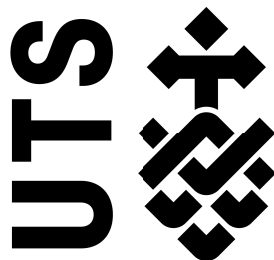


Delving into the Genetic Code of *Gambierdiscus* – the Devil is in the Detail

Thesis by
Anna Liza Kretzschmar

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy



University of Technology Sydney
Sydney, NSW, Australia

2019

© 2019

Anna Liza Kretschmar

All Rights Reserved

*Little one better find your way now,
Devil is fine,
Little one better find your way out,
Devil is kind,
Little one better run for your life,
Devil is fine,
Little one where you going with that knife?*

- Zeal & Ardor

Certificate of Original Authorship

I, Anna Liza Kretzschmar, declare that this thesis is submitted in fulfillment of the requirements for the award of Doctor of Philosophy, in the Faculty of Science at the University of Technology Sydney. This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis. This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Production Note:

Signature removed prior to publication.

Date: 2019-07-15

Anna Liza Kretzschmar

Acknowledgments

This thesis has been a harrowing journey through supervisory change, chronic mental and physical illness, and encountering road blocks rather than the mentorship which one would hope for at crucial junctions for becoming a scientist adherent to the scientific principles of transparency and objectivity.

The fact that I am still here, and smiling, is due to the love and support I was lucky enough to receive. A monumental thank you to my supervisor Aaron Darling for staying with me through this journey through the hard times, and doing so in ever present good humor.

Thank you to my partners, past and present, Kurtis Crisafulli and Leroy Jeffrey, for your love, hugs and making me remember there is more to life.

Thank you to Samantha Goyen, for having an excellent time full of laughing fits and picnics while on day and night long writing benders.

While unusual, it has been requested that I include the following section:

Acknowledgment to Shauna Murray for scholarship funding which amounted to 17.3 % of my candidature. Without your input I would not be the robust person and scientist I am today.

All other funding sources will be acknowledged with gratitude within the relevant chapters.

A. Liza Kretzschmar
Sydney, Australia, 2019.

Thesis Format

A thesis by compilation.

List of Publications

Peer reviewed

1. Murray, S. A., Ajani, P.; Kretzschmar, A. L. & Verma, A. (2017) *Response to” More surprises in the global greenhouse: Human health impacts form recent toxic marine aerosol formulations, due to centennial alterations or world-wide coastal food web”s.* Mar. pol. bulletin. 123(1-2):415-417.
2. Kohli, G.; Haslauer, K.; Chowdhury, S.; Boulter, M.; Kretzschmar, A. L.; Harwood, T.; Laczka, O. & Murray, S. (2017) *Qualitative and Quantitative Assessment of the Presence of Ciguatoxin, P-CTX-1B, in Spanish Mackerel (Scomberomorus commerson) from waters in NSW (Australia).* Toxicol. reports 4:328-334.
3. Kretzschmar, A. L.; Verma, A.; Harwood, T.; Hoppenrath, M. & Murray, S. (2017) *Characterisation of Gambierdiscus lapillus sp. nov. (Gonyaulacales, Dinophyceae): A new toxic dinoflagellate from the Great Barrier Reef (Australia).* J Phycol 53(2):283-297. **Awarded cover image.**

Under review

4. Hoppenrath, M.; Kretzschmar, A. L.; Kaufmann, M.; & Murray, S. (2018) *Morphological and molecular phylogenetic identification and record verification of Gambierdiscus excentricus (Dinophyceae) from Madeira Island (NE Atlantic Ocean).* In review, Mar. Biodiv. Records (2018).
5. Kretzschmar, A. L.; Larsson, M.; Doblin, M.; & Murray, S. *Characterisation of two toxic Gambierdiscus spp. (Gonyaulacales, Dinophyceae) from the Great Barrier Reef (Australia): G. lewisii sp. nov. and G. holmesii sp. nov.* In review, Protist (2018).
6. Kretzschmar, A. L.; Verma, A.; Kohli, G.; & Murray, S. *Development of a quantitative PCR assay for the detection and enumeration of a potentially ciguatoxin-producing dinoflagellate, Gambierdiscus lapillus (Gonyaulacales, Dinophyceae).* In review, PLoS ONE (2019).
7. Kretzschmar, A. L.; Verma, A.; Murray, S.; Kahlke, A; Fourment, M. & Darling, A. *Trial by phylogenetics - Evaluating the Multi-Species Coalescent for evolutionary inference on taxa*

with high levels of paralogy (Gonyaulacales, Dinophyceae). In review, PeerJ (2019).

Contents

Certificate of Authorship	iv
Acknowledgments	v
Thesis Format	vi
List of Publications	vii
Abstract	1
1 General introduction.	2
1.1 Ciguatera fish poisoning	3
1.1.1 Symptoms of ciguatera	3
1.1.2 Polyketide toxins produced by <i>Gambierdiscus</i> spp.	4
1.1.2.1 Ciguatoxins	4
1.1.2.2 Maitotoxins	7
1.1.2.3 Other toxins	8
1.1.2.4 Utilization of bioassays	9
1.2 The genus <i>Gambierdiscus</i>	10
1.2.1 Taxonomy	10
1.2.1.1 Morphology	10
1.2.2 Distribution of <i>Gambierdiscus</i> spp.	14
1.2.2.1 Geographic differences in <i>Gambierdiscus</i> distribution	14
1.2.2.2 Depth dependent variation of <i>Gambierdiscus</i> species	14
1.2.2.3 Local variation <i>Gambierdiscus</i> abundances	14
1.2.2.4 <i>Gambierdiscus</i> in Australia	15
1.2.2.5 Changes in distribution due to climate change	16
1.3 Dinoflagellate genetics	17
1.3.1 Polyketide synthase genes	18

1.3.1.1	Polyketide synthase genes in <i>Gambierdiscus</i>	19
1.4	Thesis aims	21
2	Characterization of <i>Gambierdiscus lapillus</i> sp. nov. (Gonyaulacales, Dinophyceae): A new toxic dinoflagellate from the Great Barrier Reef (Australia).	22
2.1	Authorship Declaration	22
2.2	Abstract	24
2.3	Introduction	24
2.4	Materials and methods	26
2.4.1	Specimen collection and culture conditions	26
2.4.2	Morphological analyses	28
2.4.3	Genomic DNA extraction	29
2.4.4	PCR amplification and sequencing	29
2.4.5	Sequence alignment and phylogenetic analysis	31
2.4.6	Toxin production	31
2.4.6.1	Toxicity via mouse bioassay	31
2.4.6.2	Toxin profile screening by liquid chromatography-mass spectrometry	32
2.5	Results	32
2.5.1	Morphology and taxonomical treatment	32
2.5.2	Phylogenetics	40
2.5.3	Toxicology	43
2.5.3.1	Toxicity via intraperitoneal injection	43
2.5.3.2	Toxicity via gavage	44
2.5.3.3	Toxin profile	44
2.6	Discussion	44
2.7	Acknowledgements	48
2.8	Conflict of interest disclosure	49
2.9	Supplementary Material	49
3	Development of a quantitative PCR assay for the detection and enumeration of a potentially ciguatoxin-producing dinoflagellate, <i>Gambierdiscus lapillus</i> (Gonyaulacales, Dinophyceae).	52
3.1	Authorship Declaration	52
3.2	Abstract	54
3.3	Introduction	54
3.4	Materials and methods	59
3.4.1	Clonal strains and culturing conditions	59
3.4.2	DNA extraction and species specific primer design	60

3.4.3	Evaluation of primer specificity	61
3.4.4	Evaluation of primer sensitivity	61
3.4.4.1	Gene based calibration curve	61
3.4.4.2	Cell based calibration curve	62
3.4.5	Determination of gene copies per cell for <i>G. lapillus</i>	62
3.4.6	Screening environmental samples for <i>G. lapillus</i>	62
3.5	Results	64
3.5.1	Evaluation of primer specificity	64
3.5.2	Evaluation of primer sensitivity	66
3.5.3	Quantification of SSU rDNA copy number per cell of <i>G. lapillus</i>	69
3.5.4	Screening environmental samples for <i>G. lapillus</i> abundance	69
3.6	Discussion	72
3.7	Conclusion	74
3.8	Acknowledgements	74
3.9	Supplementary Material	74
4	Using comparative transcriptomics for the pan-transcriptome of <i>Gambierdiscus</i>.	78
4.1	Abstract	79
4.2	Introduction	79
4.3	Methods	82
4.3.1	Transcriptome acquisition	82
4.3.2	Spliced leader search	83
4.3.3	Homolog clustering	83
4.3.4	Comparison of gene ontology annotations	84
4.3.5	Ketosynthase domain search	84
4.4	Results	84
4.4.1	Overview of the transcriptomes	84
4.4.1.1	Level 1 GO annotations between <i>Gambierdiscus</i> species	87
4.4.1.2	Level 2 GO annotations between <i>Gambierdiscus</i> species	91
4.4.2	Pan-transcriptomics of <i>Gambierdiscus</i>	95
4.4.2.1	Core transcriptome	95
4.4.2.2	Softcore transcriptome	95
4.4.2.3	The unique portion of the transcriptome	95
4.4.3	Level 1 GO annotations for pan-transcriptomes	96
4.4.4	Level 2 GO annotations for pan-transcriptomes	100
4.4.5	Ketosynthase active domain search	104
4.5	Discussion	106
4.5.1	Core, softcore and unique clusters	106
4.5.2	Ketosynthase domain detection	107

4.5.3	Areas for possible improvement in this study	107
4.5.3.1	Coverage of taxa	108
4.5.3.2	PKS active domain search	108
4.6	Conclusion	108
5	Trial by phylogenetics - Evaluating the Multi-Species Coalescent for evolutionary inference on taxa with high levels of paralogy (Gonyaulacales, Dinophyceae).	109
5.1	Authorship Declaration	109
5.2	Abstract	111
5.3	Introduction	111
5.3.1	Summary: Statistical nomenclature & errors this study seeks to address . .	113
5.3.2	Summary: Who/what are the Gonyaulacales?	114
5.4	Materials and methods	115
5.4.1	Culture conditions	115
5.4.2	RNA isolation, library preparation and Sequencing	115
5.4.2.1	Publicly available transcriptome libraries	115
5.4.3	Transcriptome processing scripts	115
5.4.3.1	Transcriptome assembly	116
5.4.3.2	Construction of multiple sequence alignments	116
5.4.4	Assembly analysis	116
5.4.5	Phylogenetic inferences	116
5.4.5.1	Ribosomal DNA based inference	116
5.4.5.2	Concatenated single copy gene based inference	117
5.4.5.2.1	Maximum likelihood with concatenated sequences . .	117
5.4.5.2.2	Bayesian inference with concatenated sequences . . .	117
5.4.5.2.3	Bayesian probability under the MSC	117
5.4.6	Stepping stone analysis	117
5.4.6.1	Generation of figures	117
5.4.6.2	Implementation	117
5.5	Results	119
5.5.1	Transcriptomes overview	119
5.5.2	Single copy gene search with BUSCOv2	120
5.5.3	Phylogenetic inferences	120
5.5.3.1	rDNA based phylogeny	121
5.5.3.2	Concatenated single copy gene based phylogeny inferred with ML	121
5.5.3.3	Concatenated single copy gene based phylogeny inferred with BI	122
5.5.3.4	Single copy gene based phylogeny under MSC	123
5.6	Discussion	124
5.6.1	Considerations for data set selection and pre-processing	125

5.6.1.1	Quantity of taxa in phylogenetic inference	125
5.6.1.2	Quality of transcriptome assemblies	125
5.6.1.2.1	Assembly parameters	126
5.6.1.3	Selection of paralogs to infer species evolution	126
5.6.1.4	Model selection for inference	128
5.6.2	Comparison to commonly employed models and data sets	128
5.6.2.1	Phylogenetic inference using ribosomal genes	128
5.6.2.2	Concatenating selected genes and using ML methods for species inference	131
5.6.2.3	Concatenating selected genes and using BI methods for species inference	133
5.6.3	Areas for possible improvement of this study	135
5.6.3.1	Contamination of other taxa	135
5.6.3.2	No representative genome for comparison	135
5.6.3.3	Different methods for RNA-seq	135
5.6.3.4	Total evidence phylogenetics	135
5.6.3.5	Model comparison methods	135
5.7	Conclusion	136
5.8	Acknowledgments	137
5.9	Supplementary material	137
6	Discussion	141
6.1	General discussion	141
6.1.1	Targeting the species at bottom of the food chain	141
6.1.2	Targeting toxin production potential directly	141
6.2	Significance and future directions	142
6.3	Thesis conclusion	142
7	Bibliography	143
8	Appendix.	170
8.1	Published manuscripts	170
8.1.1	Qualitative and quantitative assessment of the presence of ciguatoxin, P- CTX-1B, in Spanish Mackerel (<i>Scomberomorus commerson</i>) from waters in New South Wales (Australia)	170
8.1.2	Response to “More surprises in the global greenhouse: Human health im- pacts form recent toxic marine aerosol formulations, due to centennial al- terations or world-wide coastal food webs”	170
8.2	Submitted manuscripts	171

8.2.1	Morphological and molecular phylogenetic identification and record verification of <i>Gambierdiscus excentricus</i> (Dinophyceae) from Madeira Island (NE Atlantic Ocean).	171
8.2.2	Characterisation of two toxic <i>Gambierdiscus</i> spp. (Gonyaulacales, Dinophyceae) from the Great Barrier Reef (Australia): <i>G. lewisii</i> sp. nov. and <i>G. holmesii</i> sp. nov.	171
8.2.3	Development of a quantitative PCR assay for the detection and enumeration of a potentially ciguatoxin-producing dinoflagellate, <i>Gambierdiscus lapillus</i> (Gonyaulacales, Dinophyceae).	171
8.2.4	Trial by phylogenetics - Evaluating the Multi-Species Coalescent for evolutionary inference on taxa with high levels of paralogy (Gonyaulacales, Dinophyceae).	171

List of Figures

2.1	Australian sites in which <i>Gambierdiscus</i> have been detected.	27
2.2	Light micrographs of <i>Gambierdiscus lapillus</i> sp. nov.	33
2.3	Scanning electron micrographs of <i>Gambierdiscus lapillus</i> hypotheca	35
2.4	Scanning electron micrographs of the epitheca of <i>Gambierdiscus lapillus</i>	36
2.5	Scanning electron micrographs of <i>Gambierdiscus lapillus</i> showing the reticulate-foveate thecal ornamentation.	37
2.6	Line drawings showing the thecal tabulation pattern of <i>Gambierdiscus lapillus</i> . . .	39
2.7	Maximum likelihood phylogeny of <i>Gambierdiscus</i> species/phylotypes of the LSU D8-D10 rDNA region.	41
2.8	Maximum likelihood phylogeny of <i>Gambierdiscus</i> species/phylotypes of the SSU rDNA region.	42
2.9	Visual representation of the size difference between <i>Gambierdiscus</i> and <i>Fukuyoa</i> species.	46
2.10	Suppl. scanning electron micrographs of <i>Gambierdiscus lapillus</i> (strain HG26). . .	50
2.11	Suppl. scanning electron micrographs of <i>Gambierdiscus lapillus</i> showing sulcal and cingular details.	50
2.12	Suppl. scanning electron micrographs of <i>Gambierdiscus lapillus</i> (strain HG7) of morphological intra-strain variability.	51
3.1	Bioaccumulation of ciguatoxin through the food web.	55
3.2	<i>Gambierdiscus</i> sampling site on Heron Island, Queensland.	63
3.3	Cell based quantitative PCR standard curve.	67
3.4	Gene fragment based quantitative PCR curve.	68
3.5	<i>Gambierdiscus lapillus</i> detection on macroalgae across environmental sampling sites.	70
3.6	<i>Gambierdiscus lapillus</i> enumeration on macroalgae across environmental sampling sites.	71
4.1	Venn diagram of distribution of nucleotide clusters between the five <i>Gambierdiscus</i> species.	86

