important cellular functions such as: storage and transport of reduced sulphur; protection against oxidative stress; detoxification of xenobiotics and heavy metals; redox regulation of gene expression; and progression through the cell cycle. Thus, gamma-glutamyltransferases may play a role in stress tolerance. Indeed, a study performed by Paul and Cooksey (1981) showed that gamma-glutamyltransferase activity was elevated in nitrogen-limited cultures of a euryhaline *Chlamydomonas* species.

To date, aside from gamma-glutamyl cysteine synthetase, gamma-glutamyltransferases are the only enzymes known to hydrolyze the unique N-terminal amide bonds of reduced glutathione, oxidized glutathione, and glutathione S-conjugates (Martin and Slovin, 2000). Despite the presence of gamma-glutamyltransferases in various plant species, including *Arabidobsis thaliana* (Storozhenko *et al.*, 2002) and *Lycopersicon esculentum* (Martin and Slovin, 2000), plant gamma-glutamyltransferases have not been well characterized. As a result, many plant gamma-glutamyltransferase studies are based on what is known from mammalian GSH metabolism. Presently, the role of plant gamma-

University of Cape Town

glutamyltransferase activity in GSH metabolism is unclear as GSH has been reported to be a poor *in vitro* substrate for plant gamma-glutamyltransferases (Lancaster and Shaw, 1994). Enzymatic studies on an *Arabidopsis* gamma-glutamyltransferase, however, suggested that the enzyme does play an analogous role to mammalian gamma-glutamyltransferases in the catabolism of GSH (Storozhenko *et al.*, 2002). In addition, plant gamma-glutamyltransferases are believed to be involved in the synthesis of a range of gamma-glutamyl dipeptides formed during fruit ripening (Martin and Slovin, 2000). Further classification of the putative *G. gracilis* gamma-glutamyltransferase may therefore add to the limited knowledge regarding plant gamma-glutamyltransferase function and activity.

University of Cape Town

cDNA fragments 41 and 230 possess homology to the amino acid sequence of various elongation factors harbouring conserved ABC-ATPase domains. Elongation factors are involved in protein biosynthesis in that they promote the binding of aminoacyl tRNA to the A site of ribosomes in a GTP-dependent manner (Yamda *et al.*, 1993). In addition, they catalyse the translocation of the synthesised protein chain from the A to the P site of the ribosome (Yamda *et al.*, 1993). cDNA fragments 41 and 230 may therefore be up-regulated in response to nitrogen limitation in order to promote the biosynthesis of various proteins that support *G. gracilis* survival under the imposed nitrogen limitation conditions.

cDNA fragment 264 was found to have homology to a chromodomain helicase from the fungus *Aspergillus fumigatus*. Chromodomain helicase DNA-binding proteins (CHDs)

are chromatin remodelling proteins that posses an ATP-dependent remodelling activity that is distinguished by the catalytic ATPase domain (Shur and Benayahu, 2005). CHD proteins contain various DNA-binding and protein–protein binding motifs. These motifs allow for the formation of multi-protein complexes that may include subunits important for the regulation of gene transcription (Shur and Benayahu, 2005). Consequently, CHDs act to facilitate recruitment of RNA polymerase, elongation and transcriptional termination. cDNA fragment 264 may thus be involved in regulating gene expression under conditions of nitrogen limitation.

Lastly, cDNA clone 405 was shown to have 80 % similarity to a mitochondrion gene for the large subunit ribosomal RNA of the red alga *Chondrus crispus*, identified by Leblanc *et al*. (1995) while performing an analysis of mitochondrial DNA sequences of the red alga. Thus, the observed down regulation of cDNA fragment 405 following 18 days exposure to nitrogen limitation suggests that following this point, increased exposure of *G. gracilis* to nitrogen limiting conditions may result in a decrease in overall protein synthesis.

The inability to attain a significant BLAST result for the remaining 14 sequences (394, 488, 382, 10, 24, 271, 450, 1, 418, 422, 439, 604, 933 and 26) may be attributed to the limited number of *G. gracilis* and algal ESTs within the existing databases (Jackson and Degnan, 2006). In a transcriptome profiling experiment with the coccolithophore *Emiliania huxleyi*, thirty eight transcripts were found to be up-regulated in response to nitrogen limitation, of which only six could be assigned a putative function

University of Cape Town

(Dyhrman *et al.*, 2006). Thus, a number of nitrogen specific genes may exist for which the identity and function cannot yet be established. Further investigation regarding the identity and function of the fourteen unidentified *G. gracilis* cDNA sequences should therefore be performed. Acquisition of the full length sequence of each of these genes using 3' and 5' RACE may enhance the likelihood of finding comparative homology with sequences within the various databases. Additionally, assorted mutation and knockout studies may allow for the characterisation of gene expression under different conditions, thereby identifying a putative role for the unidentified genes.

Overall, a number of interesting genes belonging to a variety of functional categories were identified. Determination of the full length sequence of each of the differentially expressed cDNA fragments should allow for the identification of the role(s) of each of these genes in *G. gracilis* exposed to nitrogen limitation. In addition, western blot analysis can be used to confirm whether protein expression levels mirror transcript abundance for each of the differentially expressed cDNA fragments. With this knowledge, genetically modified crops that are more tolerant to nitrogen limiting conditions could be produced which would improve farmed crop maintenance and yield.

University of Cape Town

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

Farming of *Gracilaria* for its hydrocolloids means 'producing thousands of dry tonnes per year economically so as to be cost competitive in the raw materials world market' (Santelices and Doty, 1989). Due to the high operational and harvesting costs associated with colloid production, the economic viability of culturing this species is dependant on the collection of substantial *Gracilaria* biomass. Maximal productivity, however, is rarely achieved due to the impact of environmental stress factors such as climatic, biotic and nutrient imbalances. Such factors prevent the alga from functioning at its optimal 'genetic potential', driving it away from cellular homeostasis to a new homeostatic condition that enables the seaweed to cope with the stress condition at hand. Such defence responses require regulatory changes to the activation of multiple genes and gene pathways.

University of Cape Town

Understanding the mechanisms that govern seaweed tolerance and adaptation to external factors provides an insight into how cultivation techniques can be optimized to allow for improved field crop management and performance (Bohnert *et al.*, 2006). Stress response studies at various organizational levels, ranging from the level of the whole plant to that of a single gene, are thus fundamental in improving field crop management (Alscher *et al.*, 1990). In the past ten decades, genomic technologies have emerged that have provided a wealth of data and thus a better understanding of the changes in cellular

metabolism that are induced by biotic and abiotic stresses (Bohnert *et al.*, 2006). At present there is considerable interest in the field of genetic engineering. This field uses inducible systems for the expression of genes introduced into plants through transformation or other means. Genetic engineering therefore allows the manipulation of gene expression within a true isogenic background, leading to a more comprehensive understanding of individual gene function (Reynolds, 1999).

In terms of *Gracilaria* culture, the ability to modify and control gene expression would enhance the potential for increased yield stability under stressful conditions. The production of such crops however, first requires the identification of genes that are differentially expressed in response to the external stress. In the mid 1990s, papers were published describing the use of microarray technology for parallel surveillance of the expression of thousands of genes in a single assay (Rockett, 2004). Since then, microarray technology has emerged as one of the most promising tools in molecular biology for the detection of differential gene expression (Rockett, 2004).

University of Cape Town

In the current investigation, microarray technology was utilized to identify genes differentially expressed in response to nitrogen limitation in the red alga *G. gracilis*. Unlike model organisms such as *Arabidopsis thaliana*, *Mus musculus* or *Homosapiens sapien*, for which microarray slides are commercially available, *G. gracilis* slides cannot be purchased and thus *de novo* array design and printing is required. As a result, nitrogen limited *G. gracilis* clones from an established cDNA library were used as targets for the microarray slides and printed on the array such that each target was duplicated next to

itself on the array, and each array was duplicated on the slide itself. Thus, the microarray was designed to increase the strength of statistical significance for each gene through replication, as well as to provide a means of technical replication for comparison of hybridisation quality within the array.

Once the microarray slides were successfully printed, nitrogen deprivation experiments were conducted to determine when thallus samples become nitrogen limited. Thus, after 18 days exposure to nitrogen deprivation, thalli were found to be nitrogen limited and consequently collected as samples for microarray. To ensure that both nitrogen enriched and nitrogen limited samples were exposed to the same 'environmental' conditions and that any differences observed in gene expression could primarily be attributed to nitrogen limitation and not other stress factors such as light or CO_2 toxicity, day 18 nitrogen enriched (rather than day 0) thallal samples were chosen as the control (reference) for use in the microarray.

University of Cape Town

The process of producing, using, and analysing a microarray involves several discrete steps. There is ample opportunity for technical variability to inundate the biological variation that the tool intends to measure (Rockett, 2004). Thus, in order to reduce such technical variation and ensure improved replication of the microarray experiment, optimisation of the RNA isolation, cDNA synthesis, labelling and hybridisation steps was attempted. Originally, CAPAR cDNA synthesis and labelling procedures (adapted from Hedge *et al.*, 2000) were followed. However, Cy dye labelling was unsuccessful using this approach. Ambion's MessageAMP™ II aRNA kit was thus utilised with RNA

isolated from nitrogen limited and nitrogen enriched day 18 thallus samples prepared with TRIzol™ reagent. The labelled products were hybridised to the array, scanned and analysed using GenePix Pro software (version 6).

Even though the microarray experiments conducted in this study were satisfactory, an optimal procedure for RNA isolation was not successfully established. RNA extracted according to the TRIzol™ protocol had very low $A_{260/230}$ values, indicating possible polysaccharide contamination. In addition, variable yields of cDNA were recovered when the RNA was amplified and reverse transcribed with Ambion's MessageAMP™ II aRNA kit. Guo *et al.* (2004), however, reported that RNA purity has no significant influence on final array image quality. Thus, even though the RNA isolated from the *G. gracilis* thalli in the current investigation was not of ideal purity, it was felt that the RNA was of a suitable quality for use in microarray analysis. Nevertheless, RNA purity and integrity should not be overlooked and further optimization of an RNA isolation protocol specific for nitrogen limited and nitrogen enriched *G. gracilis* thalli is advisable. Repetition of the entire microarray experiment with higher quality RNA would not only serve as a means of comparing the influence of starting RNA quality on the overall array quality, but in addition, improve the overall level of statistical replication.

Following hybridization, image analysis of the scanned microarrays was completed and microarray data normalized using the statistical interface, R. A number of normalization algorithms were attempted in order to account for within-array variability. Overall, results indicated that the use of a robust spline approach was more efficient in

University of Cape Town

normalizing the microarray data than global and print-tip lowess adjustments. In agreement with Soler (2004), normalisation using the robust spline algorithm allowed improved control over dye bias, and in addition, provided a clearer identification of differently expressed genes within the experiment.

In total, 35 differentially expressed cDNA fragments were sequenced. The deduced amino acid sequences of the *G. gracilis* cDNA fragments were subsequently aligned with sequences in the National Centre for Biotechnology Information database in order to determine the putative identity of each sequence. Overall, 21 sequences were assigned a putative identity of which ten were redundant and thus, only 13 sequences were uniquely identified. Of the remaining 14 sequences, 6 sequences had no significant homology to any of the sequences present in any of the databases employed, while 8 sequences showed homology to various hypothetical proteins of unknown function. Of the 13 putatively identified sequences, three known nitrogen metabolic enzymes were identified; namely a putative aspartate aminotransferase, a putative asparaginase and a putative gamma-glutamyltransferase. The remaining putatively identified genes were found to have homology to either metabolic enzymes, enzymes involved in controlling gene expression and protein translation, enzymes involved in cell structure or enzymes putatively involved in cell/organism defense. The putative identification of these differentially expressed genes has thus provided a foundation for further analytical studies.