4.2	Summary of GO annotations between <i>Gambierdiscus</i> species that mapped to molecular functions at GOSUM level 1.	88
4.3	Summary of GO annotations for cellular components between <i>Gambierdiscus</i> species at GOSUM level 1.	89
4.4	Summary of GO annotations in the biological processes category, between <i>Gambierdiscus</i> species at GOSUM level 1.	90
4.5	Summary of GO annotations for biological processes between <i>Gambierdiscus</i> species at GOSUM level 2.	92
4.6	Summary of GO annotations matching to cellular components between <i>Gambierdiscus</i> species at GOSUM level 2.	93
4.7	Summary of GO annotations for molecular functions between <i>Gambierdiscus</i> species at GOSUM level 2.	94
4.8	Summary of pan-transcriptome clusters matching to cellular component GO annotations at GOSUM level 1.	97
4.9	Summary of biological processes GO annotations of pan-transcriptome at GOSUM level 1.	98
4.10	Summary of level 1 molecular function GO annotations of pan-transcriptome.	99
4.11	Summary of cellular component category GO annotations for the pan-transcriptome at level 2.	101
4.12	Summary of level 2 molecular function GO annotations for pan-transcriptome clusters.	102
4.13	Summary of biological processes annotations for the pan-transcriptome clusters at level 2.	103
4.14	Venn diagram of the ketosynthase domain clusters across five <i>Gambierdiscus</i> spp.	105
5.1	Maximum likelihood phylogenetic inference of concatenated ribosomal DNA genes.	121
5.2	Maximum likelihood phylogenetic inference of concatenated single copy gene set.	122
5.3	Bayesian phylogenetic inference of concatenated single copy gene set.	123
5.4	Bayesian phylogenetic inference of single copy gene set under multispecies coalescent.	124
5.5	Tanglegram of the single copy gene topology comparison between this study and Price et al. (2017).	128
5.6	Tanglegram comparing the topology between concatenated ribosomal DNA sequences and Bayesian inference of single copy genes under the multispecies coalescence.	130
5.7	Tanglegram comparing the topology of single copy gene inference generated with concatenated maximum likelihood vs. Bayesian inference under the multispecies coalescence.	132

5.8	Tanglegram comparing the topologies from the single copy gene set with Bayesian inference concatenated vs. under the multispecies coalescence.	134
-----	--	-----

List of Tables

1.1	Ciguatoxin congeners identified and characterized to date.	6
1.2	Maitotoxin congeners identified and characterized to date.	8
1.3	Presently characterized <i>Gambierdiscus</i> spp. and their taxonomic identifications. . .	11
2.1	<i>Gambierdiscus</i> spp. identified around Australia.	28
2.2	Primer sets used for phylogenetic analysis of <i>Gambierdiscus</i> spp.	30
2.3	Morphological measurements of <i>G. lapillus</i>	32
2.4	Morphological characteristics of sister species to <i>Gambierdiscus lapillus</i>	45
3.1	Published qPCR assays for <i>Gambierdiscus</i> and <i>Fukoyoa</i> spp.	58
3.2	Ciguatera cases reported in Queensland, 2011 to 2015	58
3.3	List of <i>Gambierdiscus</i> clonal strains used for the qPCR assay.	60
3.4	<i>G. lapillus</i> specific qPCR primer set for SSU rDNA designed in this study.	60
3.5	Quantitative PCR primer set cross reactivity.	65
3.6	Suppl. macroalgagl samples screened for <i>Gambierdiscus lapillus</i>	74
4.1	<i>Gambierdiscus</i> species transcriptomes used in this study along with associated in- formation.	83
4.2	Progression of clusters found in each <i>Gambierdiscus</i> transcriptome during processing.	85
5.1	Summary of transcriptome sequencing and assembly statistics.	120
5.2	Suppl. culturing conditions for species processed for this study.	137
5.3	Transcriptomes and associated information used for study.	138
5.4	Accession numbers for ribosomal DNA sequences.	139

Abstract

Ciguatera is a global problem but predominantly affects communities in the Pacific Island Nations, which is predicted to increase in prevalence and severity as the effects of climate change unfold. Despite the menace posed by ciguatera, very little is known about how to monitor for risk of outbreaks and report rates of the disease are estimated around 20 % at best. To address this, a global ciguatera strategy was called into action by an international body of experts and endorsed by UNESCO's Intergovernmental Oceanographic Commission, which have called for ciguatera to be considered a neglected tropical disease, which needs to be characterised as a matter of urgency.

Element 1 of the global ciguatera strategy revolves around the characterization of *Gambierdiscus* species and their distribution, as well as their toxin production. This thesis contributes to Element 1 by investigating the diversity of *Gambierdiscus* around Heron Island, a region where ciguatera is endemic, developing a monitoring tool for the species discovered there and then characterizing the evolution of the genus in the context of other closely related toxic genera as well as within the species. It contributes to the understanding of *Gambierdiscus* diversity, the species concept with description of a new species, adds information to ameliorate the scarcity of genetic data and molecular markers as well as monitoring tools for *Gambierdiscus*, and explores evolution of the Gonyaulacales.

Dissertation directed by Associate Professor Aaron E. Darling
iThree institute, Faculty of Science
University of Technology Sydney

CHAPTER 1

General introduction.

Gambierdiscus Adachi et Fukuyo is a genus of benthic harmful algal bloom (HAB) forming microalgae which are present in shallow tropical and sub-tropical marine waters, primarily associated with macroalgae, sea grass and dead corals (summarized in [99]). Some species of *Gambierdiscus* produce ciguatoxins (CTX), the causative agent of ciguatera fish poisoning (CFP) [15]. When CTX producing *Gambierdiscus* inhabit the food source of a vector, such as macroalgae for herbivorous fish, the toxin(s) enter the food chain [138]. CTXs are then passed up the food chain through bioaccumulation, a process which causes increased potency of the CTX congener through oxidation. As the abundance of CTX producing *Gambierdiscus* species is directly linked to CFP risk, an assessment of which *Gambierdiscus* species are present and in what numbers is thought to give an indication of ciguatera risk in an area [15, 105]. For this to form a functional baseline for environmental monitoring, a solid understanding of the various species of *Gambierdiscus* and their associated potential for toxin production are required. Increasing numbers of *Gambierdiscus* species have been described since 1995, and in 2018, 15 species have been characterized, and 5 unidentified but genetically distinct types are known. The genus continues to expand as sampling and characterization continue. *Gambierdiscus* have been found globally, though the species recorded differ based on location.

The dominant symptoms of CFP as well as the structure of the CTX congeners differ between the oceans. A panel of experts coordinated by the Intergovernmental Oceanographic Commissions (IOC) Intergovernmental Panel on Harmful Algal Blooms (IPHAB), as part of the United Nations Educational, Scientific and Cultural Organization, has developed a global ciguatera strategy in order to characterize the distribution of *Gambierdiscus* and the impact on the global community [105]. The strategy's aim is manifold and focuses on all aspects and impacts of *Gambierdiscus* - their distribution, toxin production and community impact. While a global issue, some of the Pacific Island Nations experience the highest number of CFP cases in the world and it is considered a critical public health issue [228]. To a lesser extent, CFP is also an endemic problem in Australia, Japan, the Canary Islands, Indian Ocean Islands, the Caribbean and the Americas [144]. In other nations, the issue of CFP risk primarily arises from imported seafood [172]. It is currently not feasible to screen for CTXs and common processing such as heating or freezing for consumption does not af-

fect the toxins [1, 138]. To understand when and where the threat of CFP originates, it is essential to identify the species that produce the toxins at the base of the food chain, and how these toxins are produced. Elucidating the species evolution of *Gambierdiscus* in the context of other closely related toxic and non-toxic genera is of interest for investigating the evolution of toxin production pathway(s). As CFP is increasing in frequency in the Pacific and in other regions, an understanding of the genetics and toxicology of these organisms is becoming increasingly important.

1.1 Ciguatera fish poisoning

CFP is an intoxication caused by the bioaccumulated presence of CTXs in a vector species, usually a type of fish, which is consumed by humans. At the base of the food chain are toxic *Gambierdiscus* cells which are ingested by the first vector species, predominantly by herbivorous fish consuming macroalgae, on which *Gambierdiscus* species are growing. The structure of CTXs is subject to oxidation during the process of digestion in fish vectors which can increase the toxin's potency 10-fold [89, 104, 144]. However a feeding study of juvenile spotted unicorn fish (*Naso brevirostris*) observed no behavioral change in the fish during long term intoxication [34]. Cases of CFP have also been reported from non-classical vectors, such as giant clams, and the filter feeder's ability to accumulate CTXs has been confirmed [230].

CFP is the most common non-bacterial seafood intoxication globally [65]. While only a small fraction of cases are believed to be reported (predicted 2 to 10% of cases globally, 20% in Australia), from a USA-based assessment the personal and economic impact of the known cases are considered to be significant [65, 144, 172]. The impacts of CFP are particularly problematic for the Pacific Island Nations, which largely rely on fish as their main source of protein [11, 141, 245].

1.1.1 Symptoms of ciguatera

CFP presents with over 175 gastrointestinal, cardiovascular and neurological symptoms and can be misdiagnosed for a range of conditions including chronic fatigue syndrome or multiple sclerosis [3, 64, 156, 243, 261]. Contracting the disease carries high risk of morbidity for the sufferer, and fatalities have been recorded and verified (e.g. 1 fatality in Australia in the last decade) [83]. Repeat exposure to CTX increases the severity of symptoms. The illness can be re-triggered by eating fish, meat, drinking alcohol or a number of other benign food stuffs even decades after symptoms abate [243]. Dominant symptoms of CFP vary between geographic regions, which could be due to the different prevalent CTX congeners [48, 142]. The correct diagnosis relies on identification of the patient having recently consumed seafood as well as an understanding of the CFP symptomology by the physician, where a universal set of diagnostic criteria still needs to be agreed on [64]. These factors have contributed to the low report rate of CFP globally.

1.1.2 Polyketide toxins produced by *Gambierdiscus* spp.

It is not yet clear which species of *Gambierdiscus* produce which toxins, or how prevalent intra-species variation of toxin production is [29, 128, 220]. Bioassays can provide an excellent indicator of the toxicity of a particular species of *Gambierdiscus*, but they provide no information regarding the precise identity of the toxin produced. In order to obtain this information, liquid chromatography-mass spectrometry (LC-MS) or tandem LC-MS/MS analysis is required [49]. Samples are extracted for separate maitotoxin (MTX) and CTX containing phases, which are then compared to a known toxin standard. *Gambierdiscus* spp. produce a suite of polyketide toxins (see below). In particular they produce MTXs, which have been detected in all species investigated bar one strain of *G. carpenteri* [92, 121, 181]. While CTXs have been linked to CFP, and have been isolated from the Pacific, Caribbean and the Indian Oceans, it has been thought that MTXs are not related to CFP [145]. A snapper (*Pagrus auratus*) feeding study found MTX uptake into the flesh of the fish [123]. Whether MTX plays a role in CFP pathophysiology warrants further investigation.

Recent reports indicate that the identification of a toxic species at the base of the food web may not be sufficient evidence for ciguatera risk. The presence of *G. polynesiensis* was considered as a strong indicator for CFP risk, until Rhodes et al. (2017) identified a non-CTX producing strain of that species which indicated intra-species variation exists in toxin profiles [220]. Chinain et al. (2010) observed a difference in CTX production depending on the culture's life cycle, as well as inter-species differences in toxin profiles [29]. Differences in the toxicity between different strains of *G. lapillus* were reported using a mouse bioassay (MBA), in which one strain was found to have an LD₅₀ (lethal dose to 50% of test subjects, a standard measure for bioassays) that was 150-fold that of another [128]. The toxins involved in this difference were not identified. Lewis et al. (2016) identified significant differences in CTX as well as MTX toxin production in different strains of the same *Gambierdiscus* species using a cell based bioassay [146].

1.1.2.1 Ciguatoxins

CTX are lipid soluble polyether ladder compounds which act as sodium channel activators [45]. They are orally effective in humans at the picomolar (pM) concentration range, causing an influx of Na⁺ ions and which cause spontaneous action potentials in the cell. These are especially active on voltage sensitive channels along the nodes of Ranvier, affecting the voltage exchange between neurons [149, 166, 174, 243].

Currently CTXs are prefixed by their origin, such that congeners from the Pacific, Caribbean and Indian Oceans are designated P-CTX, C-CTX and I-CTX respectively as they differ structurally, as do the dominant regional symptoms (Table 1.1). The congeners are further subdivided on a structural basis into type I and type II within P-CTXs, followed by all C-CTXs as type III - the molecular structures of I-CTXs have not been investigated to such an extent as to assign a

type [81, 82, 137] ([Table 1.1](#)). The CTXs P-CTX-4A, P-CTX-4B, and P-CTX-3C, which are produced by *Gambierdiscus*, were oxidized with the liver enzymes of carnivorous fish frequently involved in CFP cases, which transformed them into P-CTX-1B, 52-epi-54-deoxy-CTX-1B, 54-deoxy-CTX-1B, 2-hydroxy-CTX-3C, and 2,3-dihydroxy-CTX-3C which have been isolated from fish [104].

TABLE 1.1 CTX congeners identified to date and their associated toxicity.

Toxin	Type	Molecular weight (Da)	Source	Toxicity (LD₅₀, intraperitoneal (i.p.) mice)
P-CTX-1, P-CTX-1B	I	1111.6	Moray eel (<i>Gymnothorax javanicus</i>) [147, 179]	P-CTX-1 0.35 $\mu\text{g/kg}$ [179]; P-CTX-1B 0.25 $\mu\text{g/kg}$ [147]
P-CTX-2	I	1095.5 [147]	Moray eel (<i>Gymnothorax javanicus</i>) [147]	2.3 $\mu\text{g/kg}$ [147]
P-CTX-3	I	1095.5 [147]	Moray eel (<i>Gymnothorax javanicus</i>) [147]	0.9 $\mu\text{g/kg}$ [147]
P-CTX-4A	I	1061.6 [285]	<i>Gambierdiscus</i> sp. [285]; <i>G. polynesiensis</i> [29]	12 $\mu\text{g/kg}$ [29]
P-CTX-4B	I	1061.6 [285]	<i>Gambierdiscus</i> sp. [285]; <i>G. polynesiensis</i> [29]	20 $\mu\text{g/kg}$ [29]
P-CTX-3C	II	1023.6 [237]	<i>Gambierdiscus</i> sp. [237]; <i>G. polynesiensis</i> [29]	2.5 $\mu\text{g/kg}$ [29]
P-49-epi-CTX-3C	II	1023.6 [29]	<i>Gambierdiscus</i> sp. [237]; <i>G. polynesiensis</i> [29]	8 $\mu\text{g/kg}$ [29]
P-M-seco-CTX-3C	II	1041.6 [29]	<i>Gambierdiscus</i> sp. [237]; <i>G. polynesiensis</i> [29]	10 $\mu\text{g/kg}$ [29]
C-CTX-1	III	1141.6 [204, 271]	Horse-eye jack (<i>Caranx latus</i>) [204, 271]	3.6 $\mu\text{g/kg}$ [271]
C-CTX-2	III	1141.6 [204, 271]	Horse-eye jack (<i>Caranx latus</i>) [204, 271]	Toxic [271]
I-CTX-1	N/A	1141.6 [82]	Red bass (<i>Lutjanus bohar</i>); Red emperor (<i>Lutjanus sebae</i>) [82]	Toxic [82]

Type I P-CTXs consist of 13 rings with 60 C atoms [145, 147, 179]. The first CTX to have its structure analyzed was isolated from moray eels, and was designated P-CTX-1B or also described as CTX-1 [147, 179]. P-CTX-2 and P-CTX-3 were isolated from the same extracts but were structurally distinct and displayed different pathophysiology in MBAs [147]. P-CTX-4A and P-CTX-4B have been isolated from *G. polynesiensis* cultures, while P-CTX-1, P-CTX-2 and P-CTX-3 have only registered in vector species [29]. These observations are consistent with the oxidation of algal CTX congeners by fish liver enzymes by Ikehara et al. (2017).

The structure for type II P-CTX-3C has 13 rings and 57 C atoms and has been isolated first from *Gambierdiscus* sp. in 1993 (originally designated *G. toxicus*, however pre-dating the genus' expansion this species level designation is uncertain) [237]. In 2010, the same congener was isolated from *G. polynesiensis* [29]. Two congeners of P-CTX-3C are known and have been isolated - 49-epi-CTX-3C (or CTX-3B by Chinain et al [29]) and M-seco-CTX-3C from *Gambierdiscus* [237] and *G. polynesiensis* [29].