University of Cape Town

Although the use of microarray technology as a tool for generating and analysing gene expression data is rapidly evolving, an understanding of the general pitfalls of the technique and its associated technologies remains in its infancy (Rockett, 2004). Thus, a large amount of confirmatory analyses are required to validate the results obtained. On the whole, confirmatory studies take the form of conducting some alternative means of quantitating mRNA abundance. In general, northern blot or quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis of a limited number of differentially expressed genes identified by microarray is performed. Thus, in order to qualitatively support the microarray data in the current investigation, further confirmatory studies on a few of the genes found to be up or down-regulated in response to nitrogen limitation is required.

If the data from northern blot/qRT-PCR studies confirm that the observed changes in gene expression in response to nitrogen limitation are genuine, Rapid Amplification of cDNA Ends (RACE) of the top differentially expressed sequenced cDNA fragments can be performed. 3' and 5' RACE allows for the effective determination of the full length sequence of a gene, and consequently, its identity and potential function can be assigned with confidence. Once the complete gene sequence has been obtained, various 'knockout' and mutational studies can be performed making use of transformation and genetic engineering technologies. Such studies allow manipulation of gene expression through overproduction or deletion of a gene in order to provide a fuller understanding of individual gene function. Thus, genes found to enhance *G. gracilis's* ability to withstand nitrogen limiting conditions *in vivo* may have the potential to increase algal yield and

University of Cape Town

stability under stressful conditions if used in future field-based, inducible expression systems.

In addition to the suggested verification, identification and functional analysis studies, time/dose related experiments can be performed in order to characterise temporal expression of different sets of genes regulated in response to the abiotic stress. Additionally, such studies could aid in identifying cohorts of co-regulated genes. In general, genes that belong in a particular pathway or that respond to a common environmental challenge should exhibit similar patterns of expression, and thus with the aid of computational and statistical tools, common patterns of gene expression could be identified (Quackenbush, 2001). Furthermore, time-related microarray experiments would provide answers to questions such as how, when and to what degree a gene or gene pathway is activated or repressed in response to the abiotic stress.

Overall, the current microarray investigation has served as a platform for future research into the response of *G. gracilis* to nitrogen limitation, and ultimately, may lead to commercial aquaculture of *G. gracilis* due to increased, year-round productivity in Saldanha Bay.

University of Cape Town

APPENDIX A

MEDIA, BUFFERS AND SOLUTIONS

All media were autoclaved at 121 °C for 20 min prior to use, unless otherwise specified.

Water used for making solutions, media and diluting buffers was purified using a Milli-RO Plus (Millipore) water purification system.

A.1 MEDIA

A.1.1 Artificial Sea Water (ASW)

NaCl (Saarchem)	24.7 g
$MgCl_2.6H_2O$ (Saarchem)	4.7 g
KCl (Saarchem)	0.66 g
$CaCl_2.2H_2O$	1.9 g
$MgSO_4.7H_2O$ (Saarchem)	6.3 g
$NaHCO_3$	0.18 g
Water to	1 L

Autoclave

A.1.2 Fe-solution

$Fe(NH_4)_2(SO_4)_2.6H_2O$	702 mg
Na_2 EDTA	600 mg
Water to	1 L

A.1.3 PII metal solution

Na_2 EDTA	100 mg
H_3BO_3	114 mg
$FeCl_3.6H_2O$	4.9 mg
$MnSO_4$	16.4 mg
$ZnSO_4.7H_2O$	2.2 mg
$CoSO_4.7H_2O$	0.48 mg
Water to	100 ml

University of Cape Town

A.1.4.1 PES-enriched seawater medium (1/3 strength) (Provasoli, 1968)

$NaNO_3$	350 mg
Na_2 glycerophosphate $5H_2O$	50 mg
Fe solution	25 ml
PII	25 ml
Vitamin B_{12}	10 µg
Thiamine	0.5 mg
Biotin	5 µg
Tris buffer (Sigma Co.)	500 mg
Water to	100 ml

Adjust pH to 7.8, autoclave and store at 10 °C.
Add 6.6 ml to 1 L ASW.

A.1.4.2 PES-N (ES medium lacking nitrogen)

Prepared exactly the same as ES medium except no $NaNO_3$ is added and $Fe_2(SO_4)$ instead of $Fe(NH_4)_2(SO_4)_2.6H_2O$ when making Fe-solution.

University of Cape Town

A.1.5 Luria broth

Tryptone (Biolab)	10 g
Yeast extract (Biolab)	5 g
NaCl (Saarchem)	10 g
Water to	1 L

Autoclave

A.1.6 Luria agar

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
Agar (Biolab)	15 g
Water to	1 L

Autoclave

A.2 (ELECTROPHORESIS) BUFFERS AND DYE

A.2.1 Tris-Acetate-EDTA (TAE) buffer (50 X stock)

Tris base (Roche)	242 g
Glacial acetic acid (Saarchem)	57.1 g
EDTA (0.5 M, pH 8) (Saarchem)	100 ml
Water to	1 L

A.2.2 Tris-EDTA (TE) buffer (pH8) (1 M stock)

Tris-Cl (1 M, pH 8)	1 ml
EDTA (0.5 M, pH 8)	200 µl
Water to	100 ml

A.2.3 DNA gel tracking dye (6 X)

Bromophenol blue (Saarchem)	62.5 mg
Sucrose (Saarchem)	10 g
Water	15 ml
EDTA (0.5 M, pH 8)	1 ml

Make up to a final volume of 25 ml with water.

Autoclave

University of Cape Town

A.2.4 MOPS buffer for RNA agarose-formladehyde gels (10 X)

MOPS (Sigma)	20 g
Na Acetate (Saarchem)	1 g
EDTA (0.5 M, pH 8)	10 ml
DEPC- dH_2O	470 ml

Adjust pH to 7 with NaOH and Filter sterilize through a 0.22 µm Millipore filter using a 60 cc syringe. A 1X solution of this was used as running buffer for RNA agarose gels. Store in the dark at 4 °C.

A.2.5 RNA sample buffer and gel tracking dye

MOPS (10 X) (pH 7)	300 ul
Formaldehyde (37 %) (Sarchem)	80 ul
Formamide (Saarchem)	900 ul
Ethidium Bromide (10 mg/ml) (Sigma)	2 ul
Dye	220 ul

The dye used was made up as follows:

Xylene cyanol	50 mg
Bromophenol Blue	50 mg
DEPC- dH_2O	1 ml

RNA Sample buffer was added at a ratio of 1:2 for RNA sample:sample buffer. The samples were then heated at 65 °C for 15 min, followed by snap cooling on ice before loading into gel wells.

A.3 Solutions for Rubidium Chloride (RbCl) competent cell method

A.3.1 TFB 1

RbCl (1 M) (Saarchem)	1.21 g
$MnCl_2.4H_2O$ (HOLPRO Analytical)	0.99 g
KOAc (Saarchem)	0.294 g
$CaCl_2.2H_2O$ (750 mM) (Saarchem)	147 mg
Glycerol (50 %, w/v) (Saarchem)	30 ml

Adjust to pH 5.8 with Glacial Acetic acid, make to volume of 100 ml with dH_2O and filter sterilise.

A.3.2 TFB 2

MOPS (100 mM) (pH 7)	0.21 g
RbCl (1 M)	0.121 g
$CaCl_2.2H_2O$ (750 mM)	1.1 g
Glycerol (50 %, w/v)	30 ml

Make to volume of 100 ml with dH_2O and filter sterilise.

University of Cape Town

A.4 SOLUTIONS FOR PLASMID ISOLATION

A.4.1 Chloramphenicol (Cm) (30 mg/ml)

Chloramphenicol (Roche)	30 mg
Ethanol	10 ml

Store at -20 °C

A.4.2 Solution 1 (10 X stock)

Tris-Cl (1 M, pH 8)	25 ml
Glucose (20 % w/v) (Saarchem)	45.5 ml
EDTA (0.5 M, pH 8)	20 ml
Water	9.5 ml

Solution 2

NaOH (10 N) (Saarchem)	2 ml
Sodium Dodecyl sulphate (SDS) (25 %, w/v) (Saarchem)	4 ml
Water	94 ml

This solution is made fresh weekly

Solution 3

K-acetate (Saarchem)	147 g
Water	250 ml

pH to 4.8 – 5.0 with Acetic acid. Make up to a final volume of 500 ml.

University of Cape Town

A.5 SOLUTIONS FOR RNA ISOLATIONS

A.5.1 Diethylpyrocarbonate water (DEPC- dH_2O)

DEPC (Sigma)	1 ml
Distilled water	1 L

Shake solution vigorously until no DEPC droplets remain, and leave O/N in the fumehood. Autoclave.

A.5.2 1.2% RNA Agarose Formaldehyde gel

Agarose	0.72 g
DEPC- dH_2O	43.92 ml

Let the solution cool to 60 °C, then add:

MOPS (10 X) (pH 7)	6 ml
Formaldehyde	10.08 ml

Pour gel and leave to set.

4.5.3 Lysis solution for Chomczynski *et al* (1987) RNA extraction - Solution D

Guanidium thiocyanate (4 M) (Sigma)	4.7 g
Sodium Citrate stock (1 M) (Saarchem)	0.25 ml
Sarkosyl stock (10 %, w/v) (Sigma)	0.5 ml
β-Mercaptoethanol (Merck)	70 ul
DEPC-dH_2O	9.18 ml

The solution is made up fresh on the day it is used, and must be filter sterilized before the addition of β-Mercaptoethanol.

A.5.4 Lysis solutions for Pine Tree RNA extraction procedure

Tris HCl (1 M, pH 8)	1.5 ml
CTAB (10 %, w/v) (USB)	3.0 ml
EDTA (0.5 M, pH 8)	0.9 ml
NaCl (5 M)	6.0 ml
Spermidine (10 %, w/v) (Sigma)	75 ul
Polyvinylpolypyrrolidinone (PVPP) (Sigma)	0.3 g
β-Mercaptoethanol	0.3 ml

Both PVPP and β-Mercaptoethanol are added just before use. Proteinase K (10 mg/ml) (Sigma) is then added to a final concentration of 1.5 mg/ml.

A.5.5 10 M LiCl

LiCl (Saarchem)	42.39 g
Water	100 ml

Autoclave.

University of Cape Town

A.6 SOLUTIONS FOR MICROARRAY LABELLING AND HYBRIDISATIONS

Solutions 4, 5 and 6 are made up fresh on the day and are pre-heated to 42 °C before use.

A.6.1 SSC (20 X)

NaCl	175 g
$C_6H_5Na_3O_7{\cdot}2H_2O$ (Saarchem)	88.2 g
Water to	90 ml

Adjust PH to 7.4 and make up to 1 L. Add 2 ml DEPC and autoclave.

A.6.2 10 % BSA

BSA (Sigma)	10 g
Water to	100 ml

Filter sterilise and freeze.