Structurally C-CTXs consist of 14 rings and 62 C atoms, with multiple congeners isolated from carnivorous fish, C-CTX-1, C-CTX-2, C-CTX-1127, C-CTX-1141, C-CTX-1143, C-CTX-1157 and C-CTX-1159 [148, 203, 204, 271]. As yet, no C-CTXs have been isolated from any of the endemic *Gambierdiscus* spp., however *G. excentricus* from the Canary Islands showed activity in both the CTX and MTX phases when assessed with bioassays [62].

The most recently discovered group of CTXs is from the Indian Ocean. In samples from red bass (*Lutjanus bohar*) and red emperor (*Lutjanus sebae*) where I-CTX-1, I-CTX-2, I-CTX-3 and I-CTX-4 were isolated from carnivorous fish. Congeners from the Indian Ocean have a higher molecular weight in comparison to P-CTXs and C-CTXs [24, 81, 82]. Their structures have not yet been established, preventing classification to a subtype [81, 82].

1.1.2.2 Maitotoxins

MTXs are the largest known proteinacious natural products, and are highly potent Ca^{2+} channel activators with a lethal dose of 50% of subjects subjected to the toxin (LD_{50}) $0.05 \mu\text{g/kg}$ [181, 288]. There are very few known bacterial proteinacious toxins with higher potency [181, 288]. However the Ca^{2+} activation is a secondary process of MTX activity and the primary mode of action in mammalian cells is still unclear [264]. MTX is a water soluble, polyether ladder compound. MTX was first isolated in 1976 from the gut of a herbivorous fish species with the common name 'Maito' [284]. The complex structure and stereochemistry was elucidated in the 1990s using stereoscopic studies and partial synthesis [181, 182, 195, 235, 296].

TABLE 1.2 Currently characterized MTX congeners and their associated toxicities.

Origin	Toxin	Molecular weight (Da)	Source	Toxicity (LD ₅₀ , i.p. mice)
Pacific	MTX-1	3422 [92, 181]	<i>Gambierdiscus</i> sp. [92]	0.05 µg/kg [181]
Pacific	MTX-2	3298 [92]	<i>Gambierdiscus</i> sp. [92]	0.08 µg/kg [92]
Pacific	MTX-3	1060 [92]	<i>Gambierdiscus</i> sp. [92]; <i>G. australes</i> , <i>G. pacificus</i> [217]	0.08 µg/kg for <i>G. australes</i> [217]
Atlantic	MTX-4	1645.24 [200]	<i>G. excentricus</i> [200]	N/A

There are four known analogues of MTX (Table 1.2) - designated MTX-1, MTX-2, MTX-3, isolated from a *Gambierdiscus* sp. in Queensland, Australia [92], and MTX-4. MTX-4 was isolated from *G. excentricus* in the Atlantic and the authors postulate that this congener largely contributes to the high toxicity observed from this species [200]. MTX-1 and MTX-2 are structurally larger than the third. MTX-3 is commonly found in *Gambierdiscus*, with only one strain of *G. carpenteri* not registering the congener to date [121].

1.1.2.3 Other toxins

Suites of polyketide toxins have been co-identified with CTX and MTX *Gambierdiscus* spp. - gambieric acid, gambierol, gambierone and gambieroxide [190, 226, 236, 273]. These toxins could elicit a false positive bioassay response when the operator is probing for CTX or MTX activity, hence it is necessary to verify toxicity with a toxin profile through LC-MS/MS.

Gambierol is less toxic than CTX with 80 µg/kg i.p. and 150 µg/kg orally in MBA, but fast acting with lethality setting in within 3 hrs [107]. Gambierol could play a potential role in CFP, which to date has not been investigated. Though less toxic than CTX, it is still a highly toxic substance which could bioaccumulate in the food chain [217].

Gambieric acids A, B, C and D showed potent antifungal properties but no toxicity via MBA at 1000 µg/kg i.p. and hence are not of concern as a human health hazard [190, 217].

Gambieroxide was isolated from *G. toxicus* by Watanabe et al [273]. Gambieroxide's stereostructure has been elucidated by nuclear magnetic resonance spectroscopy and LC-MS/MS, concluding that it is structurally similar to yessotoxin, a lipophilic toxin found in filter feeding shellfish [263]. Gambieroxide needs to be further investigated for structure, toxicity and bioaccumulative potential to gauge whether it could be a health threat, including if it could contribute to CFP.

Gambierone was identified from *G. belizeanus* through NMR. The toxin has no chemical structural

similarity to CTXs, the molecular weight and bioactivity is similar to CTX-3C, with less intensity of presentation [226].

1.1.2.4 Utilization of bioassays

Studies of the toxicity of species of *Gambierdiscus* have typically used several different methods to determine the toxins present. These have been chemical, such as LC-MS, or bioassays, which are based on a characterized response to the presence of toxins from a biological entity. LC-MS/MS requires the analysis of the comparative elution times of samples as compared to those of specific chemical CTX and MTX standards. The analysis of CTX and MTX toxins using LC-MS/MS requires highly skilled operators and the use of chemical standards. However these are only available in a small number of research facilities globally. For example, a study reported 54-deoxy-CTX1B in *Fukuyoa paulensis* which would have been the first identification of CTX production outside of the genus *Gambierdiscus* and would have had repercussions on CFP monitoring efforts and risk assessments [136]. However, there were several flaws with the interpretations in the study and the type of CTX they reported is generally considered to be a fish metabolite which has cast uncertainty on this report [128].

The Caribbean species *G. excentricus* was tested with a neuroblastoma cytotoxicity assay (N2A) which indicated both MTX and CTX toxicity [62]. In an earlier study, Chinain et al. (1999) detected MTX-like toxicity in *G. polynesiensis* using MBA [31]. Rhodes et al. (2014) confirmed the production of MTX-3 in *G. polynesiensis* via LC-MS/MS [217]. *G. australes* extracts isolated by Rhodes et al. (2010) indicated CTX and MTX through MBA. However, LC-MS/MS analysis did not detect any CTXs. A follow up study on the same strain with LC-MS also found no CTXs, but potent MTX-3 like compounds were detected [217, 219]. In contrast, three different *G. australes* strains isolated from French Polynesia, tested with a receptor binding assay (RBA), produced P-CTX-3C like compounds [29].

Holland et al. [90] conducted a large scale two year toxicity screening of 56 strains from 6 species of *G. belizeanus*, *G. caribaeus*, *G. carolinianus*, *G. carpenteri* and *Gambierdiscus* ribotype 2 as well as *Fukuyoa rutzleri* via a human erythrocyte lysis assay (HELA), as MTX is haemolytic [103]. HELA activity varied between strains but was registered consistently for each strain, indicating intra-species variation of MTX production but consistent MTX production within each strain [90]. Similarly, strain specific variation was observed for CTX-like toxicity using N2A of the *Gambierdiscus* ribotypes 4, 5 and 6, where strains from a ciguateric site were up to 4 fold more toxic compared to strains from a geographically closely related non-ciguateric reference site [282].

1.2 The genus *Gambierdiscus*

Our current understanding of the geographic distribution of *Gambierdiscus* species is limited by a number of factors. Sampling events are sporadic, predominantly occurring in the Pacific with relatively few sampling points in other Oceans, and samples represent single time points [15]. The genus' taxonomic structure was extensively revised in 2009 and new species continue to be discovered, while in recent years reports of inter- and intra-species variation in morphological characters that had been considered key for species identification have further obfuscated previous identifications [15, 22, 99, 128, 220].

While CFP is a prevalent disease in Australia, the causative species in Australia has not yet been identified, as no species has yet been found to consistently produce the precise analogs of CTXs (i.e. P-CTX-3B, -3C, -4A, -4B) which are currently known to be directly related to P-CTX-1B accumulation in pelagic fish in Australia [134].

1.2.1 Taxonomy

Currently 15 species and 5 unnamed ribotypes of *Gambierdiscus* have been described (Table 1.3) based on their distinct morphological and genetic characteristics. The continued discovery of species is likely as sampling continues globally. The type species *G. toxicus* Adachi & Fukuyo was discovered in 1979 and the genus was considered monotypic before the discovery of *G. belizeanus* Faust 1995 [2, 57]. The genus was revised extensively by Litaker et al. (2009), and four new species were introduced based on more precise morphological criteria than were previously used. Reports of particular species of *Gambierdiscus* from early taxonomic papers may not be certain, as they generally did not include enough morphological or genetic information to verify the identity based on current criteria.

1.2.1.1 Morphology

The main morphological characteristics for currently described *Gambierdiscus* spp. are cell size and shape, and the shape of particular thecal plates ascertained from scanning electron microscopy (SEM). The terminology used in Table 1.3 has been conformed to the proposed terms from Hoppenrath et al [97]. The initial species description detailed *Gambierdiscus* to be large (60-100 μm), armored with a distinct plate pattern, and having a fishhook shaped accessory pore. Species initially fell into two categories - antero-posteriorly compressed, also described as lenticular, or slightly laterally compressed, or globular. Further classification is based on characteristics such as plate patterns and apical pore shape. However the globular species were re-classified into their own genus *Fukuyoa*, which comprises *F. paulensis*, *F. rutzleri* and *Fukuyoa yasumotoi*. As species of the genus *Fukuyoa* produce MTXs, and a bioassay has registered CTX production in *F. rutzleri*, they will be discussed alongside species of *Gambierdiscus* where

appropriate [146].

Initially the morphological features of individual species of *Gambierdiscus* were proposed as diagnostic tools for species identification, however a number of more recent studies have found that some morphological characteristics can show considerable variation between strains. For example, Kohli et al. (2014) found that *G. toxicus* could not be definitively distinguished from *G. carpenteri* based on morphological features, as these showed some variation [121]. A detailed report of the extensive change in cell morphology and volume in a species of *Gambierdiscus*, based on life cycle stage was conducted by Bravo et al. (2014) (the strain was later described as *G. balechii* [63]), and a difference in observed characteristics was dependent on the culturing medium used [22]. Variability in thecal plate shapes as well as overall cell shape were noted for *G. lapillus* in the species description [128]. As variability in intra-species morphology in *Gambierdiscus* is extensive, molecular genetic information is also crucial to verify the identity of species of *Gambierdiscus* [128].

TABLE 1.3 Presently characterized *Gambierdiscus* spp. and their taxonomic identifications.

Species	Cell size (μm) (depth-width-length)	Morphological characteristics	References
<i>G. australes</i> M.Chinain & M.A.Faust	(72.5 $\pm$ 3.8)- (63.4 $\pm$ 5.0)- (38.7 $\pm$ 3.8)	Antero-posteriorly compressed species; narrow 1p plate; smooth cell surface; 2' rectangular shaped; smaller than <i>G. excentricus</i>	[31, 157]
<i>G. balechii</i> Fraga, F. Rodríguez & I. Bravo	(57.4 $\pm$ 0.3)- (60.4 $\pm$ 0.4)- (N/A)	Antero-posteriorly compressed species; narrow 1p plate; 2' hatchet-shaped; heavily ornamented theca	[63]
<i>G. belizeanus</i> M.A.Faust	(61.7 $\pm$ 3.1)- (60.0 $\pm$ 4.5)- (48.1 $\pm$ 4.2)	Antero-posteriorly compressed species; narrow 1p plate; different 2' plate symmetry and size; heavily reticulate-foveate cell surface	[57, 157]

<i>G. caribaeus</i> Vandersea, Litaker, Faust, Kibler, Holland & Tester	(82.2±7.6)- (81.9±7.9)- (60.0±6.2)	Antero-posteriorly compressed species; broad 1p plate; 2' rectan- gular shaped; symmetric 4''	[157]
<i>G. carolini- anus</i> Vandersea, Litaker, Faust, Kibler, Holland & Tester	(78.2±4.8)- (87.1±7.1)- (51.4±5.2)	Antero-posteriorly compressed species; broad 1p plate; 2' hatchet shaped; dorsal end 1p oblique; larger cell size than <i>G. polynesiensis</i>	[157]
<i>G. carpenteri</i> Kibler, Litaker, Faust, Holland, Vandersea & Tester	(81.7±6.4)- (74.8±5.9)- (50.2±6.1)	Antero-posteriorly compressed species; broad 1p plate; 2' rectan- gular shaped; asymmetric 4''	[157]
<i>G. cheloniae</i> Smith, Rhodes & Murray	(60)-(62)-(33)	Antero-posteriorly compressed species; 2' plate was hatchet shaped; smaller than <i>G. toxicus</i> & <i>G. carpenteri</i>	[248]
<i>G. excentricus</i> S.Fraga	(97.8±8.0)- (83.0±10.0)- (37.0±3.0)	Antero-posteriorly compressed species; narrow 1p plate; smooth cell surface; 2' rectangular shaped; cell size bigger than <i>G. australes</i> (1.5 times wider and deeper)	[157]
<i>G. honu</i> Rhodes, Smith & Murray	(51.5)-(52.9)- (33.8)	Antero-posteriorly compressed species; lenticular; smooth cells	[218]
<i>G. lapillus</i> Kretzschmar, Hoppenrath & Murray	(40.6±3.3)- (39±3.2)-(N/A)	Antero-posteriorly compressed species; narrow 1p plate; strong reticulate-foveate thecal orna- mentation; smaller than all other <i>Gambierdiscus</i> species	[128]
<i>G. pacificus</i> M.Chinain & M.Faust	(58.5±3.9)- (53.6±4.1)- (40.4±3.6)	Antero-posteriorly compressed species; narrow 1p plate; smooth cell surface; 2' hatchet shaped	[31, 157]

<i>G. polynesiensis</i> M.Chinain & M.Faust	(66.3±3.0)- (60.5±5.9)- (44.3±5.1)	Antero-posteriorly compressed species; broad 1p plate; 2' hatchet shaped; dorsal end 1p oblique; smaller cell size than <i>G. carolini-</i> <i>anus</i>	[31, 157]
<i>G. scabrosus</i> T.Nishimura, Shinya Sato & M.Adachi	(63.2±5.7)- (58.2±5.7)- (37.3±3.5)	Antero-posteriorly compressed species; narrow 1p plate; reticulate- foveate; asymmetric 4" plate	[130, 193, 194]
<i>G. silvae</i> S.Fraga & F.Rodríguez	(69±8)-(64±9)- (46±5)	Antero-posteriorly compressed; narrow 1p plate; heavily reticulate- foveate	[61, 158]
<i>G. toxicus</i> R.Adachi & Y.Fukuyo	(93.0±5.5)- (83.0±2.3)- (54.0±1.5)	Antero-posteriorly compressed species; broad 1p plate; 2' hatchet shaped; dorsal end 1p pointed	[2, 33, 157, 221]
<i>F. paulensis</i> F. Gómez, D. Qiu, R.M. Lopes & S. Lin	(50±3)-(45±2)- (56±3)	Globular species smaller than <i>F. ya-</i> <i>sumotoi</i> and larger than <i>F. rutzleri</i>	[73]
<i>F. rutzleri</i> Faust, Litaker, Vander- sea, Kibler, Hol- land & Tester	(45.5±3.3)- (37.5±3.0)- (51.6±4.9)	Globular species smaller than <i>F. ya-</i> <i>sumotoi</i> ; cell size less than 42µm	[157]
<i>F. yasumotoi</i> Holmes	(56.8±5.6)- (51.7±5.6)- (62.4±4.3)	Globular species larger than <i>F. ruet-</i> <i>zleri</i> ; cell size exceeds 42 µm	[91, 157]
Genetically described phylotypes			
<i>Gambierdiscus</i> ribotype 2	N/A	N/A	[158]
<i>Gambierdiscus</i> sp. type 2	N/A	N/A	[130, 193]
<i>Gambierdiscus</i> sp. type 3	N/A	N/A	[193]

<i>Gambierdiscus</i> sp. type 4	(65.9-72.4±4.1-4.2)-(64.5-68.9±5.0)-(N/A)	Antero-posteriorly compressed species; evenly distributed pores; larger than <i>Gambierdiscus</i> sp. type 5; 2' hatchet shaped; broad 1p plate	[282]
<i>Gambierdiscus</i> sp. type 5	(54.8±4.6)-(53.7±6.3)-(N/A)	Antero-posteriorly compressed species; evenly distributed pores; smaller than <i>Gambierdiscus</i> sp. type 4; 2' rectangular shaped; narrow 1p plate	[282]

1.2.2 Distribution of *Gambierdiscus* spp.

Gambierdiscus inhabit tropical and sub-tropical benthic areas, commonly attached to macroalgal or coral rubble substrates, though free swimming cells have been reported as well [15, 44]. *Gambierdiscus* abundance and species composition can vary even over short distances, sampling depth, and the type of substrate collected.