A.6.3 20 % SDS

SDS	20 g
Water to	100 ml

A.6.4 Pre-hybridisation buffer

SSC (20 X)	50 µl
SDS (20 %, w/v)	1 µl
BSA (10 %)	20 µl
Water to	129 µl

A.6.5 Hybridisation buffer (2 X)

SSC (Sigma) (20 X)	100 µl
SDS (Life Technologies) (20 %)	2 µl
Formamide (Life Technologies)	100 µl

University of Cape Town

A.6.6 Low stringency wash buffer

20 X SSC	40 ml
25% SDS	10 ml
Water to	350 ml

A.6.7 Medium Stringency wash buffer

20 X SSC	10 ml
Water to	390 ml

Made up fresh on the day

A.6.8 High stringency wash buffer

SSC (20 X)	0.5 ml
Water to	199.5 ml

Made up fresh on the day.

A.6.9 Probe hybridisation mixture

Cy3 Probe	1.0 µg
Cy5 Probe	1.0 µg
COT1-DNA (Life Technologies) (20 ug/ul)	1.0 µl
Poly(A)-DNA (Sigma) (20 ug/ul)	1.0 µl
Hybridisation buffer	60.0 µl

The mixture is heated at 90 °C for 3 minutes before injecting onto slide.

University of Cape Town

APPENDIX B

DNA PRIMER SEQUENCES AND PCR CYCLE PROFILES

B.1 DNA PRIMER SEQUENCES

All synthetic oligonucleotide sequences were supplied by the Oligonucleotide synthesising service of the Biochemistry Department, University of Cape Town, Cape Town, South Africa.

B.1.1 Primers used in colony PCR

M13(-21)-F	5’	TGT AAA ACG ACG GCC AG	3’
M13(-21)-R	5’	CAG CTA TGA CCA TGT TCA C	3’

B.1.2 Primers used in 18S rRNA gene confirmatory PCR

E18R	5’ CTA CGG AAA CCT TGT TAC GAC TTC TCC 3’
R18F	5’ CCT GGT TGA TCC TGC CAG TGG 3’

B.1.3 Primers used for sequencing

Lib-F	5’	AAC GAC CGA GCG CAG CGA GTC	3’
M13(-21)-R	5’	CAG CTA TGA CCA TGT TCA C	3’

University of Cape Town

B.2 PCR CYCLE PROFILES

B.2.1 Colony PCR

1 cycle	denaturation	96 °C	5 min
25 cycles	denaturation	94 °C	30 s
	annealing	61 °C	30 s
	extension	72 °C	3 min
1 cycle	extension	72 °C	7 min
1 cycle	cool down	4 °C	10 min

B.2.2 18S rRNA gene PCR

1 cycle	denaturation	94 °C	2 min
40 cycles	denaturation	94 °C	30 s
	annealing	57.5 °C	1 min
	extension	68 °C	2 min
1 cycle	extension	68 °C	7 min
1 cycle	cool down	4 °C	∞

B.2.3 PCR cycle for sequencing

1 cycle	denaturation	96 °C	1 min
25 cycles	denaturation	94 °C	30 s
	annealing	50 °C	5 s
	extension	60 °C	4 min
1 cycle	cool down	4 °C	10 min

University of Cape Town

APPENDIX C

STANDARD METHODS

C.1 Agarose gel electrophoresis

Agarose gel electrophoresis of DNA was carried out according to Sambrook *et al.* (1989). TAE buffer (Appendix A) was used to dissolve Agarose (D1 LE, Whitehead Scientific) and 2 µl ethidium bromide (10 mg/ml) was added per 50 ml agarose solution. DNA electrophoresis loading buffer (Appendix A) was added to DNA samples prior to loading DNA into the wells at a ratio of 1:3. Visualisation of DNA was performed using a Universal Hood II 264 nm transilluminator and Camera (BIORAD) using Quantity One version 4.5.2 software.

C.1.2 Preparation of competent *E. coli* DH5 α cells

E. coli DH5α cells were streaked onto a fresh Luria agar plate (Appendix A) and left to incubate at 37 °C overnight. Using a sterile loop, a single colony from the freshly streaked bacterial plate was resuspended in 5ml Luria broth (Appendix A). The culture was left to incubate with agitation overnight at 37 °C. A one in two hundred dilution of this culture was made in a volume of 200 ml of pre-heated Luria broth and incubated for 3 h at 37°C with agitation until the OD_{550} reached 0.35 (early log phase). The culture was transferred into a GSA (Beckman) centrifuge tube and was subsequently left on ice for 15 min. The cells were pelleted at 4 °C by centrifugation at 2 500 rpm for 5 min and the supernatant poured off. The cells were gently resuspended in 21 ml ice-cold TFB1 (Appendix A), and left to incubate on ice for 90 min. The resuspended cells were centrifuged as before at 2 500 rpm for 5 min at 4°C, and the supernatant decanted. The cells were gently resuspended in 3.5 ml ice-cold TFB2 (Appendix A). Once resuspended, the cell solution was aliquoted into 100 µl amounts and either used directly for transformation of ligated DNA or stored at -70°C in 50 % sterile glycerol.

C.1.2.2 Transformation of pDNR-LIB clones into competent *E. coli* DH5 α cells

A transformation mix was prepared consisting of 2 µl DNA (pDNR-LIB clones) and 100 µl (aliquot) of the *E. coli* DH5α competent cells (Appendix C.1.2). The mixture was left to incubate on ice for 10 min following which the cells were heat shocked at 42 °C for 2 min. A volume of 0.9 ml Luria broth supplemented with chloramphenical (30 mg/ml) (Appendix A) was added to the cells. The cells were left on a shaker for 60 min at 37 °C before being plated. Luria Agar plates were used as viability control plates, while Luria agar supplemented with chloramphenical (30 mg/ml) was used to select transformants. Transformation of the *E. coli* DH5α competent cells with PSK vector

University of Cape Town

served as an additional control to determine the competency of the *E. coli* DH5α cells. These transformants were selected for on Luria agar supplemented with ampicillin (10 mg/ml).

C.1.3 Isolation of plasmid DNA

Isolation of the pDNR-LIB clones from transformed *E. coli* DH5α was carried out according to Sambrook *et al*. (1989). A 5 ml culture of transformed *E. coli* DH5α cells were incubated overnight at 37 °C with agitation in Luria broth supplemented with chloramphenical (30 mg/ml) (Appendix A). A 2 ml volume of the culture was transferred into a 2 ml microfuge tube, microfuged for 30 s at 12 000 rpm, and the supernatant removed. An additional 2 ml of culture was added to the pelleted cells, and the microfuge process repeated. The supernatant was removed and the pellet retained. The cell pellet was then resuspended in 0.2 ml solution 1 (Appendix A) and left at room temperature for 10 min to allow for cell lysis. Following this 0.4 ml solution 2 (Appendix A) was added and the cell mixture was left to incubate on ice for 10 min. Pre–cooled solution 3 (Appendix A) was then added to precipitate chromosomal DNA, and the cell mixture was left on ice for 10 min. Consequently, plasmid DNA was recovered through microfugation at room temperature for 5 min. The supernatant containing the plasmid DNA was transferred to a fresh 1.5 ml microfuge tube to which 0.6 ml isopropanol was added. The mixture was left for 2 min at room temperature and subsequently microfuged at 12 000 rpm for 10 min. The supernatant was decanted and the pellet retained. The pellet was washed in 70 % ethanol, rinsed by inversion, air-dried and re-dissolved in 30 µl sterile Milli Q water (Appendix A).

University of Cape Town

C.1.4 Preparation of 2:3 Aminoallyl dUTP:dTTP mix

The dNTP mix was prepared as follows:

dATP (100 mM)	5ul
dCTP (100 mM)	5ul
dGTP (100 mM)	5 ul
dTTP (100 mM)	3 ul
aa-dUTP (100 mM)	2 ul

The mixture was stored at -20 ºC.

C.1.5 Procedure for DNase Treatment (PROMEGA)

The following constituents and reagents were mixed together as follows:

RNA	25 ug
Dnase Buffer (Promega)	8 ul
$MgCl_2$	7.2 ul
DNase (Promega)	4 ul
Water to	25 μl

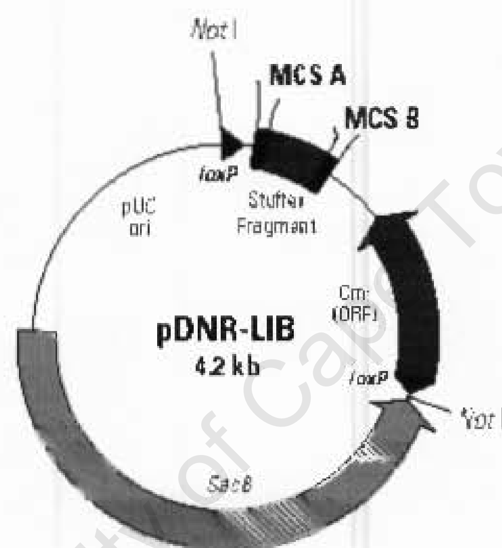
The reaction mix was left to incubate at 37 °C for 1 hr following which 2 μl stop buffer was added and left to incubate at 65 °C for 10 min.

University of Cape Town

APPENDIX D

CLONING VECTORS

D.1 pDNR-LIB – Cloning vector



MCS A

45

loxP

TTA TCA GTC GAC GGT ACC GGA CAT ATG CCC GGG AAT TCG GCC ATT ACG GCC TGC AGG ATC C [Stuffer fragment]

SalI NdeI SmaI EcoRI SfiI PstI BamHI

MCS B 282

[Stuffer fragment] GG ATC CGG CCG CCT CGG CCC TCG AGA AGC TTT CTA GAC CAT TCG TTT GGC GCG CGG GCC CAG TAG GTA ACT GAA

BamHI SfiI XhoI HindIII XbaI Bsp120 I ApaI stops

Diagram from Creator™ SMART™ cDNA Library Construction kit manual, (CLONTECH Laboratories, Inc.).