1.2.2.1 Geographic differences in *Gambierdiscus* distribution

Some *Gambierdiscus* species have been classified as endemic to either the Pacific or Atlantic Oceans, while others have been found globally distributed. With more extensive sampling all *Gambierdiscus* species may turn out to be cosmopolitan [15, 158]. Issues for species identification have come from a comparatively lower sampling frequency in the Atlantic ocean and Caribbean compared to the Pacific, as well as historic uncertainty of species designation to historic samples [15, 157]. Localized benthic HABs of *Gambierdiscus* species have been noted in the literature in both the Pacific and Atlantic regions [32, 41, 191, 278].

1.2.2.2 Depth dependent variation of *Gambierdiscus* species

In the Republic of Kiribati as well as the U.S Flower Garden Banks National Marine Sanctuary, the type of species present changed with the depth sampled [165, 259]. Cells have been detected at depths of 45m [259] and blooms have been recorded at depths down to 20m [162]. Yoshimatsu et al. (2016) tested the light tolerance of 2 species and 2 ribotypes of *Gambierdiscus* and estimated that the theoretical depth limit at which cells could survive is 86m [289].

1.2.2.3 Local variation *Gambierdiscus* abundances

Reports of *Gambierdiscus* cell numbers have not shown a correlation to the traditionally high concentrations associated with harmful blooms to CFP risk. In both the Pacific and Atlantic the

most commonly reported cell concentrations are between 100 to 1000 cells per gram wet weight algae, with a relatively small number exceeding 1000 cells (summarized in [15]). Sampling within the Republic of Kiribati showed comparatively low colonization of *Gambierdiscus* species on the macro algae *Halimeda* with 0 - 174 cells/g⁻¹ wet weight algae [165]. Yet 90% of fish sampled in the area, which spanned every trophic level of the local food web, were considered ciguateric as they exceeded the recommended quarantine threshold of 0.01 µg/kg⁻¹ P-CTX-1 equivalent [28, 282]. The authors postulate that *Gambierdiscus* spp. could preferentially colonize another substrate, which is supported by Parsons et al. (2015) who reported *Gambierdiscus* species specific preferences in macroalgal host colonization [213]. Xu et al. (2014) suggested that an unidentified *Gambierdiscus* sp. could be present alongside the 6 low- to medium-toxic strains reported [18, 282]. Another explanation could be that continuous low level exposure to CTX causes the concentration of CTXs to increase in fish over long periods of time. Hence for a nuanced assessment of CFP risk, factors other than just *Gambierdiscus* cell numbers may need to be considered.

There are a number of other factors which could contribute to CFP risk. The toxicity of the species at the base of the food web is essential in determining risk, however this could be subject to intra-strain variation [29, 128, 220]. Furthermore, the type of macroalgal substrate is key to determining the likelihood of CTX entering the food chain for two reasons: *Gambierdiscus* displays species specific preferential colonization [213]; and the properties of the substrate that is consumed by the vector species can influence toxicity [38]. Cruz-Rivera et al. (2006) report several of the commonly inhabited macroalgae are not palatable for herbivores, which suggests that secondary metabolites such as CTX from the protist communities residing on their surface would not enter the food chain [38]. They suggest that chemically and structurally defended macroalgae could serve as a refugium for toxic *Gambierdiscus* spp., which none the less do not contribute to CFP. Conversely, fast growing palatable hosts would serve as a consistent delivery mechanism of toxic *Gambierdiscus* spp. for bioaccumulation, while nutritionally low macroalgae are less likely to be consumed [38]. Low cell numbers in this scenario could still serve as the basis for a highly ciguateric trophic system, depending on the speed of uptake [38].

1.2.2.4 *Gambierdiscus* in Australia

Around 50% of Australia's coastline is tropical or sub-tropical, and therefore the water temperatures could contain suitable habitats for *Gambierdiscus* populations. The prevalence of CFP in Queensland (QLD), and since 2014 in New South Wales (NSW), suggests that toxic species are resident within these shorelines (e.g. [55, 56, 143]). As yet, no CTX producing species has been verified with LC-MS/MS, however several candidate species have recently been identified. Kretzschmar et al. (2017) described a new species, *G. lapillus*, from Heron Island in the Great Barrier Reef (GBR), QLD, which registered an uncharacterized peak in the CTX phase of the analysis

(Chapter 2) [128]. Larsson et al. (2018) found *G. lapillus* extracts to display activity in the CTX phase, as well as two as yet unnamed species also originating from Heron Island, which could be part of the ciguateric web [134]. For a breakdown of the intricacies between LC-MS/MS toxin profiles and bioassay-based species toxicity please refer to the toxin production [Section 1.1.2.4](#). Currently six species of *Gambierdiscus* have been recorded in Australia: *G. belizeanus* [184], *G. carpenteri* [121, 134, 250], *G. lapillus* [128, 134], *G. cf. toxicus* [80] as well as two uncharacterized species [134], as well as *F. yasumotoi* [184], from NSW and QLD. Heron Island on the GBR has been a commons sampling site due to the pristine research station on the island. Another species, *G. honu* isolated from Rarotonga, Cook Islands, may have a range reaching the Australian shores as the genetic sequence from the type species matches a short part of a LSU sequence recorded by Richlen et al. (2008) [218, 221]. The presence of *Gambierdiscus* was also recorded in a community pyrosequencing study of a site in Western Australia but only identified to genus level [122]. In the 1990s, several strains of *Gambierdiscus* were isolated and tested for toxin production, however identification to the species level is uncertain [92, 94]. Our current state of knowledge of the *Gambierdiscus* species inhabiting Australian waters is based on single or short term sampling events with limited spatial distribution, which means it is likely that other species are present at those sites and along the shorelines which were not registered.

1.2.2.5 Changes in distribution due to climate change

The impact of climate change on the prevalence and distribution of *Gambierdiscus*, and hence CFP, is of major concern [15, 70, 105, 228]. *Gambierdiscus* grow on both coral rubble and macroalgae [75, 115]. Coral rubble commonly occurs from reef destruction during extreme weather events such as cyclones, and the death of the reef allows for proliferation and expansion of macroalgae into previously occupied areas. Extreme weather events and associated reef deterioration are observed repercussions of climate change, as well as increases in ambient ocean temperature, which are both predicted to increase in frequency and severity. This is likely to change the habitable range of *Gambierdiscus* species and expand the occurrence of CFP into new regions [15, 70]. Kibler et al. (2012) investigated the temperature dependent growth tolerances of six *Gambierdiscus* species. Their findings demonstrated that *Gambierdiscus* species have varying optimal and tolerable growth temperature ranges, which indicates a potential shift in species composition as the water temperature changes [115]. An assessment of *Gambierdiscus* species composition and associated CFP risk in the Gulf of Mexico, USA, similarly predicts a change in species composition. Tester et al. (2013) found the optimal species growth correlated with annual water temperature fluctuations, and they suggested that a small increase in water temperature would substantially shift the range of the *Gambierdiscus* species [259]. Should the shift in species composition favor the dominance of toxic species, it would have substantial impact on local CFP risk. In the Western Pacific Warm pool, a prolonged shift to warmer water temperatures was linked to an increase in CFP cases [227]. Further, comparing CFP cases for three time periods

between 1973 and 2008 showed an overall 60% increase. What extent of the observed increase can be attributed to improvements of reporting in the later time period(s) is unknown [245]. Yet in the Cook Islands, since 2006 the number of reported CFP cases started to decrease, where the lowest case load in a 20 year period were reported in 2011. The authors link this decrease to the negative shift in the Pacific Decadal Oscillation [228].

The impacts of climate change on CFP outbreaks have already been observed in Australia. The Eastern Australian Current (EAC) spans most of Australia's East Coast, running north to south. The EAC has been observed to continue further south over the last century, delivering warmer water temperatures with higher salinity at more southerly sites [222]. Investigating the salinity and temperature tolerance range of native *G. carpenteri*, found in both QLD and NSW, indicated that the species' resilience to a wide range of conditions could aid their expansion into new habitat [250]. As CTX production in *G. carpenteri* has not been found by LC-MS/MS or bioassays, this species may not be directly involved in Australian CFP cases however they could still play a role in the local community structure and indirectly affect the ciguateric web [134].

While the change in *Gambierdiscus* spp. distribution along the Australian East Coast is hypothetical, the observation of CFP cases is well documented. Historically, cases of CFP tended to be caused by fish from QLD or the Northern Territory [92, 144]. However, in 2014 an outbreak of CFP occurred from Spanish Mackerel (*Scomberomorus commerson*) caught in NSW waters, originating close to the QLD/NSW border [56]. This was the first of three outbreaks from this vector to date, each further south into NSW, affecting a total of 17 people with different severity. Whether the outbreaks are linked to habitat expansion of toxic *Gambierdiscus* spp. or Spanish Mackerel populations into NSW is unclear, however it is likely that CFP outbreaks in NSW will continue [56, 118].

1.3 Dinoflagellate genetics

To understand the potential for toxin production, habitat expansion and tolerance to climate change induced conditions, among many others, exploring the genetic content of *Gambierdiscus* is of interest. However, there are a number of factors that have made this endeavor difficult - *Gambierdiscus* is part of the phylum Dinoflagellata, which have some unique structural and regulatory genetic features [188]. Dinoflagellates have unusually large genomes (between 1.5 and 245 Gbp depending on species), uniquely arranged in a permanently condensed liquid crystalline state [131, 152, 224]. Instead of genomes packaged by histones, dinoflagellate-viral-nucleoproteins play a role in chromosomal organization and their acquisition may be responsible for the loss of their histones due to toxicity from incompatibility of the two DNA packaging strategies [74, 106, 112, 224]. In the genome, up to 70% of the nucleotide base thymine is replaced by 5-hydroxymethyluracil [211, 212]. Genes from the plastid and mitochondrial genomes have predominantly been incor-

porated into the nuclear genome and genes for highly expressed proteins appear to be organized in repetitive tandem arrays as a form of expression regulation [8, 9]. These features have confounded current sequencing technologies as assembly becomes unfeasible, though headway has been made for Symbiodiniaceae genomes which are on the lower end of the dinoflagellate genomic size spectrum [132, 153, 241].

However, these same peculiar features of the dinoflagellate regulatory machinery offer an alternate avenue of inquiry - the transcriptome may offer an adequate approximation of the genomic content for many studies. Unlike most eukaryotes, the transcriptome of dinoflagellates does not undergo extensive editing between transcription and translation which indicates a potential unlinking of transcribed mRNA and the observed protein expression [9]. For example under both nitrogen and phosphorous stress, cultures of *Alexandrium monilatum* as well as *Prorocentrum minimum* showed significant biochemical changes in comparison to nutrient replete conditions. However, the change in transcriptome profile was only 0.1 to 1% [85]. Based on observations of this ilk, combined with an extremely long mRNA half life, a common conception follows that the protein translation is regulated at a post-transcriptional level and that the amount of gene expression is linked to copy numbers within tandem gene arrays, rather than the traditional eukaryotic up and down regulation [9, 151, 177, 188, 277, 293]. Some transcripts have a trans-spliced leader (dinoSL): a 22-nt sequence at the 5'-end of transcripts along with a poly-adenylated tail on the 3'-end [151, 293]. The frequency of dinoSLs in a transcript library varies and their function is as yet unknown, however they provide a definitive way of identifying the dinoflagellate origin of a transcript even from non-axenic samples [155]. This was demonstrated in case studies in a grazer-prey setting [154], a natural aquatic ecosystem [155] and in a mixed system with *Fugacium kawagutii* [132, 295].

1.3.1 Polyketide synthase genes

As discussed in previous sections, the production of polyketide toxins, the causative organisms, and their subsequent impact on health and the economy are a major concern for harmful algal research. In *Gambierdiscus* the structure of the polyketide toxins CTXs and their modes of action vary between species, and within species as well (see [Section 1.1.2](#)). Polyketides are produced as secondary metabolites which have undergone several steps of transformation from a polyketide precursor produced by polyketide synthases (PKS) [198, 215]. PKS are a family of enzymes with a number of active domains which are part of many essential pathways, such as the cell wall's fatty acid synthesis (FAS) [120]. PKS are comprised of a number of catalytic domains which synthesize or modulate the polyketide backbone. At the very least, five catalytic domains need to be present for producing a functional polyketide molecule. Required for synthesis and release of the backbone are ketosynthase (KS), acyltransferase (AT) and the acyl carrier protein (ACP); for chain modification a ketoreductase (KR), dehydrase (DR) and enoylreductase (ER) are necessary; and finally the polyketide is released from the PKS complex by thioesterase (TE) [54, 253]. There are three classical types of PKS: Type I, which contains all the catalytic domains consecutively on

one peptide; Type II, which are comprised of mono-functional units with each catalytic domain on a separate peptide which assemble into complexes; and Type III which is a smaller version of type II (eg. [54]). Type I PKS are sub-divided into two functional sub-groups - modular and iterative. In modular Type I PKS the catalytic domains are organized into modules according to each step, and that module is only used once during polyketide synthesis. In iterative type I PKS the same set of catalytic domains are located on one protein and used cyclically for chain elongation. Unsurprisingly and in keeping with their peculiar genetics, dinoflagellates cross the classical classification of PKS types. Phylogenetic inference has consistently placed dinoflagellate KS and KR domains within the type I PKS classification, yet predominantly only one active domain is found per transcript which is a feature of type II PKS [54, 116, 117, 119, 176]. Hence dinoflagellate PKS are termed type I-like [175].

Separating the PKS responsible for maintaining cellular function from PKS involved in polyketide toxin production is paramount for deciphering the toxin production pathway(s) and subsequent monitoring efforts. Studies focusing on PKS from dinoflagellates have been conducted for *Alexandrium ostensefeldii* [54], *Amphidinium* spp. [135, 187], *Azadinium spinosum* [170], *Gambierdiscus* spp. [117, 119, 198], *Hererocapsa* spp. [54, 233], *Karenia* spp. [116, 176, 232, 265], and symbiodiniacean taxa [13, 14, 132]. Comparative transcriptomics between toxic and non-toxic strains of dinoflagellates has been suggested as an approach to differentiate between PKS essential for cell functionality and PKS involved in toxin production [119].

1.3.1.1 Polyketide synthase genes in *Gambierdiscus*

Developing monitoring tools that directly target toxin production pathway(s) rather than species presence could markedly improve the identification and assessment of ciguatera risk. To pursue this avenue for CFP monitoring, an understanding of the PKS involved in CTX production must first be attained.

G. polynesiensis has been studied in order to investigate the relevant CTX production pathway(s) as this species is known to produce CTXs [117, 198]. Pawlowicz et al. (2014) recovered 22 transcripts with sequence similarity to type I-like PKS, however no FAS conserved domains were found [198]. A later study suggested the sequence depth and coverage may have been too shallow to detect these FAS conserved domains [119]. Approaching the pathway identification from a MTX perspective instead, Kohli et al. (2015) compared the MTX-1 producer *G. australes* to *G. belizeanus*, whose toxin profile did not show any MTX-1 but both strains produced MTX-3. The study compared these transcriptomes with sequencing data from the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) to analyze KS domains. They identified 5 clades and postulated that clades A and B corresponded to polyketide production, while clades C, D and E were part of common cellular pathways such as FAS. The identification of transcripts with modular single catalytic domains as well as sequential domains on a single contig, as per the classic type II PKS

type, have been identified in *Gambierdiscus* [117]. While the distinction between PKS and FAS pathways are starting to be worked out, the identification of toxin specific pathway(s) is still elusive.

1.4 Thesis aims

Gambierdiscus species are the known culprit for ciguatera outbreaks. This disease is prevalent in Australia, with no known causative species to date, and no functional understanding of the mode of toxin production or associated evolution. This thesis sought to address the knowledge gaps in the *Gambierdiscus* and ciguatera fields, exploring the genetic capacity of *Gambierdiscus* and related species inform the evolution of these species and of ciguatoxins with a focus on the species endemic to Australia. [Chapter 2](#) sought to address the diversity of *Gambierdiscus* around Heron Island in the GBR, which is a research station within a well established ciguateric web. The majority of the isolates turned out to be a novel, toxic species which were then characterized. To assess the prevalence of this new species around Heron Island, in [Chapter 3](#) a molecular, species specific probe was designed and validated. This probe was then used to explore the distribution and abundance of the new species in environmental community samples from around Heron Island. This chapter revealed the variation in rDNA gene copies, which are a common target for molecular probes. Due to the complexity of *Gambierdiscus* genetics, comparing the overall gene expression potential between toxic and non-toxic *Gambierdiscus* was considered a possible avenue for insight. [Chapter 4](#) explored the commonalities and differences between 6 transcriptomes of *Gambierdiscus*, elucidating a core set of transcripts which can aid future *de novo* sequencing efforts. Then the unique fractions of the transcriptomes were explored, as well as clustering and comparisons between ciguatoxic and non-ciguatoxic members. Due to the convoluted nature of the genetics and no clear toxin production pathways identified, [Chapter 5](#) sought to place *Gambierdiscus* within the greater context of closely related, toxic taxa - the gonyaulacales. Organisms within this order, and more broadly the phylum dinoflagellata, are ancient lineages with extensive paralogy. The chapter sought to explore their evolution with cutting edge methods robust to paralogy, and compared the resulting phylogeny to the current standards in the field for data and model selection.

CHAPTER 2

Characterization of *Gambierdiscus lapillus* sp. nov. (Gonyaulacales, Dinophyceae): A new toxic dinoflagellate from the Great Barrier Reef (Australia).

2.1 Authorship Declaration

By signing below I confirm that for the paper titled “Characterization of *Gambierdiscus lapillus* sp. nov. (Gonyaulacales, Dinophyceae): A new toxic dinoflagellate from the Great Barrier Reef (Australia))” and published by the Journal of Phycology, that:

Anna Liza Kretzschmar conceived and designed the experiment, performed the phylogenetic experiments, performed the light microscopy and cell size experiments and analyzed the data, wrote the paper, prepared figures and tables, reviewed drafts of the paper.

Arjun Verma isolated the clonal cultures of *Gambierdiscus lapillus*, prepared figures, analyzed data and reviewed the manuscript.

Tim Harwood performed the LC-MS/MS analysis and arranged for the acute toxicological assessment to be performed using a mouse bioassay, reviewed drafts of the paper.

Mona Hoppenrath performed the morphological experiments, analyzed the morphological data, wrote the morphological sections of the manuscript, reviewed the manuscript.

Shauna Murray conceived and designed the experiment, allocated funding for the research, and reviewed drafts of the paper.

Production Note:

Signature removed prior to publication.

Anna Liza Kretzschmar

Production Note:

Signature removed prior to publication.

Arjun Verma

Production Note:

Signature removed prior to publication.

Tim Harwood

Production Note:

Signature removed prior to publication.

Mona Hoppenrath

Production Note:

Signature removed prior to publication.

Shauna Murray

Date: 2019-07-15

2.2 Abstract

Gambierdiscus is a genus of benthic dinoflagellates which is found worldwide. Some species produce neurotoxins (maitotoxins and ciguatoxins) which bioaccumulate and cause ciguatera fish poisoning (CFP), a potentially fatal food-borne illness that is common worldwide in tropical regions. The investigation of toxigenic species of *Gambierdiscus* in CFP endemic regions in Australia is necessary as a first step to determine which species of *Gambierdiscus* are related to CFP cases occurring in this region. In this study we characterised five strains of *Gambierdiscus* collected from Heron Island, Australia, a region in which ciguatera is endemic. Clonal cultures were processed using (i) light microscopy; (ii) scanning electron microscopy; (iii) DNA sequencing based on the nuclear encoded ribosomal 18S and D8-D10 28S regions; (iv) toxicity via mouse bioassay and; (v) toxin profile as determined by Liquid Chromatography-Mass Spectrometry. Both the morphological and phylogenetic data indicated that these strains represent a new, toxic species of *Gambierdiscus*, *Gambierdiscus lapillus* sp. nov (plate formula Po, 3', 0a, 7'', 6c, 7-8s, 5''', 0p, 2'''' and distinctive by size and hatchet shaped 2' plate). Culture extracts were found to be toxic using the mouse bioassay. Using chemical analysis, it was determined that they did not contain maitotoxin (MTX1) or known algal-derived ciguatoxin analogs (CTX3B, 3C, CTX4A, 4B), but that they contained putative MTX3, and likely other unknown compounds.

2.3 Introduction

Gambierdiscus Adachi et Fukuyo is a genus of benthic dinoflagellate genus, epiphytic on many substrates in shallow tropical and sub-tropical waters [100]. Species of *Gambierdiscus* have been found to produce the ciguatoxins (CTX), which are potent neurotoxins thought to be responsible for ciguatera fish poisoning (CFP) [15]. Since the discovery of the type species *G. toxicus* R.Adachi & Y.Fukuyo in 1977 [286], extensive research has established that certain species of the genus are the primary sources of CTXs, the causative agent of CFP [33, 91]. Maitotoxins (MTXs) are also commonly produced, however their role in CFP is yet to be established [123].

For over 15 years, *Gambierdiscus* was considered to be a monotypic taxon, but since 1995 new species have been discovered with more extensive sampling in the Atlantic Ocean, Indian Ocean and Pacific [31, 57, 62, 91, 157, 194]. Currently 11 species, and six unnamed clades of *Gambierdiscus* have been described based on their distinct morphological and genetic characteristics [2, 31, 57, 62, 157, 194]. The genus *Fukuyoa* F. Gómez, D. Qiu, R.M. Lopes & S. Lin was separated from *Gambierdiscus* in 2015 [73]. *Fukuyoa* spp. produce MTX [90, 91] and hence may contribute to CFP [123], they will be referred to alongside *Gambierdiscus* where appropriate. Ascertaining the distribution of species and their toxin production is essential for assessing risk of CFP to local populations as well as the seafood industry.

CFP is the most common non bacterial illness associated with seafood consumption [65], how-

ever the correct diagnosis of the illness is difficult due to extensive list of neurological, gastrointestinal and cardiovascular symptoms [243]. The United Nations Educational Scientific and Cultural Organization's international panel for harmful algal blooms has proposed a coordinated global ciguatera strategy to address the impact of the disease through *Gambierdiscus* spp. detection and monitoring, improved toxin detection in seafood and epidemiological data collation [105]. The suite of toxins, if any, produced by *Gambierdiscus* varies between species. Bioassays are commonly used to assess toxicity, however liquid chromatography-mass spectrometry (LC-MS/MS) analysis is required to unambiguously identify the toxins present [49]. Some species exclusively produce MTX (*G. australes* [217]) while others solely produce CTX (*G. polynesiensis* [217]) or neither (*G. carpenteri* [121]), as confirmed by LC-MS/MS.

Some *Gambierdiscus* spp. have been classified as endemic to either the Pacific or Atlantic Oceans, while others have been found globally distributed [15, 158]. However, the current understanding about *Gambierdiscus* distribution and abundance is fragmentary due to the paucity of studies worldwide, the continued discovery of new species, difficulty with species identification with light microscopy and the comparatively lower sampling frequency in the Atlantic Ocean [15, 194].

Different species of *Gambierdiscus* usually co-occur at sample sites [158]. As species can show intra-specific morphological variability [22], but can also be highly morphologically similar to one another [121], molecular genetics techniques are necessary to complement the analysis of morphology for species determination. The taxonomy of the genus was poorly defined until relatively recently (ie [157, 221]) which means that species from studies pre-dating 2009 are difficult to evaluate based on current species delimitations [15].

Cases of CFP have been reported in Australia, mostly around the Great Barrier Reef (GBR), but a consistent CTX-producing species of *Gambierdiscus* has not yet been confirmed [144]. This is likely due to a combination of factors: Australia has a coastline of 66,000 km (Figure 2.1), of which at least 50% could be considered tropical or sub-tropical, most of which has never been sampled for *Gambierdiscus* species presence; species identifications have often been uncertain; culturing *Gambierdiscus* requires considerable time and effort; and measuring CTX analogs is also very difficult. To date, the four species *G. belizeanus* M.A.Faust, *G. carpenteri* Kibler, Litaker, Faust, Holland, Vandersea & Tester, *G. toxicus* and *F. yasumotoi* (M.J.Holmes) F.Gómez, D.X.Qiu, R.M.Lopes & Senjie Lin have been identified in Australia (Table 2.1), in Queensland and New South Wales. *Gambierdiscus* was also detected as part of a benthic dinoflagellate community via pyrosequencing in Western Australia, however could only be identified to genus level [122]. Three analogues of MTX, designated MTX-1, MTX-2 and MTX-3, isolated from a *Gambierdiscus* sp. in Queensland, Australia [92]. As this discovery pre-dates the understanding that *Gambierdiscus* includes more species than *G. toxicus*, it remains unclear from which species these analogues were isolated. It is highly likely that many more *Gambierdiscus* species were present even at those particular sites examined, as in each case the studies represent only single or short term sampling

events at one location within a site. Therefore the toxicity and distribution of *Gambierdiscus* species in Australian coastal waters remains largely unknown.

The aim of the current study was to further the continued search for the causative agent/s of CFP in Australia by characterizing a species of *Gambierdiscus* from a ciguateric region using molecular and morphological methods, and investigate the toxicology of the isolates.

2.4 Materials and methods

2.4.1 Specimen collection and culture conditions

Specimens were collected from Heron Island (23° 4420' S, 151° 9140' E), which is part of the southern GBR region. Heron Island constitutes site 3 of previous specimen collection in Australia (Figure 2.1). In October 2013, a sample of *Halimeda* sp. was taken when the average sea surface temperature was 23.8 °C. A second sample of *Padina* sp. was collected in July 2014 at a time when the average sea surface temperature was 19.8 °C. Macroalgal samples were taken around the Heron Island research station (Figure 2.1). The macroalgal samples were stored in 500ml bottles with seawater and processed within 48 hours. The samples were shaken to dislodge protists, then concentrated through a 125 μm sieve [158]. Single cells were isolated from the samples using a micropipette under an inverted light microscope (Nikon Eclipse Ni), washed three times in sterile seawater, and transferred into F/10 medium [94] to establish clonal cultures. The culture HG1 was isolated from the sample from October 2013, while the cultures HG4, HG5, HG6, HG7, HG26 were isolated from the sample from July 2014. The cultures were maintained F/10 medium at 27 °C, $60\mu\text{mol.m}^{-2}.\text{s}^{-1}$ light in 12hr:12hr light to dark cycles.

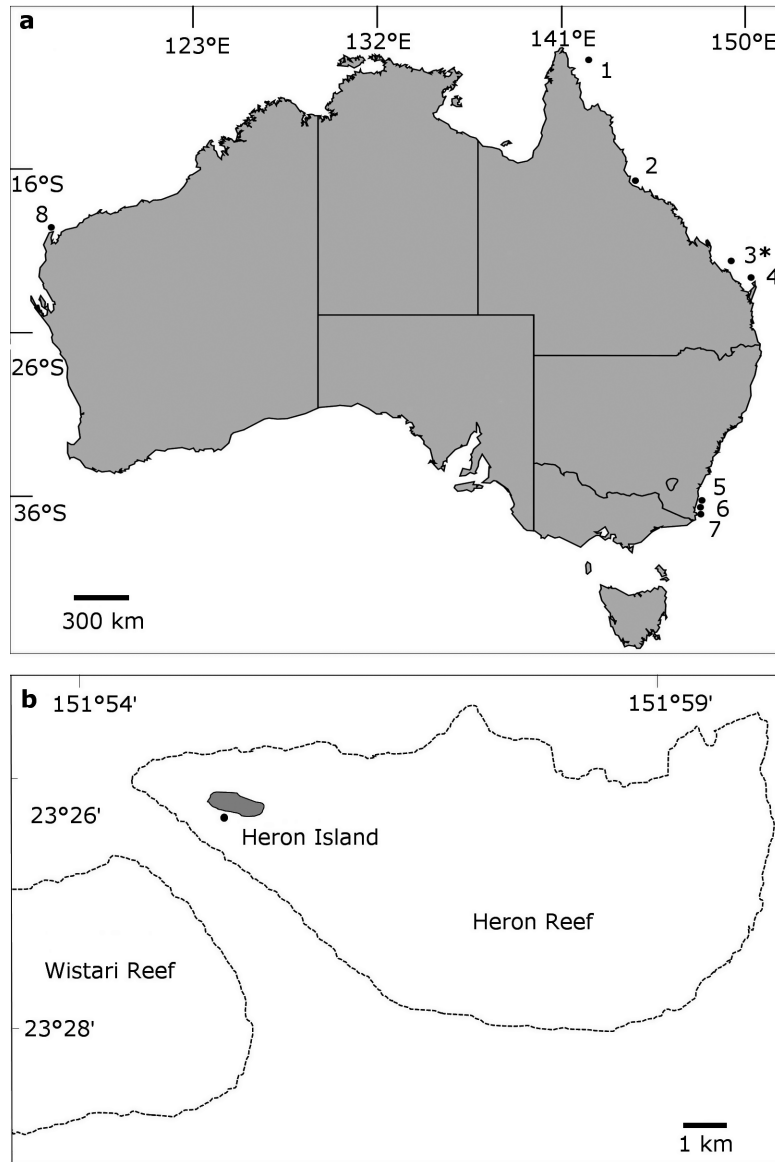


FIGURE 2.1 (a) Sites where *Gambierdiscus* isolates have been reported in Australia: Site 1 indicates Raine Island, QLD; site 2 indicates Townsville, QLD; Site 3 indicates Heron Island (asterix denotes site of panel B), QLD; Site 4 indicated Platypus Bay, QLD; site 5 indicates Wapengo Lagoon, NSW; site 6 indicates Merimbula, NSW; site 7 indicates Eden, NSW; and site 8 indicates Exmouth, WA; (b) Heron Island, Great Barrier Reef (site 3): Sampling site for this study as indicated by black dot.

TABLE 2.1 *Gambierdiscus* spp. identified around Australia, including toxicological data available. N/D denotes not detected. NA denotes not attempted.

<i>Gambierdiscus</i> sp.	Site identified from Figure 2.1	Toxicity	LC-MS profile
<i>G. belizeanus</i>	Site 3: Heron Island, GBR [185]	CTX +ve via RBA [29]; MTX +ve via HELA [90]	NA
<i>G. carpenteri</i>	Site 5: Wapengo Lagoon, NSW [121]; site 6: Merimbula, NSW [121]	MTX +ve via MBA [121]	N/D [121]
<i>G. toxicus</i>	Site 2: Townsville, GBR [80]; Site 7: Eden, NSW [80]	CTX +ve via RBA [29]; MTX +ve via MBA [31]	NA
<i>Gambierdiscus</i> sp.	Site 8: Exmouth, WA [122]	NA	NA
<i>F. yasumotoi</i>	Site 1: Raine Island, GBR [185]; Site 2: Townsville, GBR [185]; site 3: Heron Island, GBR [185]	MTX +ve via MBA [91]	N/D [216]

2.4.2 Morphological analyses

The cultures HG4, HG6, HG7 and HG26 (during exponential growth phase) were fixed with 10 % Lugol's iodide and examined with an inverted light microscope (Nikon Eclipse TS100 equipped with an Infinite Luminera 1 camera) for cell size. Depth (D: measured along dorso-ventral axis) and width (W: measured between the lateral axis) was measured for 30 to 36 cells per strain. The culture HG26 was fixed with 10 % Lugol's iodide for plastid and nucleus diagnosis with an inverted light microscope (Nikon Eclipse Ni equipped with an Nikon DS-Qi2 camera). For plate pattern identification LM with calcofluor staining of the type strain HG7 was carried out under UV (Olympus BX51 microscope equipped with Olympus DP73 camera, Australia). The culture was centrifuged at 300g, concentrated and fixed in 4% Lugol's iodide, then stained with 1 μ L of calcofluor stock solution (1 mg.mL⁻¹) of Fluorescent Brightener 28 (Sigma, St Louis, MO, USA) for illumination. For HG4, HG5, HG6, HG7 and HG26 30 ml of culture was centrifuged at 15,000 x g and concentrated to 2 ml, preserved with 10 % Lugol's iodide, and analyzed by scanning electron microscopy (SEM) at the Senckenberg am Meer (Wilhelmshaven, Germany). Cells were placed on a 5 μ m Millipore filter, rinsed in distilled water, and dehydrated in a series of increasing ethanol concentrations (30, 50, 70, 85, 90, 100 %), followed by chemical drying with hexamethyldisilazane at room temperature. When completely dry, the sample was mounted on a stub and sputter coated with gold palladium (SCD 050 Bal-Tec, BAL-TEC Präparations-gerätevertrieb, Wallof, Germany). Cells were observed using a Tescan VEGA3 microscope (Elektronen-Optik-Service GmbH, Dort-

mund, Germany) at 15 kV.