University of Cape Town

APPENDIX E

R PROGRAMMING LANGUAGE USED TO ANALYSE MICROARRAY DATA

C.1 R SCRIPT USED TO SPECIFY LIMMA PACKAGE STATISTICAL ANALYSIS.

```
## Data input from specified working directory on Computer.
setwd("C:/Documents and Settings/Tanya/Desktop/Microarray Results/R Work
Analysis Space/GPR files")

## Instillation of limma package and target folder from working directory
library(limma)
targets <- readTargets()
targets

## Read Image Files incorporating weighting of flagged spots into a data object
in R
RG <- read.maimages(targets$FileName, source="genepix",wt.fun=wtflags(0))
RG

## Read gene information (GAL file)
RG$genes <- readGAL()
RG$genes[1:20,]

## Set print layout
RG$printer <- getLayout(RG$genes)
spottypes<-readSpotTypes()
RG$genes$Status<-controlStatus(spottypes,RG)

## Diagnostic analysis of raw data - MA-plots of raw data for each array
par(mfrow=c(2,2))
plotMA(RG,array=1,ylim=c(-5,5),zero.weights=FALSE)
plotMA(RG,array=2,ylim=c(-5,5),zero.weights=FALSE)
plotMA(RG,array=3,ylim=c(-5,5),zero.weights=FALSE)
## zero.weights = FALSE excludes spots with a zero/-ve weight - Spots are not
printed by default.

## Diagnostic analysis of raw data - Density plot
MA.b<-normalizeWithinArrays(RG,method="none",bc.method="none",controlspots=200)
plotDensities(MA.b)

## Normalize within arrays and background correct
MA.p<-
normalizeWithinArrays(RG,method="robustspline",bc.method="normexp",controlspots
=200,df=4)
plotDensities(MA.p)

## MA plots AFTER Within Array normalization
x11()
par(mfrow=c(2,2))
plotMA(MA.p,array=1,ylim=c(-5,5))
plotMA(MA.p,array=2,ylim=c(-5,5))
plotMA(MA.p,array=3,ylim=c(-5,5))
```

University of Cape Town

```
## Normalize between Arrays
MA.pAq<-normalizeBetweenArrays(MA.p,method="Aquantile")
plotDensities(MA.pAq)

## MA plots AFTER normalization
par(mfrow=c(2,2))
plotMA(MA.pAq,array=1,ylim=c(-5,5))
plotMA(MA.pAq,array=2,ylim=c(-5,5))
plotMA(MA.pAq,array=3,ylim=c(-5,5))

## Determination of differential expression

## sort by gene ID to get spacing regular

i <- order(MA.pAq$genes$ID)
MA.pAq<- MA.pAq[i,]

## remove controls: Bactin,rc,cc,uc,nc
MA.pAq$M<-MA.pAq$M[MA.pAq$genes$Status=="gene",]
MA.pAq$A<-MA.pAq$A[MA.pAq$genes$Status=="gene",]
MA.pAq$genes<-MA.pAq$genes[MA.pAq$genes$Status=="gene",]

#average over 4 replicate spots on each array (M and A values)
fit1<-lmFit(MA.pAq,design=diag(ncol(MA.pAq)),ndups=4,spacing=1,correlation=0)
fit2 <- lmFit(MA.pAq$A, design=diag(1, dim(MA.pAq)[2]), ndups=4,
spacing=1,correlation=0)

MA.pAq$M<-fit1$coef
MA.pAq$A<-fit2$coef
MA.pAq$weights<-NULL
MA.pAq$genes<-fit1$genes

## Design for each array specified as a modelmatrix
design<-modelMatrix(targets,ref="Enriched")

## Fitting normalized data for differential expression determination
fit<-lmFit(MA.pAq,design)
fit<-eBayes(fit)
topTable(fit,adjust="none",number=100)
toptable<-topTable(fit,adjust="none",number=100)
x<-topTable(fit,adjust="none",number=100)
x[x[ ,6]=="gene",]
xx<-x[x[,6]=="gene",]
xx[1:100,]

## Multiple comparison correction - p-value adjustment using Bonferroni
correction.
topTable(fit,adjust="bonferroni",number=100)
x<-topTable(fit,adjust="bonferroni",number=100)
xx<-x[x[,6]=="gene",]
xx[1:100,]

## Save object Table to a (TAB-delimited) text file
Table<-xx
write.table(Table,file="c:\\Table.txt",row.names=FALSE,sep="\t")
```

University of Cape Town

APPENDIX F

PUTATIVELY IDENTIFIED SEQUENCES

F.1

All conserved hypothetical proteins, unknown proteins or unidentifiable sequences are not included in this list. In addition, only one sequence representative for those sequences found to be redundant is given below.

642 – Putative aspartate aminotransferase

```
1       GGGTTCCGTC GCTAACTACA GAGACCCCTT GGCCGACCGT CTAGGCCGCC TTGGTACTGA
61      GACTGCGTAT GCCGTCGCCT TGGAGGCCGG TCAAATTAAG GCCACTGGTA AGACTGTGTA
121     TCCCTTCCAC ATCGGTGACC TCAACTTCCG CACACCCAAG GTCGTCGTGG ACGCCTGCAA
181     ACAGGCGCTG GATGATGGCT TCACTGGTTA CTGTCCCGCC GCTGGTATCC CGTCGCTGAG
241     AGCCGCTCTC GCCGACCACT ACAATCAGGT GTTCGGACTG GAGTACTCCG CTTCCAACGT
301     GTCCATCCAG AGCGGAGGAA AGCCTGGCAT TGGAAAATTC CTCATGTGTG TGTGCAATGA
361     GGGCGATGAG GTGCTTTTCC CCAGCCCGGG TTACCCCATC TACGAGAGCA TGGCTCGCTT
421     CCTCAACACC AAGCCGGTGC CATACTGCTA CAAGGAGACT GAGGACGGCT TCGTGTTGGA
481     CGTTTCTGAA ATGGAGTCCA AGATTACCAG CCGTACCAAG GCTATCTTCC TCAACAACTT
541     CCAGAACCCA ATGGGTGTGG CTCACACCAT TCAGGAATTG GAGGACATTG CGGCGCTCTG
601     CATCGAGAAG GACTTGTTCG TGTTCTCAGA TGACCCGTAC TACCAGATTA GGTTCTCGGA
661     CTTTCCGCAG AATGACTTCA TGCACATCGC CAAAATGCCT GGGAATGCTG TCGCGCAC
```

41 – Putative elongation factor alpha

```
1       GACACCCACG TCCTGTACTA CTCCCCCGCC CCCTGCTCCT TCCGAAGCCG GCTGCCTCAC
61      GTTCCACCTA TCTAGTTTGC CATCTTTTTT GCATCATGAC TGAGGGAAAG AAGCACGTCT
121     CCATCGTCAT CTGTGGACAT GTCGATGCCG GTAAATCCAC CACCACCGGC CGACTCATCT
181     TCGAGCTTGG CGGCATTTCC GATCGTGAAA TGCAGAAGCT CAAGGATGAG GCCAAGGCTC
241     TCGGAAAGGA TTCCTTCGCT TTCGCCTTCT ACATGGACAA GTCCAAGGAG GAGCGTGCTC
301     GCGGTGTCAC CATCGCCTGT ACCACCAAGG AGTTTTTCAC AGACTCCTAC CACTACACCA
361     TCATCGATGC ACCGGGCCAT CGTGATTTCA TCAAGAACAT GATTACTGGT GCCTCGCAGG
421     CTGATGTTGG TCTCCTCATG GTTCCCGCCG ATGGTAACTT CATTACCTCC ATCGCCAAGG
481     GTGACCACAA GGCTGGTCAA GTGCAGGGTC AGACTCGCCA GCATGCCCGT CTGCTCAACC
541     TTTTGGGTGC CAAGCAACTT ATCGTCGGTG TCAACAAGAT GGACTGTGAT GTCGCCAAGT
601     ACGGCAAGGA TAGATATGAT GAGATCCGCG CTGAGGTCAC CACCATGCTC ATGAAGGTCG
661     GATGGAAGAA GGATTTCATT GGCAAGAGTG TGCCGGTCAT TCCGATTTCC GGATGGATGG
721     GTGACAACCT TATCACCAAT CCAGCAAA
```

University of Cape Town

230 – Putative elongation factor 3 homolog

```
1       CATCTCGAGA AGACACCTAA TGAGTACATT CGTTGGCGGT ATGCCTATGG TGAGGATCGT
61      GAGGCCTTGC AGAAAGATAC TATGCAAGTC AGTGATGAAG AGTTGAAGCT CATGAAGACT
121     CCGATTCAGT ATGTGTGGCT GACTGAAGAC GGAAGGGAGA ATCGCGAAAA GCTTGTTATC
181     GACCGCCTCA CTACCCAACG CCGAACCAAG AAGGGAACTA AGTCCGACTA CGAGTACGAA
241     GTGAAGTGGG TCGGAAAGCC CTCTACCTTC ACTGCTTGGT TGGCTCTTGA GAAGCTTGAG
301     AAGCTAGGAT GGGGCAAGGC CGTCAAGACT GTTGATGAGA AGGTTGCTTC CCGTGCTTCG
361     GCACACACTA TGCCGCTCAC GTCCTCAAAC GTCGAGAGAC ATCTTGAGAA CGTTGGACTT
421     GATCGTGAGT TTGGAACTCA TTGCCGTCTC GGTGCTCTGT CCGGTGGTCA GAAGGTCAAG
481     GTCGCCCTTG CTGCTGCGCT TTGGAACTGC CCTCATATTA TTATTCTAGA TGAGCCTACT
541     AACTATCTTG ACCGCGATTC TCTCGTGGCT CTCGCTAACG CTATCGAGGA GTTCGATGGT
601     GGTGTTGTGA TTATTTCTCA TTACTCCGAA TTCACAGGTG AGGTGTGTCC AGAAATCTGG
661     AACTTGACTC CTGCCACTGA TACCGAGCCA GCAAGACTCA ACCTTGAAGG CGATTTGGAT
721     TGG
```

524, 381 and 3 – Putative peptidase C14 caspase catalytic subunit p20

```
1       GACCCACCAC TCCGACGGTG AGAGTGTCCC GCTGCTCGGA GTGTTGTGCT TCCGGGACTC
61      TCAAAACGCG TAGGGCGGTT GCCCTACGGT GCAGCAGACC CATATGAGGA GCATCTACCA
121     CTAGCAACAC CGAAGTGCTC AGTAATAGCG CCAGTGCAAT GCGGCGCGGT GGGCAGCCGG
181     CCCACGGGGA TTTGGATGAA CTTTTGTGTC ATCGCTCGCA TCGTGTCAGC CTGACGTACG
241     CAGTGGTATC TGCGCCGGGG CGAGCGATGC AAGGTGGCAC AACACACAAC GCGTGTGCGC
301     GACGCGGCAA CGGCGTCGGG CAACCGCTCC AAGGACACGC AGACTTGGTG AAGTGCGAGA
361     CAGCGAGCGT TGACGGCACT CGCGTCGAGT GAGCGTCTTG GGACAAGAAG GTGCGCCTGT
421     GTGACGCTCA AAGTGGCAAC GCCGTCGGAC AACCGCCCCA AGGACACAAA GACATGGTCA
481     AGTGCGTGAT AGTGAGCGCT GACGGCACTC GCATTGTGTC GGTGTCTTGG GAC
```

685 – Putative peptidase T2 asparaginase

```
1       GGCTAAACAG CTGCCTTCTT CCACCGCTCA CGTGGAGGCA GCAGTCATCT CACCGCCCGA
61      CGTGCTTCCT TCTATTCATA AACGCTGCCT AGTCTCCCCA GCACCCTGCC TCGTCCCAAC
121     GCTAGAGCTC TAGCTTACTT TGCAAATCCT TCAACAACCA CGCAAACAGC ATGACCAACC
181     TTCCTGCTCA TTCCCGGTCG CACCATCGCA CCCCAGTCAT CATTGTCCAT GGTGGCGCAT
241     GGGCTATTCC GGACGCAACA GCCGACGCCA CAGAAGCATC AGTCCGTCGT GCCGCCTCCA
301     TCGGCCATGC CTTACTGACG CAATCGCGTT CCGCGTTAGA TGCCGTCGAG GCTGCGGTGC
361     GCCATCTCGA GGAAGATCCG CTGTTCGATG CGGGTATCGG CTCCTGTCTC ACTGAGACCG
421     GTACCGTCGA GATGGATGCC GCCGTCATGT ACGATGCACC GTCGGGTCTC CGCTCCGGCG
481     CTGTAGCATG TGTGTCAAAT TCTGTACACC CTATTAGCGT GGCGAGGGGA GTGATGGAGC
541     AGACTCAGCA TTGTCTGCTG GTCGGCCGTG GAGCCGATGC GTTCTCAAAT CATATGGGAC
601     TCGGAGGTGC CTCGAGCGAT CAGTTGGTGT CTTCGGAGGC TATGAAAGAG TGGCGGAACT
661     TGCGTCGCTT TGGTAATGCT GTGCAGACGC TGTTTAACGG TCATGATACC GTTGGCGCTG
721     TCGCGATGGG ACGTGCATGG CAATTTTGCC AGCGCTACTA GTACTGGTGG TATCACTGGG
781     CAAAAAGTTG GGACGAGGTG GGGAAAATTC CCC
```