2.4.3 Genomic DNA extraction

Genomic DNA was extracted using the CTAB method [297]. Purity and concentration of the extract was measured by Nanodrop (Nanodrop2000, Thermo Scientific), the integrity was visualised on 1% agarose gel.

2.4.4 PCR amplification and sequencing

Extracted DNA was used as a template to amplify rDNA sequences in 25 μ l mixture in PCR tubes. Reactions consisted of a final concentration of 0.6 μ M forward and reverse primer, 0.4 μ M BSA, 2 - 20 ng DNA, 12.5 μ l 2xEconoTaq (Lucigen Corporation, Middleton, WI, USA) and 7.5 μ l PCR grade water.

The PCR cycling comprised of an initial 10 min step at 94 °C, followed by 30 cycles of denaturing at 94 °C for 30 s, annealing at 55 °C for 30 s and extension at 72 °C for 1 min, finalized with 3 minutes of extension at 72 °C. The LSU D8-D10 and SSU rDNA regions were amplified with the FD8-RB and 18ScomF1-18ScomR1 primer sets, respectively ([Table 2.2](#)).

Sanger sequencing was conducted by MacroGen Inc. (Seoul, Korea).

TABLE 2.2 List of primers used for phylogenetic analysis of *Gambierdiscus* strains, synthesised by Integrated DNA technologies.

Name	Sequence (5'-3')	Purpose	Reference
LSU D8-D10 region			
FD8	GGATTGGCTCTGAGGGTTGGG	Amplification & sequencing	[31]
GLD8_-421F	ACAGCCAAGGGAACGGGCTT	Sequencing	[193]
GLD8_-677R	TGTGCCGCCCCAGCCAACT	Sequencing	[193]
RB	GATAGGAAGAGCCGACATCGA	Amplification & sequencing	[31]
SSU region			
18ScomF1	GCTTGTCTCAAAGATTAAGCCATGC	Amplification & sequencing	[290]
18ScomR1	CACCTACGGAAACCTTGTTACGAC	Amplification & sequencing	[290]
Dino18SF2	ATTAATAGGGATAGTTGGGGGC	Sequencing	[292]
Dino18SR1	GAGCCAGATRCDCACCCA	Sequencing	[292]
G10'F	TGGAGGGCAAGTCTGGTG	Sequencing	[193]
G18'R	GCATCACAGACCTGTTATTG	Sequencing	[159]

2.4.5 Sequence alignment and phylogenetic analysis

Sequencing data were aligned with *Gambierdiscus* spp. data from the GenBank reference database (accession numbers as part of Figure 2.7 and Figure 2.8). The alignment algorithm MUSCLE, with a maximum of eight iterations [53], was used through the Geneious software, version 8.1.7 [113]. Alignments were truncated to the same length (D10-D8 and SSU at 799bp and 1690bp, respectively), and discrepancies removed. Alignments are available on request. Phylogenetic trees were calculated with Bayesian inference (BI) as well as maximum likelihood (ML). BI, via Mr. Bayes 3.2.2, was used to estimate the posterior probability (PP) distribution with Metropolis-Coupled Markov Chain Monte Carlo simulations [229]. A random starting tree with three heated and one cold chain(s) with a temperature set at 0.2. Trees were sampled every 100th generation for the 2,000,000 generations generated. PHYML was used for ML analysis with 1,000 bootstraps (BS) [76]. The general time reversal model with an estimated gamma distribution was used for both analyses.

2.4.6 Toxin production

2.4.6.1 Toxicity via mouse bioassay

The cultures HG4, HG6 and HG7 (1,248,000, 3,982,000, and 1,323,200 cells, respectively) were freeze-dried in late stationary phase (day 36 of culture). The weights of the samples were 57.0 mg, 155.0 mg and 23.0 mg, respectively. The samples were extracted exhaustively with methanol. The extracts were evaporated to a small volume at 35°C using a rotary evaporator, and the solution aliquotted into glass vials. The remaining methanol was removed under a stream of nitrogen, and the extracts freeze-dried overnight. The total weight of the extracts, which were all brown green colour, was as follows: HG4, 22.5 mg HG6, 42.9 mg HG7, 15.1 mg. The median lethal doses of the test materials by intraperitoneal (i.p.) injection were determined according to the principles of the Organisation for Economic Co-operation and Development Guideline 425. Samples were dissolved in 1% Tween 60 in normal saline immediately before dosing. For toxicity by i.p. injection, aliquots of this solution, made up to a total volume of 1 ml with the same solvent, were injected into female Swiss mice, of initial body weight 18-22 g. The toxicity of the extract of HG6 by gavage was also investigated, when aliquots of the extract solution were made up to 200 μ l with Tween-saline. The mice were monitored intensively during the day of dosing. Survivors were examined and weighed each day for the following 13 d, after which they were killed and necropsied. Tap water and food (Rat and Mouse Cubes, Speciality Feeds Ltd, Glen Forrest, Western Australia) were available at all times.

2.4.6.2 Toxin profile screening by liquid chromatography-mass spectrometry

After 36 days in culture 2 litres of HG4, HG6 and HG7 were centrifuged (15,000g) and the resultant pellet freeze-dried. Pellets were extracted with methanol and screened for the presence of maitotoxin (MTX-1) and algal CTXs (CTX-3B, CTX-3C, CTX-4A, CTX-4B) using a LC-MS/MS methods developed by the Cawthron Institute [123]. The limit of detection of the analysis was 1 ng.mL⁻¹ for MTX-1 and <1 ng.mL⁻¹ for the various CTX analogs monitored.

2.5 Results

Gambierdiscus lapillus sp. nov. Kretzschmar, Hoppenrath et Murray

2.5.1 Morphology and taxonomical treatment

Description: Cells were anterioposteriorly compressed and round to ellipsoid, sometimes dorsally pointed in apical view (Figure 2.2, Figure 2.4). Cells were small compared to other *Gambierdiscus* spp., with a dorsoventral depth of 40.6 μm (range of 34.4 - 50.9 μm ; SD 3.3 μm) and width of 39 μm (range of 32 - 46.9 μm ; SD 3.2 μm) with a depth to width ratio of 1.04 (SD 0.06 μm) (n=128, Table 2.3).

TABLE 2.3 Morphological measurements of *G. lapillus* sp. nov. strains collected from Heron Island, Australia. Between 30 to 36 cells were counted per strain. Values in parentheses are $\pm$ SD

Strain	D (μm)	W (μm)	D:W ratio
HG4	39 ($\pm$ 2.6)	38.1 ($\pm$ 3.1)	1.03 ($\pm$ 0.06)
HG6	43.2 ($\pm$ 3.0)	40.9 ($\pm$ 3.0)	1.06 ($\pm$ 0.04)
HG7	39.2 ($\pm$ 2.8)	38.4 ($\pm$ 2.8)	1.02 ($\pm$ 0.07)
HG26	40.7 ($\pm$ 2.7)	38.5 ($\pm$ 3.2)	1.06 ($\pm$ 0.05)

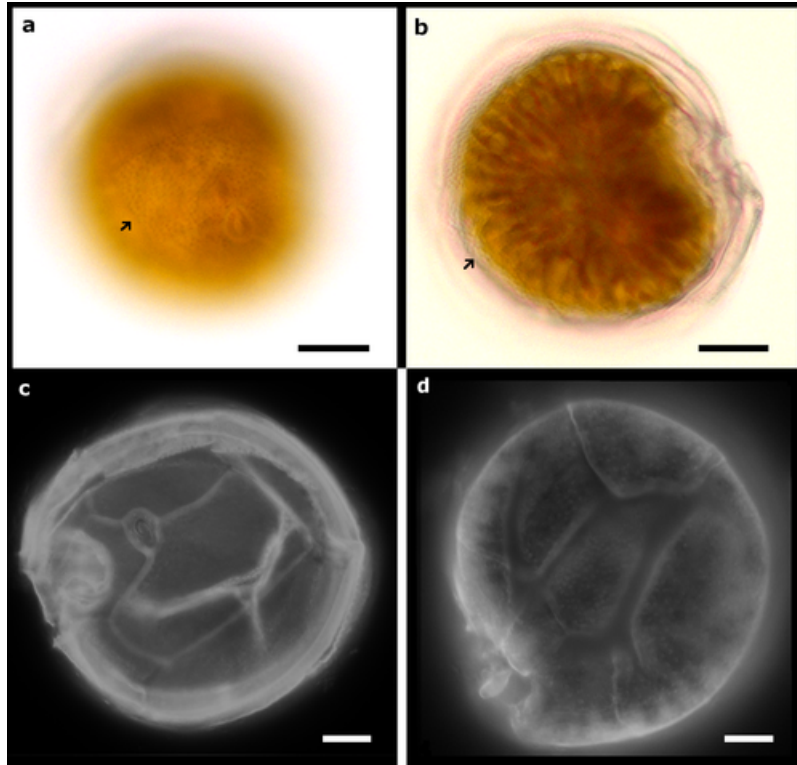


FIGURE 2.2 Light micrographs of *Gambierdiscus lapillus* sp. nov. (a, b) Bright field, arrow in (a) shows the 2' apical plate with characteristic aerolated surface and pores; (b) shows the distinctive ribbed appearance due to areolation and pores; (c, d) epifluorescence images (Calcofluor staining) show plate tabulation, (c) epitheca with 2; plate and (d) hypotheca with the distinctive heavily reticulate-foveate surface. Scale bar equal 10 μm .

The hypotheca was slightly higher than the epitheca, rarely significantly higher (Figure 2.3f and suppl. Figure 2.10a). The cell surface had a strong reticulate-foveate ornamentation (Figure 2.2, Figure 2.4, Figure 2.3), apical and antapical plates with more regularly arranged deep depressions nearly all containing a pore, and pre- and postcingular plates more irregularly reticulated with scattered pores (Figure 2.5).

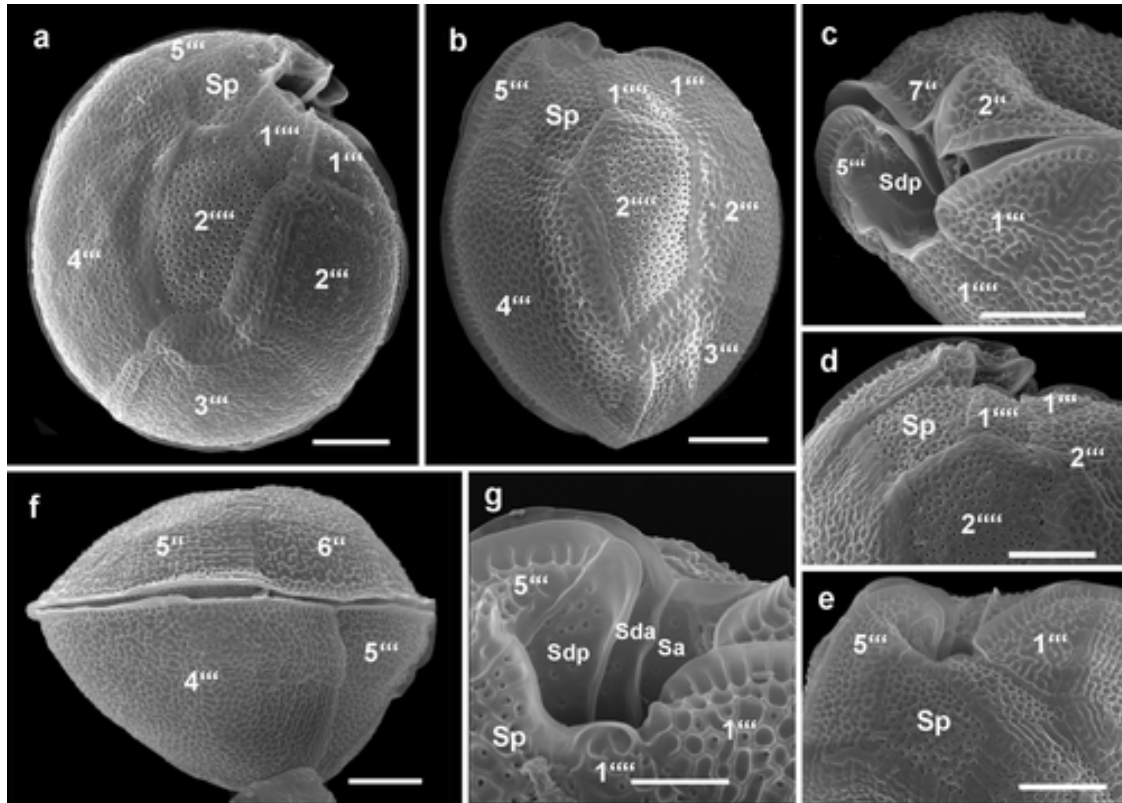


FIGURE 2.3 Scanning electron micrographs of *Gambierdiscus lapillus*. (a, b) Hypotheca in antapical view. (c) Left lateral to ventral view of the sulcal opening. (d, e) Antapical view of the ventral area, note the posterior sulcal plate appearing like a hypothecal plate. (f) Right lateral view of a cell with relatively high hypotheca. (g) Antapical view into the sulcus. Scale bars (a-f) 10 μm , (g) 5 μm .

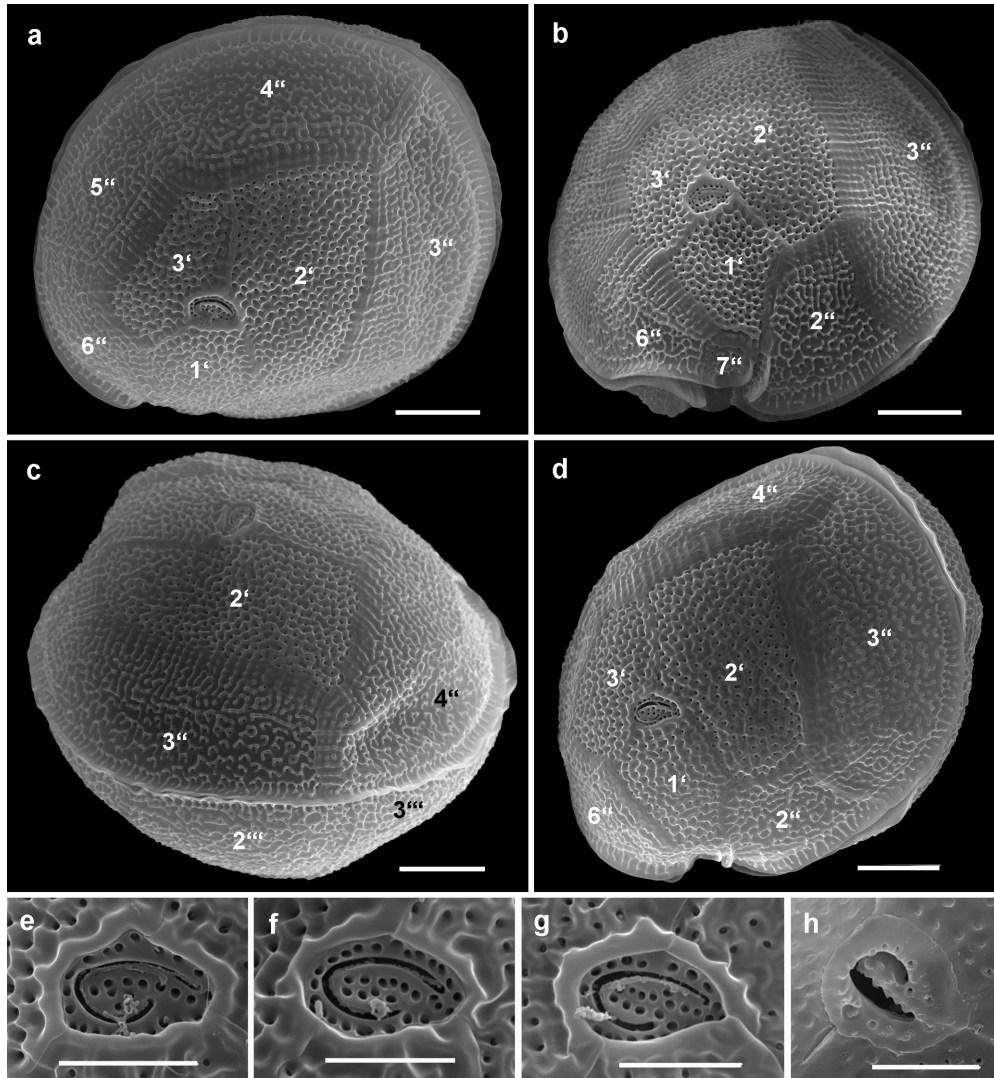


FIGURE 2.4 Scanning electron micrographs of the epitheca of *Gambierdiscus lapillus*. (a) Apical view. (b) Apical ventral view. (c) Left lateral view. (d) Apical to left lateral view. (e–h) Apical pore complex. (e–g) Outside view of the pore plate (Po), note the variable arrangement of large depressions. (h) Inside view of Po. Scale bars (a–d) 10 μm , (e–h) 5 μm .

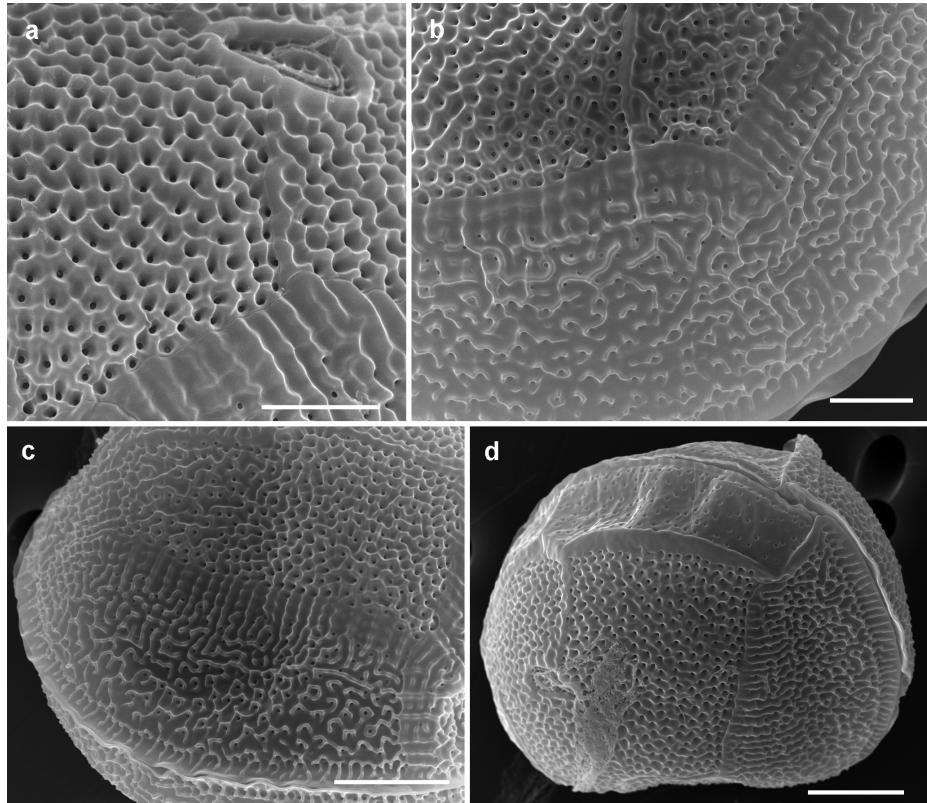


FIGURE 2.5 Scanning electron micrographs of *Gambierdiscus lapillus* showing the reticulate-foveate thecal ornamentation. (a) Apical plates with regularly arranged deep depressions, nearly all containing a pore. (b, c) Wide transversely “ribbed” intercalary bands separating apical plates from precingular plates with more irregular reticulation with scattered pores. (d) Cell after division with newly developing plates that have a smooth surface with scattered pores. Scale bars (a, b) 5 μm , (c, d) 10 μm .

Wide intercalary bands had a ribbed appearance (Figure 2.5). Newly developing plates after cell division had a smooth surface with scattered pores (Figure 2.5d). The plate formula (Kofoid tabulation) was Po, 3', 0a, 7'', 6c, 7-8s, 5''', 0p, 2'''' (Figure 2.4, Figure 2.3, suppl. Figure 2.10, suppl. Figure 2.11). The apical pore plate (Po) is oval with a fish-hook shaped, slit-like apical pore, located close to the center of the epitheca, slightly shifted ventrally (Figure 2.4). Po was about 6.1-5.2 x 3.6-3.7 μm in size (Figure 2.4e-h). twenty-eight to thirty-four large depressions were distributed over the Po plate and arranged in no constant pattern (Figure 2.4e-g); one (Figure 2.4e) or two (Figure 2.4f, g) rows of depressions intruded into the hook. As seen by an inside view the depressions contained a pore (Figure 2.4h). Also a tiny pore was visible on the Po-plate (Figure 2.4c). The largest apical plate was the hatchet-shaped 2' plate, followed by the pentagonal 3' plate and the hexagonal 1' plate (Figure 2.2c, Figure 2.4a-d, Figure 2.6a).

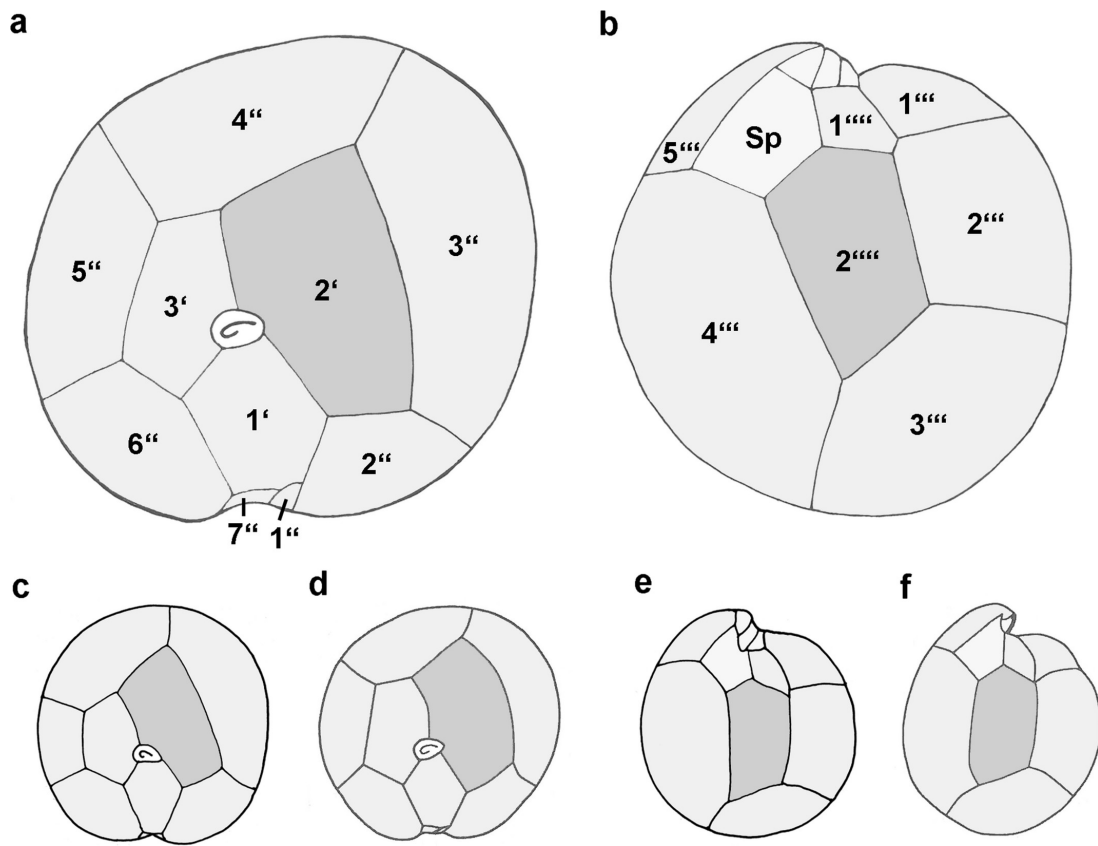


FIGURE 2.6 Line drawings showing the thecal tabulation pattern of *Gambierdiscus lapillus* (a, b); *Gambierdiscus belizeanus* (c, e); and *Gambierdiscus scabrosus* (d, f). Epitheca (a, c, d). Hypotheca (b, e, f)

The largest precingular plate was plate 3'' (four-sided, covering the left lateral precingular area) and the pentagonal 4'' plate (dorsal) was nearly of the same size (Figure 2.4a-d, Figure 2.6a). Plates 5'' and 6'' were on the right lateral side (Figure 2.4a, Figure 2.6a). Plates 1'' and 7'' were the smallest (Figure 2.4b, Figure 2.3c, Figure 2.6a) located ventrally. The largest postcingular plate was the 4''' plate (covering most of the left hypotheca), followed by the plates 2'', 3'', 5'' and 1'' (Figure 2.3a-b, Figure 2.6b). The quadrangular first antapical plate (1''') was small and the larger, relatively narrow pentagonal 2''' plate that covered the antapex (Figure 2.3a-b, Figure 2.6b). The sulcus was a deep excavation (Fig. Figure 2.3c, g) and composed of seven to eight plates (Suppl. Figure 2.11). The narrow and deep cingulum was difficult to observe and consisted of six plates (Suppl. Figure 2.11c). A few cells displayed shape variability of plate 2' (Suppl. Figure 2.12c-e). The shape varied from hatchet shaped (main morphology) to nearly rectangular (Suppl. Figure 2.12c) and six-sided (Suppl. Figure 2.12e). Some of the cells were tear-drop shaped rather than ellipsoid (Suppl. Figure 2.12a-c).

Etymology: The epithet refers to the diminutive size and heavily ornamented thecal plates of the species. Latin, lapillus (= little gem).

Holotype: Figure 2.4a (strain HG7); SEM-stub (designation CEDiT2016H58) deposited at Senckenberg am Meer, German Centre for Marine Biodiversity Research, Centre of Excellence for Dinophyte Taxonomy, Germany.

Isotype: Lugol-fixed subsample of strain HG7 (designation CEDiT2016I59) deposited at the Senckenberg am Meer, German Centre for Marine Biodiversity Research, Centre of Excellence for Dinophyte Taxonomy, Germany.

Molecular diagnosis: strain HG7 SSU (KU558932) and LSU (KU558928)

Type locality: Heron Island (23.4420° S, 151.9140° E), GBR, Australia, South Pacific Ocean.

Molecular characterization: *Gambierdiscus lapillus* can be genetically identified by rDNA sequences deposited in GenBank SSU: HG1 KU558929, HG4 KU558930, HG6 KU558931, HG26 KU558933; and D8-D10 LSU: HG1 KU558925, HG4 KU558926, HG6 KU558927.

Further information: The strain HG7, from which the holotype is derived, has been deposited in the Australian National Algae Culture Collection under strain designation CS-1146.

2.5.2 Phylogenetics

Phylogenetic trees for SSU and D8-D10 LSU rDNA regions were inferred using both BI and ML methods. Strains were marked according to their geographic origin, as per figure description (Figure 2.7, Figure 2.8).

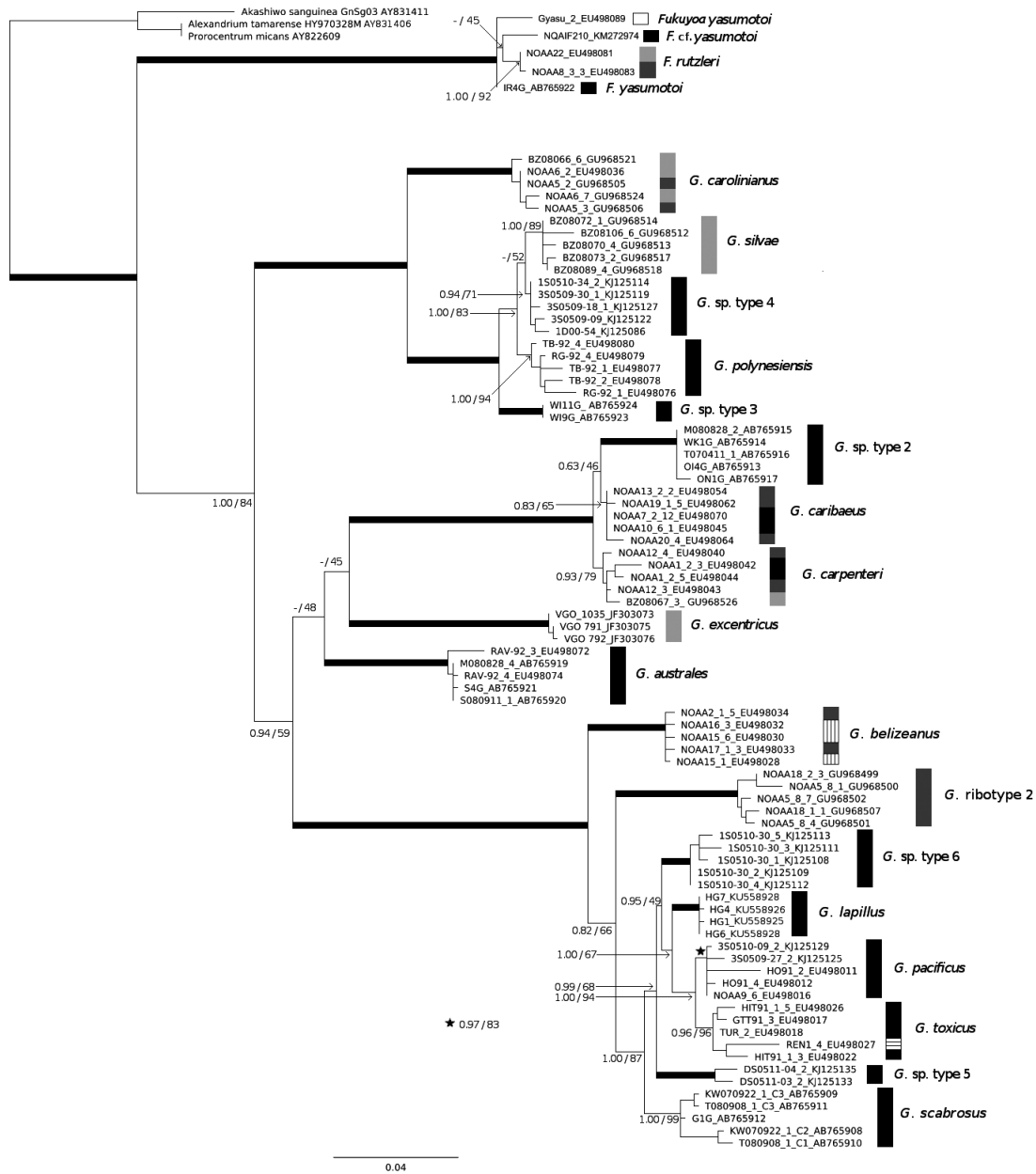


FIGURE 2.7 Maximum likelihood phylogeny of *Gambierdiscus* species/phylotypes of the LSU D8-D10 rDNA region. Nodal support are Bayesian posterior probability (PP) and bootstrap (BS) values obtained from Bayesian inference analysis and maximum likelihood analysis, respectively. Nodes with strong support (PP/BS = 1.00 / 100) are shown as thick lines. The geographic origin of the isolates is coded as follows: Black = Pacific Ocean; dark grey = Caribbean Ocean; light grey = Atlantic Ocean; white = South China Sea; horizontal white stripes = Indian Ocean; vertical white stripes = Gulf of Mexico.