17, 19, 58, 9, 301, 207 and 537 – Putative chloroplast hypothetical protein

```
1       ACAATTAAGG GCTTATGGTG GATACCTAGG CATTTAGAAG CGATGAAGGG CGTGACTACC
61      AACGAAATTC TTCGAGGAGC TGGAAGTAAG CTATGATTCG GAGGTACCCG AATGAGGAAA
121     CTCTTTATAC TACCTACTAA ATTCATAAGT AGGCAAGAGC TAACTTGGCG AACTGAAACA
181     TCTTAGTAGC CAAAGGAAAA GAAAGCAAAC GCGATTCCCT CAGTAGTGGC GAACGAAAAG
241     GGAGAAGCCT AAACCACTTT AATTAGTTTT AGTGGGGTTG GGGGACAGTA TTTCAAAATT
301     TTGTTATGCT AGATGAATCA ATTGAAAAAT TGTACCGCAG AAGGTTATAG TCCTGTAGTC
361     GAAAGCAAAA CAATATTTTT ATTGAATCCC AAGTAGCATG GGACACGCGA AACCCCGTGT
421     GAATCAACGA GGACCACCTC GTAAGGCTAA ATATTCCTAA ATGACCGATA GTGAACTAGT
481     ACCGTGAGGG AAAAAAAAAA AAAAAAAAAA AAAAAAAAA
```

University of Cape Town

970 – Possible high light inducible protein

```
1      GAGTCAATCA GTCGAACTTC TTCACACCTT ATCCTTCAGA CCATTCCAAC TCTCTCACCT
61     TGTACCTCAA AGCAACAATG GTCTCCACAG CTTTCGTCGC CGCCACTCCG GTCCTCTCCC
121    GCACCACAAC CTCTGTTGCC TCTACATCTT TCACATCCCA CCGCACTTGC GTTAACCCAC
181    CTCGTCGTTC GGCCACGCTC ATGATGGCCG ATAAAGAGAA GATCCCCCAA GGCTTTACAG
241    CCTTCTCTGA AGTTCTCAAC GGTCGTGCCG CTATGCTCGG ACTCGTACTT GCAATCGCCA
301    CCGAGGCCAT CACCGGAAAG GGCATCATTG GCCAACTTGC TGCCCTTGGG GATATCTCCG
361    CCATTACGCA CGCATTGGGA CTTTGAGCCT GACCGCCACG CTGTGTATGC TACGTGGTGG
421    ATCTAAATTG TATGAGTAAA TAAGCTTTCC TATACCTATG TTCAGTCCAA AAAAAAAAAA
481    AAAAAAAAAA AAAAAAA
```

181 – Putative uridine phosphorylase 1

```
1      GGCCATTACG GCCGGGGAAC TCGACTTGTT TCGATATCGT AACAACCTTG CTTCTTCTAT
61     ACTCTGCTTA ATCTCATCAA GCACACGAAA GGGTACGTTC TGTCAAGAAC CATCGCACCA
121    ACTTAATTCA GCTCCTGCTC GTCAGCTATT GGCCTTTTAT TGCACTTCGA CTTCAAGGGA
181    CCCGTTTTGC GAATGCCTGC CGATCCAAAA AGCCAAGGCT CAAGCGCATG TCTGCCAACG
241    CCCAGCCTCA ATGAAGTTAT CTTGCTTCAT CTAGGCTATC GACTTCCAGC CGAGACACCG
301    TACCTGCTTC AACACCTCAA CGCGGTAAGT CACGTAATTC TCTGTGGTTC CTTTAATCGA
361    GCTCGTAAAA TAGCGGAAGA ATTCTCGTCT GAGCCAGCCA TAAACTTCTG TCGCACTGAT
421    CGATATCTGC TGCTACAACC TCTGCCTTCA GTGTTAGTTG CAGCTCATGG AATTGGAACT
481    GGATCAATAG ACGTTCTATT ACACGAGATA TATATCGCGT TGAAGGCAGC ACAGGCTTCC
541    GAATGGTGCT TTATTCGAAC AGGCAGTTGT GGTGGGCTAG GAGTTCCCCC TGGCACATTG
601    GTGGTCACAC GTCGCCCATT GGGAGGAGAC CTGACTCCCT CGCTGCGTAT GTTTGTTCTA
661    GGGGAGGACA AACGCTATCG CGCTGAGCTC GACCAGTCTC TGAGCGAATC TCTGTTTCAA
721    AGCGGA
```

281 – Putative ribulose-phosphate-3-epimerase

```
1      GGCCGACCTC AAACCCATCG CTCCCTCGAT CGCTCGCACC TCCCCCGCGC GTCGTCTTCT
61     TCCTAACTCA CAGCCACGCC ACCATCATCT CGCACACACC ACCCTCAAGT TGCGCCTTCC
121    ACAGCCAAAG ATGGGAATCA CTGCGTTCGC CGCATCATCC TCGCTGATTC AGCGCTCTGT
181    CCTCGCGCCA GCTACCAGTC AACAGCCCGC CACATGCGCT CGGCCACAGC GCACCGTGCC
241    CACAGCCGCC GTTGACGTGC CTGCTCCAGC AAAAGGTCAG GTCGTCATCT CTCCTTCCAT
301    CCTCTCGGCC GATTTTTCAA AGCTCGGAGA CGAAGTGCGT GCAATCGATA CGGCAGGATG
361    TGATTGGGTT CACGTTGATG TTATGGACGG TCGTTTTGTG CCAAACATTA CCATTGGCCC
421    ATTGGTTGTG AGTGCGCTTC GACCGGTCAC AAATAAGCCG CTTGACGTTC ATCTCATGTT
481    TGTTGAACCA GAGCTCTTGA TTCCTGAGTT TGCCAAGGCC GGTGCTGATA TCATTTCTGT
541    ACATTGTGGA GGAGAGAGTA CAATTCATCT CCACAGAACT CTGAACCTCA TCAAGGATCT
601    CGGATGCAAG GCTGGCGTTG TTCTCAACCC CGCTACTCCA GCGTCCGCCA TTGAATACGT
661    TATTGAGCTC TGCGACCTCA TACTCGTCAT GTCAGTAAAC CCCGGCTTCG GCGGACAAGG
721    TTTCATTGAC TC
```

252 – Possible cyclase dehydrase

```
1      GATCGCACTT GACGACGTGC TGTTGAGGAG CGTGGCGCGT TGACTCGTCC ACTATCTCTC
61     CGTGATAAAG TTCTCGTTTC GACTGCGGCC TGTGCTTCCT TCCTTGTCTC CGCGCGCTAC
121    GACATGACTG CCACTGCTGC GTACGTCACT CAGCCTGCGT TCCGACCGTC GCGCGTGCTA
181    CGGCGCTCCG CCAGCCCCGC CGCGCAGCGG CGAACATGCG TGATAACGAT GCGACGTGCG
241    CGCGCTGTGC AGCGTGAGTC GAGCTTATTA TCCGTGCAAG CTGCAATTCC TACCCCGCGT
301    ACGCGCGTCA CAATCGAAGC ACCGGAATCC AATCGGCGGC TGGTATCGGC CTGCAACCTC
361    ATCAACGCCC CGCTGGCAAC GGTGTGGGCA CTTCTCTCAG ACTACGCCAA CCTGGCCACG
421    CATATTCCGA ACCTGCTCAT GTCGGAGAGA AGGATGCATC CGAATGGCGG CATTCGTGTA
481    GAGCAATGTG GCGCGCAGAG CGTTTTTGGC TTTCAGTTTC GTGCCAGTTT GGTTATGGAT
541    ATGATTGAGG TCAACGCCTC GTCGTCCAAC TGGCGCGCCA TCAACTTTGA TCTTGTGTCC
601    AGTCGAGATT TTCGCCAGTT TGAGGGCGTA TGGAGGATGG AGAAGGTTGA TGACAGTAAG
661    ACCGCTTTGT ATTATACCGT CTCGATTGTA CCGAAGGGAC TTGTACCGGC CAAGGCCATT
721    GAATGGAGGA TTTCGGAAGA TGTTCCC
```

264 – Chromodomain helicase

```
1       GTATCGAAGA AGATCTTGCG AAATCGCTTC TTGACGACTG CCTGTCTCAA GCTCAAGAGG
61      CTGTTGAGAG CTCAAAAAGG AAGGGAAGGA ACTTGAACAA GGACCCCGAA TACGGTGATC
121     ATAGAATGAA TGGAAAGGAT TCAAAGTCCA AGGCTCTACG TGTGCAAATC GACATTCTTG
181     GCGAGAGGAG GGTAGATGCT CACGATCTCT TGAAGAGATG CCGAGACCTG AAGATGCTAC
241     GAGATGCTAT CGGATCATTT AATTCAGATC TGCAGTTCAG ATTGCCCGGA GTCATTAGGC
301     CACCGTCGTT TGGTATTCGA TGGAAGCGGT ATCATGACGC AATGCTGCTG GTTGGAACTT
361     GTAGGCATGG ATTCGGGAAC TGGACACAGA TCGCCAAGGA CGACCAGCTG GATTTAGGTG
421     ATAAGATGAA TGTAGCTGGT AATTCGGCTC AGGCAGGAGC GCCAGATACG ACGAAACTGG
481     CACGGCGAAT CACGGCGCTG CTCCGAGAGC TTGAGCGCGA ATCCCGCTTG CGAGCTGCTG
541     CCGATCGCAG GGGCAAATCT CGGGCCAAGG GCCAAAAGAG GCAAAGGGAC TCGTCGGCGA
601     AGGGAACCGT AGAGAGGGCC AGCAAACGTT CGAAACCCGG TCGTCCGGAC AAGAACAAGA
661     AACAGGCGAT GCGGTTAGAG ATTAAGAGAA GCAATCTAAC CAATTGCGTG AGCTGCGCTC
721     TCTCTCTAAG CAGAGCTACA TCTCGATGCC GCGGAGCGGA TTTCGAGACC AAGCAG
```

766 – Possible gamma-glutamyltransferase

```
1       GACAGACTTA AAAACTTGTT CGAAATGAAA GTGTTTAGAA GGTCAGCCGC ACAGGAAACG
61      AACTCGTTCC AAAAGGACGA GTTAAACTTA CTCCGTTAAC TAATAACTAT GTCTGCTGCT
121     CATGACGAAA ATATTGCTCT TCGCATTCCC CTTGCCTCAG CAGAAAACGC GGAAGACGAA
181     GAAGATCTTG ATGCCGATGG CATAATTCAG AATGAAAACC CACACTTTGT CCCTTTACAA
241     TCCACATATG ATGAGTTACA GGTTTGGGAA GTCGTCTCTA ACCACTTCTC TCGACTATGG
301     TCCAAAATTC CTCCGTCATG GTCTGTCGAA AGAGATCCTG CAACTGCTCT ACGCCAAGTC
361     ATGTATCTCG TAGGATTTAT AAGCTTATGT ATCTTTGTAT TTTCAGTTCC CTCCTTACTT
421     CACGTCATTT CTCCCCGTGC CCCAACCCCA CCAGGCACAG CCCCCGACTT GGATGAAGCA
481     GCTTCTATGA TAAGCGGACC AAATGGTGCT GTAGCTGCGG ATCATCCGTT GTGCAGTGAG
541     CTTGGCGTCA AGGTAATGCG AGATATGCAC GGAAACGCTG TAGACGCAAT GGTTTCAACA
601     GTGTTATGTC AAGGTGTGCT CGCTCCGTTC GCGTCTGGGC TTGGTGGGGG CGCTTTCATT
661     CTCATACATG ACACACGGTC AGCGACATCA CGGTTTTATG ACGCAAGAGA AACAGCACCT
721     ATGGCTGCTA CTCCAAAGTT ATTTACAAAC
```

405 – Mitochondrion gene for the large subunit ribosomal RNA

```
1       TACAAAGGTA AGTAACAGTT ATTTATAATA AATAATGCAG AATATAATGA TATAAACTTG
61      CTTGACAATA AGATTTATAA ATCGAATTGC GACGAAAGTC AGTCATAGTG ACCCGATATT
121     AAAGCGTGGA ATTGATATCG TTTAACAGAT AAAAGGAACT CTAGGGATAA CAGATTCATC
181     GTGATTAAGA GTTCGTATTG ACGTCACGGT TTGATACCTC GATGTCGACT CATCTTATCC
241     TGAAATTGAA GTAGATTTCA AGGGTCTAGT TGTTCGCTAG TGAAAAAGGT ACGTGAGTTG
301     GGTTCAGAAC GTCGTGAGAC AGTTCGGTCC CTATCTACTG TAAGAAAAGA AAAATAAAAA
361     GTGTATTCTT AGTACGAGAG GACCAGAATA CTCTAACCTC TGGTCGATTG ATTGTTATGC
421     CAATAGCATA GTCAAGTAGC TACGTTAGTT TTTAATAAAT ACTGAACGAA TCAAGTGTAT
481     TAAGTATTCT TTAATGTATT TTTCAAGAAT AATCTAAAAA AAAAAAAAAA AAAAAAAAAA
541     A
```

University of Cape Town

LITERATURE CITED

Alscher, R. G. and Cumming, J. R. 1990. *Stress response in plants: Adaptation and acclimation mechanisms*. Wiley-Liss. New York, Chister, Brisbane, Toronto, Singapore. ix – xii.

Anderson, R. J., Bolton, J. J., Molloy, F. J. and Rotmann, K. W. G. 2003. Commercial seaweeds in Southern Africa. In *Proceedings of the 17th international Seaweed Symposium*. Oxford University Press. 1-12.

Anderson, R. J., Smit, A. J. and Levitt, G. J. 1999. Upwelling and fish-factory waste as nitrogen sources for suspended cultivation of *Gracilaria gracilis* in Saldanha Bay, South Africa. *Hydrobiologia* **398/399**: 455-462.

Anderson, B. C., Smit, A. J. and Bolton, J. J. 1998. Differential grazing effects by isopods on *Gracilaria gracilis* and epiphytic *Cermanium diaphanum* in suspended raft culture. *Aquaculture* **169**: 99-109.

Anderson, R. J., Monteiro, P. M. S. and Levitt, G. J. 1996a. The effect of localised eutrophication on competition between Ulva lactuca and a commercial resource of *Gracilaria verrucosa* (Gracilariaceae, Rhodophyta). *Hydrobiologia* **326/327**: 291-296.

Anderson, R. J., Levitt, G. J. and Share, A. 1996b. Experimental investigations for the mariculture of *Gracilaria* in Saldanha Bay, South Africa. *Journal of applied Phycology* **8**: 421-430.

Anderson, R. J., Levitt, G. J., Keats, D. W. and Simons, R. H. 1993. The role of herbivores in the collapse of the *Gracilaria* resource at Saldanha Bay, South Africa. *Hydrobiologia* **260/262**: 285-290.

Anderson, R. J., Simons, R. H. and Jarman, N. G. 1989. Commercial seaweeds in Southern Africa: a review of utilization and research. *South African Journal of Marine Science* **8**: 277-289.

Aquaculture association of Southern Africa http://www.aasa-aqua.co.za/

Armisen, R., 1995. World-wide use and importance of *Gracilaria*. *Journal of Applied Phycology*. **7**: 231-243.

Attwood, T. K., Bradley, P., Flower, D. R., Gaulton, A., Maudling, N., Mitchell, A. L., Moulton, G., Nordle, A., Paine, K., Taylor, P., Uddin, A. and Zygouri, C. 2003. PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Research*. **31**: 400-402. http://www.biochem.ucl.ac.uk/bsm/dbbrowser/PRINTS/PRINTS.html

University of Cape Town

Azevedo, H., Lino-Neto, T. and Tavares, R. M. 2003. An improved method for high-quality RNA isolation from needles of adult Maritime pine trees. *Plant Molecular Biology Reporter*. **21**: 333-338.

Badiee, A., Eiken, H. G., Steen, V. M. and Lovlie, R. 2003. Evaluation of five different cDNA labeling methods for microarrays using spike controls. *BMC Biotechnology*. **3**: 23-27.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K. L., Marshall, M., Sonnhammer, E. L. L. 2002. The Pfam Protein Families Database. *Nucleic Acids Research*. **30**: 276-280. http://www.sanger.ac.uk/Software/Pfam/ or http://pfam.wustl.edu/ or http://www.cgr.ki.se/Pfam/

Belenghi, B., Salomon, M. and Levine. 2004. Caspase-like activity in seedlings of *Pisum sativum* eliminates weaker shoots during early vegetative development by induction of cell death. *Journal of Experimental Botony*. **55**(398): 889-897.

Bird, C. J. 1995. A review of recent taxonomic concepts and developments in the Gracilariaceae (Rhodophyta). *Journal of Applied Phycology* **7:** 255−267.

Bird, K. T. and Benson, P. H. 1987. *Seaweed ultivation for renewable resources*. pp241-276, Elsevier, Amsterdam, New York.

Bird, K. T., Habig C. and DeBusk, T. 1982. Nitrogen allocation and storage patterns in *Gracilaria tikvahiae* (Rhodophyta). *Journal of Phycology* **18**: 344-348.

Bohnert, H. J., Gong, Q., Li, P. and Ma, S. 2006. Unravelling abiotic stress tolerance mechanisms – getting genomics going. *Current Opinions in Plant Biology*. **9**: 180-188.

Branch, M. and Branch, G. 1981. *The Living Shores of Southern Africa*. pp251-261, C.Struik, Cape Town.

Buschmann, A. H., Correa, J. A., Westermeier, R., Hernández-Gondález, M. C. and Norambuena, R. 2001. Red algal farming in Chile: a review. *Aquaculture*. **194:** 203-220.

Bzowska, A., Kulikowska, E. and Shugar, D. 2001. Purine nucleoside phosphorylases: properties, functions and clinical aspects. *Pharmacology and Therapeutics*. **88**: 349-425

Cao, D., Pizzorno, G. 2004. Uridine phosophorylase: an important enzyme in pyrimidine metabolism and fluoropyrimidine activation. *Drugs Today*. **40**: 431-443

Chan, C. X., Teo, S. S., Ho, C. L, Othman, R. Y. and Phang, S. M. 2004. Optimization of RNA extraction from *Gracilaria changii* (Gracilariales, Rhodophyta). *Journal of applied Phycology*. **16**: 297-301.

University of Cape Town

Chapman, V. J. and Chapman, D. J. 1980. Gracilaria (Rhodophyta, Gracilariales): An Economically Important Agarophyte. In *Seaweeds and their uses* (3rd edition), pp89-107, Chapman and Hall, USA, London & New York.

Chapman, V. J. and Chapman, D. J. 1980. Introduction: Seaweed Resources. In *Seaweeds and their uses* (3rd edition), pp1-6, Chapman and Hall, USA, London & New York.

Chapman, V. J. and Chapman, D. J. 1980. Mariculture of seaweeds. In *Seaweeds and their uses* (3rd edition), pp241-252, Chapman and Hall, USA, London & New York.

Chiang, Y. M. 1981. Cultivation of *Gracilaria* (Rhodophyta, Gigartinales) in Taiwan. In *Proceedings of the Xth International Seaweed Symposium* (Levring, T., ed), pp570-574, Walter de Gruyter and Co., Berlin.

Chichkova, N. Y., Sang, H. K., Titova, E. S., Kalkum, M., Morozov, V. S., Rubtsov, Y. P., Kalinina, N. O., Taliansky, M. E. and Vartapetian, A. B. 2004. A plant caspase-like protease activated during the hypersensitive response. *Plant Cell.* **16**(1): 157-171.

Chomczynski, P. and Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry.* **162**: 156-159.

Crawford, N. M. and Glass, A. D. M. 1998. Molecular and physiological aspects of nitrate uptake in plants. *Trends in plant science reviews.* **3**: 389-395.

Critchley, A. T. 1993. *Gracilaria* (Rhodophyta, Gracilariales): an economically important agrophyte. In A. T. Critchley, & M. Ohno, *Seaweed cultivation and marine ranching* (1st edition), pp89-112, Japan International Cooperation Agency (JICA).

Danon, A., Botari, V. I., Gordon, A., Mailhac, N. and Gallois, P. 2004. Ultraviolet-C overexposure induces programmed cell death in *Arabidopsis*, which is mediated by caspase-like activities and which can be suppressed by caspase inhibitors, p35, and defender against apoptotic death. *Journal of Biological Chemistry.* **279**(1): 779-787.

Dawes, C. D. 1995. Suspended cultivation of *Gracilaria* in the sea. *Journal of Applied Phycolology.* **7**: 303-313.

De Oliveira, E. C. and Plastino, E. M. 1994. Gracilariaceae. In *Biology of Economic algae* (Akatsuka, I., ed), pp185-226, SPB Academic Publishing, The Hague, Netherlands.

Dopazo, J., Zanders, E., Dragoni, I, Amphlett, G. and Falciani, F. 2001. Methods and Approaches in the analysis of gene expression data. *Journal of Immunological Methods.* **250**: 93 – 112.

University of Cape Town

Draper, J., Scott, R., Armitage, P. and Walden, R. 1988. Vectors of the transformation of plant cell using agarobacterium. In *Plant genetic transformation and gene expression*, A laboratory manual, p47, Blackwell Scientific Publications, Oxford.

Dudoit, S., Yang, Y. H., Callow, M. J. and Speed, T. P. 2002. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*. **12**: 111-140

Dunlop, P. C. and Roon, R. J. 1975. L-Asparaginase from *Saccharomyces cerevisiae*: an extracellular enzyme. *Journal of Bacteriology*. **122**:1017-1024.

Duret, L. and Abdeddaim, S. 2000. Multiple alignments for structural, functional, or phylogenetic analyses of homologous sequences. In *Bioinformatics: Sequence, structure and databanks* (Higgins, D. and Taylor, W., eds), pp51 - 76. Oxford University Press. Oxford and New York.

Dyhrman, S. T., Haley, S. T., Birkeland, S. R. and Wurch, L. L. 2006. Long serial analysis of gene expression for gene discovery and transcriptome profiling in the widespread marine coccolithophore *Emiliania huxleyi*. *Applied and Environmental Microbiology*. **72**: 252-260.

Edding, M., Macchiavello, J. and Black, H. 1987. Culture of *Gracilaria* sp. in outdoor tanks: productivity. *Hydrobiologia*. **151/152**: 369-373.

Edwards, P. and Demaine, H. 1998. Rural Aquaculture: Overview and Framework for Country Reviews, Food and Agriculture Organisation, United Nations, Regional office publication, Bangkok, Thailand.

Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J. A., Hofmann, K. and Bairoch, A. 2002. The PROSITE database, its status in 2002. *Nucleic Acids Research*. **30:** 235-238. http://www.expasy.ch/prosite/ or http://www.isrec.isb-sib.ch/profile/profile.html

Ferrario-Méry, S., Besin, E., Pichon, O., Meyer, C. and Hodges, M. 2006. The regulatory PII protein controls arginine biosynthesis in *Arabidopsis*. *FEBS Letters*. **580**: 2015-2020.

Friedlander, M. and Levy, I. 1995. Cultivation of *Gracilaria* in outdoor tanks and ponds. *Journal of Applied Phycology*. **7**: 315-324.

Gebrekiros, S. T. 2003. Characterisation of nitrogen stress response genes of the marine alga *Gracilaria gracilis*. MSc Thesis. University of Cape Town, South Africa.

Green, P. J. 1994. The Ribonucleases of higher plants. *Annual Review Plant Physiology*. **45**: 421-445.

University of Cape Town

Guena, F., Hartings, H. and Scienza, A. 1998. A new method for rapid extraction of high-quality RNA from recalcitrant tissues of grapevine. *Plant Molecular Biology Reporter*. **16**: 61-67.

Guo. D., Cutri, B. and Catchpoole, D. R. 2004. The influence of RNA integrity, purity and cDNA labelling on glass slide cDNA microarray image quality. *Genome Biology*. **5**: 13-38.

Hanisak, M. D. 1990. The use of *Gracilaria tikvahiae* (Gracilariales, Rhodophyta) as a model system to understand the nitrogen nutrition of cultured seaweeds. *Hydrobiologia* **204/205**: 79-87.

Heddad, M. and Adamska, I. 2000. Light stress-regulated two-helix proteins in *Arabidopsis thaliana* related to the chlorophyll a / b-binding gene family. *The Proceedings of the National Academy of Science*. **97**: 3741-3746.

Hegde, P., Qi, R., Abernathy, K. Gay, C., Dharap, S., Gaspard, R., Earle-Hughes, J., Snesrud, E., Lee, N. and Quackenbush, J. 2000. A concise guide to cDNA microarray analysis. *Biotechniques*. **29**: 548-562

Hildebrand, M. and Dahlin, K. 2000. Nitrate transporter genes from the diatom *Cylindrotheca fusiformis* (Bacillariophyceae): mRNA levels controlled by nitrogen source and by the cell cycle. *Journal of Phycology*. **36**: 702-713.

Hugot, K., Ponchet, M., Marais, A., Ricci, P. and Galiana, E. 2002. A tobacco S-like RNase inhibits hyphal elongation of plant pathogens. *Molecular Plant Microbe Interactions*. **15**: 243-250.

Huguenin, J.E. 1976. An examination of problems and potentials for future large-scale intensive seaweed culture systems. *Aquaculture*. **9:** 313-342.

Hutchinson, C. R. 2003. Polyketide and non-ribosomal peptide synthases: falling together by coming apart. *The Proceedings of the National Academy of Science*. **100**: 3010-3012

Ikeda, Y. and Taniguchi, N. 2005. Gluthione transferases and gamma-glytamyl transpeptidases. *Methods in Enzymology*. **401**: 408-425.

Inokuchi, R., Kuma, K., Miyata, T. and Okada, M. 2002. Nitrogen-assimilating enzymes in land based plants and algae: phylogenic and physiological perspectives. *Physiologia Plantarum*. **116**: 1-11.

Isaac, W. E. 1956. The ecology of *Gracilaria confervoides* (L.) Grev. in South Africa with special reference to its ecology in the Saldanha-Langebaan Lagoon. In *The Second International Seaweed Symposium, Trondheim*, (Braarud, T. and Sørensen, N. A., eds), pp173-185, Pergamon, London.

University of Cape Town

Jackson, D. J. and Degnan, B. M. 2006. Expressed sequence tag analysis of genes expressed during development of the tropical abalone *Haliotis asinine*. *Journal of shellfish research*. **25**: 225-231.

Jaffray, A. E., Anderson, R. J. and Coyne, V. E. 1997. Investigation of bacterial epiphytes of the agar-producing red seaweed *Gracilaria gracilis* (Stackhouse) Steentoft, Irvine *et* Farnham from Saldanha Bay, South Africa, and Luderitz, Namibia. *Botanica Marina*. **40**: 569-576.

Jaffray, A. E. and Coyne, V. E. 1996. Development of an *in situ* assay to detect bacterial pathogens of the red alga *Gracilaria gracilis* (Stackhouse) Steentoft, Irvine *et* Farnham. *Journal of Applied Phycology*. **8**: 409-414.

Juhnke, H., Krems, B., KoÈtter, P. and Entian, K. D. 1996. Mutants that show increased sensitivity to hydrogen peroxide reveal an important role for the pentose phosphate pathway in protection of yeast against oxidative stress. *Molecular and General Genetics*. **252**: 456-464.

Kerr, M. K. and Churchill, G. A. 2001, Statistical design and analysis of gene expression microarray data. *Genetics Research Cambridge.***77:** 123-128.

Kooperberg, C., Sipione, S., LeBlanc, M., Strand, A. D., Cattaneo, E. and Olson, J. M. 2002. Evaluating test statistics to select interesting genes in microarray experiments. *Human Molecular Genetics*. **11**: 2223-2232

Kopp, J., Kopriva, S., Suess, K. H. and Schultz, G. E. 1999. Structure and Mechanism of the Amphibolic Enzyme D-Ribulose-5-phosphate 3-Epimerase from Potato Chloroplasts. *Journal of Molecular Biology*. **287**: 761-771.

Korf, I., Yandell, M. and Bedell, J. 2003. *BLAST* (1st edition), pp1-70, O'Reilly and Associates, Inc. United States.

Krutovskii, K. V. and Neale, D. B. 2001. Forest genomics for conserving adaptive genetic diversity. *Forest Genetic Resources Working Papers*. Working Paper FGR/3.

Kuehn, G. D. and Phillips, G. C. 2005. Role of Polyamines in apoptosis and other recent advances in plant polyamines. *Critical Reviews in Plant Sciences*. **24**: 123-130.

Kusian, B., Yoo, J. G., Bednarski, R. and Bowien, A. B. 1992. The Calvin cycle enzyme pentose-5-phosphate 3-epimerase is encoded within the cifx operons of the chemoautotroph *Alcaligenes eutrophus*. *Journal of Bacteriology*. **174**: 7337-7344.

Lancaster, J. E. and Shaw, M. L. 1994. Characterization of purified gamma-glutamyl transpeptidase in onions: evidence for *in vivo* role as a peptidase. *Phytochemistry*. **36:** 1351–1358.

University of Cape Town

Lapointe, B. E. 1985. Strategies for pulsed nutrient supply to *Gracilaria* cultures in the Florida Keys: Interactions between concentration and frequency of nutrient pulses. *Journal of Experimental Marine Biology and Ecology*. **93**: 211-222.

Lapointe, B. E., Williams, L. D., Goldman, J. C. and Ryther, J. H. 1976. The mass outdoor culture of macroscopic marine algae. *Aquaculture*. **8:** 9-21.

Leblanc, C., Boyen, C., Richard, O., Bonnard, G., Grienenberger, J. M. and Kloareg, B. 1995. Complete sequence of the mitochondrial DNA of the Rhodophyte *Chondrus crispus* (Gigartinales). Gene content and genome organization. *Journal of Molecular Biology*. **250**: 484-495.

LeBrasseur, N. D., MacIntosh, G. C., PeÂrez-Amador, M. A., Saitoh, M. and Green, P. J. 2002. Local and systemic wound-induction of RNase and nuclease activities in Arabidopsis: RNS1 as a marker for a JA-independent systemic signaling pathway. *The Plant Journal*. **29(4)**: 393-403.

Lee, H., Vázquez-Bermúdez, M. F. and de Marsac, N. T. 1999a. The global nitrogen regulator NtcA regulates transcription of the signal transducer P_{II} (GlnB) and influences its phosphorylation level in response to nitrogen and carbon supplies in the cyanobacterium *Synechocccus* sp. strain PCC 7942. *Journal of Bacteriology*. **181**: 2697-2702.

Lee, R. E. 1999b. Evolution of the chloroplast: Rhodophyta. In *Phycology* (3rd edition), pp159-160, Cambridge University Press, Cambridge.

Lee, Y. H., Huang, G. M., Cameron, R. A., Graham, G., Davidson, E. H., Hood, L. and Britten, R. J. 1999c. EST analysis of gene expression in early cleavage-stage sea urchin embryos. *Development*. **126**: 3857-3867.

Lignell, A. and Pedersén, M. 1987. Nitrogen metabolism in *Gracilaria secundata* Harv. *Hydrobiologia* **151/152**: 431-441.

Lindell, D., Erdner, D., Marie, D., Prášil, O., Koblížek, M., Le Gall, F., Rippka, R., Partensky, F., Scanlan, D. J. and Post, A. F. 2002. Nitrogen stress response of *Prochlorococcus* strain PCC 9511 (oxyphotobacteria) involves contrasting regulation of *ntc*A and *amt*1. *Journal of Phycology*. **38**: 1113-1124.

Lobban, C. S. and Harrison, P. J. 1994. *Seaweed Ecology and Physiology* (1st edition), pp178-202, Cambridge University Press, Cambridge.

Lopes, P. F., Cabral de Oliveira, M. and Colepicolo, P. 2002. Characterization and daily variation of nitrate reductase in *Gracilaria tenuistipitata* (Rhodophyta). *Biochemical and Biophysical Research Communications*. **295**:50-54.

University of Cape Town

Luque, I., Zabulon, G., Contreras, A. and Houmard, J. 2001. Convergence of two global transcriptional regulators on nitrogen induction of the stress-acclimation gene *nbl*A in the cyanobacterium *Synechococcus* sp. PCC 7942. *Molecular Microbiology.* **41**: 937-947.

Markham, J. W. and Hagmeier, E. 1982. Observations on the effects of germanium dioxide on the growth of macro-algae and diatoms. *Phycologia.* **21(2)**:125-130.

Martinez, L. A. and Buschmann, A. H. 1996. Agar yield and quality of *Gracilaria chilensis* (Gigartinales, Rhodophyta) in tank culture using fish effluents. *Hydrobiologia* **326/327**: 341-345.

McHugh, D. J. 2003. A guide to the seaweed industry, Food and Agriculture Organisation, United Nations, Fisheries Technical Paper 441, Rome, Italy, 123 pp.

McHugh, D. J. 2002. Prospects for seaweed production in developing countries. FAO fisheries circular. No 968. Rome, Italy, 28p.

Morgan, P. W. 1990. Hormone systems. In *Stress responses in plants: Adaptation and acclimation mechanisms* (Alscher, R.G. and Cumming, J.R. eds), pp128-138, Wiley-Liss, New York.

Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A. N., Orchard, S., Pagni, M., Ponting, C. P., Quevillon, E., Selengut, J., Sigrist, C. J., Silventoinen, V., Studholme, D. J., Vaughan, R. and Wu, CH. 2005. InterPro, progress and status in 2005. *Nucleic Acids Research.* **33**: Database Issue: D201-5.

Muro-Pastor, M. I., Reyes, J. C. and Florencio, F. J. 2001. Cyanobacteria perceive nitrogen status by sensing intracellular 2-Oxoglutarate levels. *The Journal of Biological Chemistry.* **276**: 38320-38328.

Naderi, A., Ahmed, A. A., Morias, N. L. B., Aparicio, S., Brenton, J. D. and Caldas, C. 2004. Expression microarray reproducibility is improved by optimizing purification steps in RNA amplification and labelling. *BMC Genomics.* **5**:9-22

Neori, A., Shpigel, M. and Ben-Ezra, D. 2000. A sustainable integrated system for culture of fish, seaweed and abalone. *Aquaculture* **186**: 279-291.

Ninfa, A. J. and Jiang, P. 2005. PII signal transduction proteins: sensors of α-ketoglutarate that regulate nitrogen metabolism. *Current Opinion in Microbiology.* **8**: 168-173.

University of Cape Town

Olafsen, J. A. 2001. Interactions between fish larvae and bacteria in marine aquaculture. *Aquaculture.* **200**: 223-247

Oliveira, E. C., Alveal, K. and Anderson, R. J. 2000. Mariculture of the agar-producing Gracilarioid red algae. *Reviews in Fisheries Science* **8(4)**: 345-377.

Patrzykat, A. and Douglas, S. E. 2003. Gone gene fishing: How to catch novel marine antimicrobials. *Trends in Biotechnology.* **21**: 362-369.

Paul, J. H. and Cooksey, K. E. 1981. Regulation of asparaginase, glutamine synthetase, and glutamate dehydrogenase in response to medium nitrogen concentrations in a Euryhaline *Chlamydomonas* species. *Plant Physiology.* **68**: 1364-1368.

Provasoli, L. 1968. Media and prospects for the cultivation of marine algae. In *Cultures and collections of algae* (Watanabe, A. and Hattori, A. eds.), pp63-75, Proc. U.S.-Japan Conf., Hakone.

Quackenbush, J., 2001. Computational analysis of microarray data. *Nature Reviews, Genetics.* **2**: 418-427.

Resnick, A. D. and Magasanik, B. 1976. L-Asparaginase of *Klebsiella aerogenes* - activation of its synthesis by glutamine synthetase. *Journal of Biological Chemistry.* **251**: 2722-2728.

Reynolds, P. H. S. 1999. Inducible control of gene expresion: an Overview. In *Inducible gene expression in plants* (Reynolds, P.H.S. ed), pp1-9, CAB International, UK Cambrige.

Rockett, J. C. and Hellmann, G. M. 2004. Confirming microarray data – is it really necessary? *Genomics.* **83**: 541- 549

Rothstein, S. J. and Sivasankar, S. 1999. Nitrate inducibility of gene expression using the nitrite reductase gene promoter. In *Inducible gene expression in plants* (Reynolds, P.H.S. ed), pp83-96, CAB International, UK Cambrige.

Rotmann, K. W. G. 1990. Saldanha Bay, South Africa: recovery of *Gracilaria verrucosa* (Gracilariales, Rhodophyta). *Hydrobiologia* **204/205**: 325-330.

Rotmann, K. W. G. 1987. The collection, utilization and potential farming of red seaweeds in Namibia. *Hydrobiologia* **151/152**: 301-305.

Salem, K. and van Waasbergen, L. G. 2004. Light Control of *hliA* transcription and transcript stability in the *Cyanobacterium Synechococcus elongates* strain PCC 7942. *Journal of Bacteriology.* **186**: 1729-1736.

University of Cape Town

Sambrook, J., Fritsch, E. F. and Maniatis, T. 1989. *Molecular cloning*. A laboratory manual, (2nd edition), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Santelices, B. and Doty, M. S. 1989. A review of *Gracilaria* farming. *Aquaculture* **78**: 95-133.

Sauer, J., Dirmeier, U. and Forchhammer, K. 2000. The *Synechococcus* strain PCC 7942 *gln*N product (Glutamine Synthetase III) helps recovery from prolonged nitrogen chlorosis. *Journal of Bacteriology*. **182**: 5615-5619.

Sharma, A. D., Gill, P. K. and Singh, P. 2003. RNA isolation from plant tissues rich in polysaccharides. *Analytical Biochemistry*. **314**: 319-321.

Shimizua, T., Inoue, T. and Shiraishi, H. 2001. A senescence-associated S-like RNase in the multicellular green alga *Volvox carteri*. *Gene*. **27**:227-235.

Shur, I. and Benayahu, D. 2005. Characterization and functional analysis of CReMM, a novel chromodomain helicase DNA-binding protein. *Journal of Molecular Biology*. **352**: 646-655.

Silvente, S., Camas, A. and Lara, M. 2003. Molecular cloning of the cDNA encoding aspartate aminotransferase from bean root nodules and determination of its role in nodule nitrogen metabolism. *Journal of Experimental Biology*. **54**: 1545-1551.

Simon, R. M. and Dobbin, K. 2003. Experimental Design of DNA Microarray Experiments. *Bio Techniques*. **34**:16-21.

Simon, R., Korn, E. L., McShane, L. M., Radmachyer, M. D., Wright, G. W. and Zhao, Y. 2003. *Design and analysis of DNA microarray investigations*. pp1-55, Springer, New York.

Smit, A. J. 2002. Nitrogen uptake by *Gracilaria gracilis* (Rhodophyta): Adaptations to a temporally variable nitrogen environment. *Botanica Marina*. **45**:196-209.

Smit, A. J. and Bolton, J. J. 1999. Organismic determinants and their effect on growth and regeneration in *G. gracilis*. *Journal of Applied Phycology*. **11**: 293-299.

Smit, A. J., Robertson, B. L. and du Preez D. R. 1997. Influence of ammonium-N pulse concentrations and frequency, tank condition and nitrogen starvation on growth rate and biochemical composition of *Gracilaria gracilis*. *Journal of Applied Phycology* **8**: 473-481.

Smyth, G. K. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*. **3**, No. 1, Article 3.

University of Cape Town

Smyth, G.K. and Speed, T. 2003. Normalization of cDNA Microarray Data. *Methods* **31**: 265-273.

Smyth, G. K., Yang, Y. H. and Speed, T. 2002. Statistical issues in cDNA microarray data analysis. *Methods in Molecular Biology*. **244**: 111-136.

Soler, J. M. P., da Rosa, F. H. F. P., Chiavegatto, S., Aneas, I. and Krieger, J. E. 2004. Use of splines for normalization of microarray gene expression data. *Bioscience Journal*, Especial: 101-116.

Storozhenko, S., Belles-Boix, E., Babiychuk, E., Hérouart, D., Davey, M. W., Slooten, L., Van Montagu, M., Inzé, D. and Kushnir, S. 2002. γ-Glutamyl transpeptidase in transgenic Tobacco Plants. Cellular localization, processing, and biochemical properties. *Plant Physiology*. **128**:1109-1119.

t Hoen P. A. C., de Kort, F., van Ommen G. J. B. and den Dunnen, J. T. 2002b. Flourescent labelling of cRNA for microarray applications. *Nucleic Acids Research*. **31, No.5**: e20.

Taniguchi, M., Kobe, A., Kato, M. and Sugiyama, T. 1995. Aspartate aminotransferase in *Panicum miliaceum* L., an NAD-malic enzyme type C_4 plant: comparison of enzymatic properties, primary structures, and expression patterns. *Archives of Biochemistry and Biophysics*. **318**: 295-306.

Temple, S. J., Vance, C. P. and Gantt, J. S. 1998. Glutamate synthetase and nitrogen assimilation. *Trends in Plant Science reviews*. **3**: 51-56.

Treshow, M. 1970. *Environment and Plant Response*. McGraw-Hill publishers. New York, St Louis, San Francisco.

Troell, M., Halling, C., Nilsson, A., Buschmann, A. H., Kautsky, N. and Kautsky, L., 1997. Integrated marine cultivation of *Gracilaria chilensis* (Gracilaria, Rhodophyta) and salmon cages for reduced environmental impact and increased economic output. *Aquaculture* **156**: 45-61.

Vasquez, J. A. and Guerra, N. 1996. The use of seaweeds as bioindicators of natural and anthropogenic contaminants in northern Chile.*Hydrobiologia*. **326/327:** 327-333.

Vasquez, J. A. and Westermeier, R. 1993. Limiting factors in optimizing seaweed yield in Chile. *Hydrobiologia*. **260/261**: 313-320.

Yamada, T., Fukuda, T., Tamura, K., Furukawa, S. and Songsri, P. 1993. Expression of the gene encoding a translational elongation factor 3 homolog of *Chlorella Virus* CVK2. *Virology*. **197**: 742-750.

University of Cape Town

Yang, Y. H., Buckley, M. J. & Speed, T. P. 2001b. Analysis of cDNA microarray images. *Briefings in Bioinformatics*. **2**: 341-349.

Yang, Y. H., Dudoit, S., Luu, P. and Speed, T. P. 2001. Normalization for cDNA microarray data. In *Microarrays: Optical Technologies and Informatics* (Bittner, M.L., Chen, Y., Dorsel, A.N., and Dougherty, E.R. eds), Volume 4266 of proceedings of SPIE.

Ye, J., Dickens, M. L., Plater, R., Li, Y., Lawrence, J and Strohl, W. R. 1994. Isolation and sequence analysis of polyketide synthase genes from the daunomycin-producing *Streptomyces* sp. Strain C5. *Journal of bacteriology*. **176**: 6270-6280

Yona, G and Brenner, S. E. 2000. Comparison of protein sequences and practical database searching. In *Bioinformatics: Sequence, structure and databanks* (Higgins, D. and Taylor, W. eds), pp. 167– 190, Oxford University Press. Oxford and New York.

Yu, J., Othman, M. I., Farjo, R., Zareparsi, S., MacNee, S. P., Yoshida, S. and Swaroop, A. 2002. Evaluation and optimization of procedures for target labeling and hybridization of cDNA microarrays. *Molecular vision*. **8**: 130-137.

Zeng, Y. and Yang, T. 2002. RNA isolation from highly viscous samples rich in polyphenols and polysaccharides. *Plant molecular Biology Reporter*. **20**: 417a-417e.

Zhila, N. O., Kalacheva, G. S. and Volova, T. G. 2005. Influence of nitrogen deficiency on the biochemical composition of the green alga *Botryococcus*. *Journal of applied Phycology* **17**: 309-315.

University of Cape Town