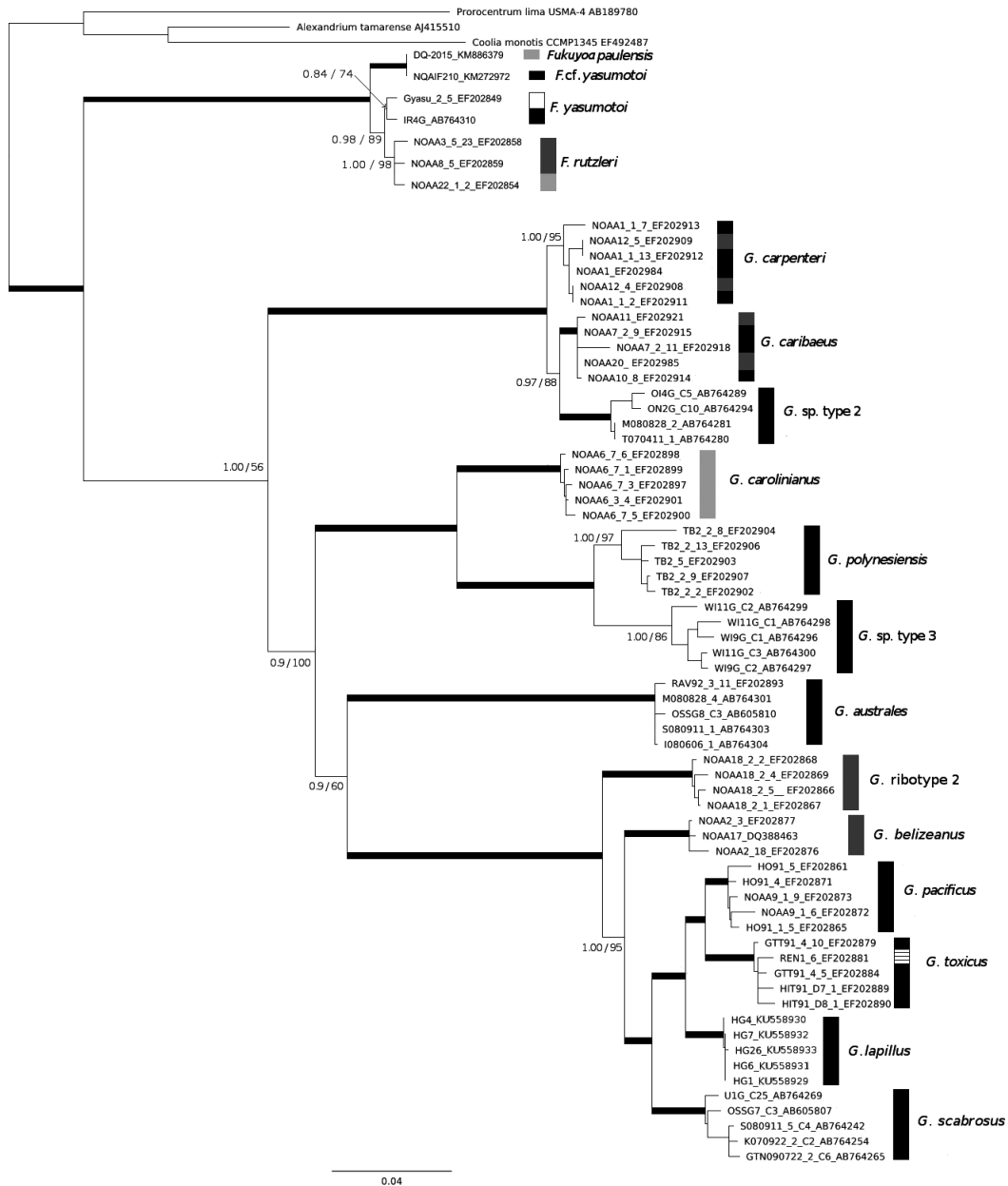


FIGURE 2.8 Maximum likelihood phylogeny of *Gambierdiscus* species/phylotypes of the SSU rDNA region. Nodal support are Bayesian posterior probability (PP) and bootstrap (BS) values obtained from Bayesian inference analysis and maximum likelihood analysis, respectively. Nodes with strong support (PP/BS = 1.00 / 100) are shown as thick lines. The geographic origin of the isolates is coded as follows: Black = Pacific Ocean; dark grey = Caribbean Ocean; light grey = Atlantic Ocean; white = South China Sea; white with black horizontal stripes = Indian Ocean.

Branches were analyzed thus: 100%/1.0 was fully supported, above 90%/0.9 was very well supported, above 80%/0.8 was relatively well supported and below 50%/0.5 was unsupported for ML and BI, respectively. The SSU rDNA topologies using BI and ML were the same with all major branches fully supported (1.00 / 100, PP and BS, respectively). Within *Gambierdiscus*, the majority of the branches were fully supported, with all except one at least relatively well supported. The branch splitting *G. australes* M.Chinian & M.A.Faust from the major clade including *G. pacificus* M.Chinian & M.Faust, *G. toxicus* and *G. lapillus* had the lowest support at 0.95 / 60. For the SSU rDNA region, the major clades of *Gambierdiscus* were fully supported based on ML BS and BI PP. The major clade support for the D8D10 LSU rDNA varied from fully supported to unsupported depending on the clade and support between ML BS and BI PP varied (eg. the branching for the major clade containing *G. caribaeus* Vandersea, Litaker, Faust, Kibler, Holland & Tester and *G. carpenteri* Kibler, Litaker, Faust, Holland, Vandersea & Tester). The same relationships of the Heron Island isolates HG1, HG4, HG6, HG7 and HG26 to the previously described *Gambierdiscus* spp. as well as uncharacterised ribotypes and species types were found in the phylogenies based on both D8D10 LSU rDNA (Figure 2.7) and SSU rDNA (Figure 2.8). The phylogenies showed that these strains are clearly separated from other *Gambierdiscus* spp. and fall into a new, fully supported clade. The new clade falls within the most diverse major clade (8 species, Figure 2.7). *Gambierdiscus lapillus* is closely related to *G. sp. type 6*, *G. toxicus* and *G. pacificus*. The closest following relatives from within their major clade are *G. sp. type 5*, *G. scabrosus*, *G. ribotype 2* and then *G. belizeanus* in descending order. No differentiation within species based on their geographic origin was determined for the SSU rDNA or LSU D8-D10 rDNA regions (Figure 2.7 and Figure 2.8).

2.5.3 Toxicology

2.5.3.1 Toxicity via intraperitoneal injection

The i.p. LD50 of the extract of HG6 was 0.78 mg.kg⁻¹, or 1,448 cells.per MU, with 95% confidence limits between 0.40 and 1.60 mg.kg⁻¹. At high dose-levels, stretching movements were observed immediately after injection of the test material, most likely due to irritation. This was rapidly followed by cessation of movement and abdominal breathing. Respiration rates declined and death occurred within 1.5-4 hours after dosing. At necropsy, erythema was observed throughout the gastrointestinal tract of these animals. At lower doses, death occurred at up to 24 hours after dosing. The stomachs of these animals were grossly distended with gas and with abnormally fluid contents. An abnormally large amount of material was present in the duodenum and upper jejunum, which was reddish-brown in colour and gelatinous in consistency. Caeca were also enlarged. Erythema in the glandular stomach was noted. No gross changes were observed in any other organ. The acute i.p. toxicity of the extract of HG7 was lower than that of HG6, with an LD50 of 12.5 mg.kg⁻¹, or 219,073 cells.per MU, with 95% confidence interval between 10.1 and 15.3 mg.kg⁻¹.

The symptoms of intoxication and the gross pathology in mice injected with HG7 were the same as those recorded with HG6. The acute i.p. toxicity of the extract of HG4 was even lower. No deaths were recorded at 100 mg.kg⁻¹. A dose of 150 mg/kg was lethal, however, with the same effects as seen with the other two extracts. Insufficient material was available to determine a precise LD50, though from the available data it may be concluded that it lies between 100 and 150 mg.kg⁻¹, or between 1,109,333 cells.per MU and 1,664,000 cells.per MU.

2.5.3.2 Toxicity via gavage

HG6 was much less toxic by oral administration than by i.p. injection. No effects were observed at a dose of 300 mg.kg⁻¹, which is 385 times the median lethal dose by i.p. injection. Insufficient material was available for dosing at higher levels.

2.5.3.3 Toxin profile

Analysis of extracts from *G. lapillus* isolates showed no detectable MTX-1 or known algal CTXs, CTX3B, 3C, 4A, 4B. However, all extracts were positive for a compound tentatively described as MTX-3, which represent a sulfated polyether produced by many *Gambierdiscus* species. Also, unassigned chromatographic peaks were observed in the CTX transition and these warrant further investigation.

2.6 Discussion

Gambierdiscus lapillus cells are antero-posteriorly compressed, poses a narrow 2''' plate (1p plate equivalent) and heavily areolated cell surface (strong reticulate-foveate thecal ornamentation), as does *G. belizeanus* [157]. *Gambierdiscus scabrosus* T.Nishimura, Shinya Sato & M.Adachi is also morphologically similar to *G. belizeanus*, but can be distinguished by the asymmetry of the 4'' plate (designated 3'' by Nishimura et al. (2014)) compared to the symmetry of the 4'' plate in *G. belizeanus* (Table 2.4, asymmetric in Fig. 42A in [99]).

TABLE 2.4 Morphological character of closely related *Gambierdiscus* species and *G. lapillus* sp. nov. Common features preceding diagnostic features are Po, 3', 0a, 7'', 6c, 7-8s, 5''', 0p, 2''' and cells antero-posteriorly compressed.

Species	<i>G. lapillus</i>	<i>G. belizeanus</i>	<i>G. scabrosus</i>	<i>G. toxicus</i>	<i>G. pacificus</i>
Cell depth (μm)	40.6 $\pm$ 3.3	63 $\pm$ 2.2	63.2 $\pm$ 5.7	93 $\pm$ 5.7	70 $\pm$ 4.7
Cell width (μm)	39 $\pm$ 3.2	58 $\pm$ 2.5	58.2 $\pm$ 5.7	83 $\pm$ 2.3	63 $\pm$ 3.6
Theca ornamentation	Reticulate-foveate	Reticulate-foveate	Reticulate-foveate	Smooth	Smooth
Plate 2'	Hatchet-shaped	Nearly rectangular	Rectangular	Hatchet-shaped	Hatchet-shaped
Plate 4''	Symmetric	Symmetric (Fig. 3 [57]); Asymmetric (Fig. 42A [99])	Asymmetric	Asymmetric	Symmetric
Plate 2'''	Narrow	Narrow	Narrow	Broad	Narrow
References	This study	[31, 32, 57, 157]	[130, 193, 194]	[2, 33, 157, 221]	[31, 32, 157]

Hence, the closest morphological relatives to *G. lapillus* are *G. belizeanus* and *G. scabrosus*. Comparing the distinguishing morphological characteristics between *G. lapillus* sister species (Table 2.4), is the *G. lapillus*'s diminutive size (Figure 2.6, Figure 2.9).

Size is a feature that has been previously used to distinguish species of *Gambierdiscus* from each other [157]. *Gambierdiscus lapillus* also differs to *G. scabrosus* due to its symmetric 4'' plate (Figure 2.4a), hence it is closer morphologically to *G. belizeanus*. The morphological point of difference between the two species lies in the 2' plate (Figure 2.6). The 2' plate of *G. lapillus* is hatchet shaped (Figure 2.4a-d) whereas the 2' plate of *G. belizeanus* is rectangular (Fig. 3 in [57]). In terms of the genetic relatedness of the new species and *G. belizeanus*, they are relatively distantly related (Figure 2.7 and Figure 2.8).

A small subset of isolates displayed some variability morphological features, and were different to the majority of cells, in the 2' plate as well as cell shape (Suppl. Figure 2.12). Bravo et al. (2014) recorded a change in cell shape and size during the life cycle of a clonal strain of *Gambierdiscus* sp., where a small proportion of the cells observed displayed this variability. They postulate that this was previously unrecorded due to their large number of observations compared to the rest of the literature [22]. We suggest that the variability we observed is similar due to screening numerous cells of four strains of *G. lapillus*. The variability in 2' plate is negligible; the hatchet shape of this plate is still a diagnostic tool for identifying *G. lapillus*.

Kohli et al. (2014a) observed a similar difference from the original species description in the mor-

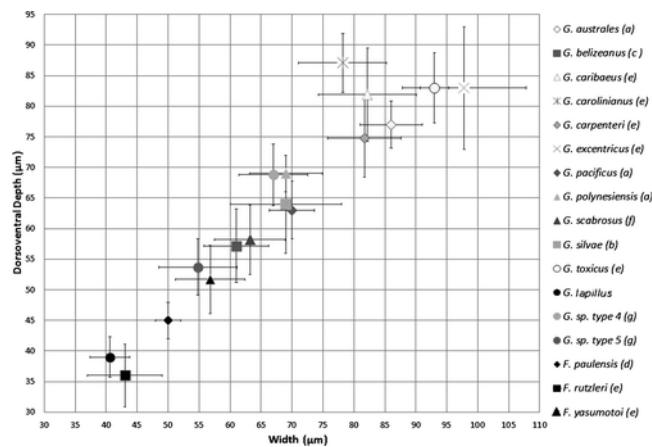


FIGURE 2.9 Visual representation of the size difference between *Gambierdiscus* and *Fukuyoa* species. Error bars denote standard deviation of measurements. Data taken from publications as follows: a) Chinain et al. (1999b) [31]; b) Fraga and Rodriguez (2014) [61]; c) Faust et al. (1995) [57]; d) Gómez et al. (2015) [73]; e) Litaker et al. (2009) [157]; f) Nishimura et al. (2014) [194]; and g) Xu et al. (2014) [282]

phology of strains of *G. carpenteri* [121]. Strains of *G. carpenteri* isolated from southern NSW, Australia, were identified by a combination of morphological features and genetic sequencing. Genetically these strains were 99% identical to the type strains of *G. carpenteri*. However, their morphological features differed somewhat from those originally described for *G. carpenteri*, and were more similar to those of the very similar species *G. toxicus* (broad 2nd plates, pointed dorsal ends, and the lack of a dorsal rostrum) [121, 157]. The potential intraspecific morphological variability of *Gambierdiscus* spp. indicates that the identification of *Gambierdiscus* spp. needs to be approached with caution, and preferably verified with molecular analysis.

The sister groups to *G. lapillus* in the phylogenetic analyses were found to be *G. pacificus* and *G. toxicus*, which was fully supported (Figure 2.7 and Figure 2.8). For the D8-D10 LSU region, the confidence of the interspecies relationship within the major clade containing *G. lapillus* was lower. While the resolution of *G. lapillus* as a separate species is fully supported, the other branches within that major clade were more ambiguous and support differed between BI and ML with the latter consistently lower than the former. Branches were well supported or better in BI phylogeny, while in the ML phylogeny some were unsupported. The topology of the clade resembles that of the phylogeny by Xu et al. (2014), in which the *G. sp. type 5* and *G. sp. type 6* were presented, hence it can be assumed that the BI phylogeny is accurate [282]. The discrepancy in confidence between SSU and D8-D10 LSU could be due to the length of the fragment, as the SSU fragment is almost twice the length as the LSU fragment. Alternatively it could be due to lower selection pressure, in effect higher mutation rate, in the D8-D10 region compared to the SSU region. Murray et al. (2015) found that the D1-D6 LSU region in dinoflagellates had a substitution rate 4-8% times that of the SSU region [183]. Whether this faster substitution rate applies to the D8-D10 region remains unknown, however the possibility should be taken into consideration.

Morphologically, *G. lapillus* is distinct from *G. pacificus* due to its smooth cell surface [31],

whereas *G. lapillus* is discrete from *G. toxicus* based on the broad 2'' plate of *G. toxicus* [157].

Berdalet et al. (2012) postulated that species of *Gambierdiscus* may be cosmopolitan, and the fact that some species were reported from only one or few locations was likely due to under sampling [158]. It has been postulated that *Gambierdiscus* strains of the same species in warmer areas display higher toxicity than those in cooler regions [19]. To observe whether genetic divergence exists between strains from the same species from different localities, where possible a range of isolates from a variety of geographic locations were included in the phylogenetic study. The phylogenetic analysis of the D8-D10 LSU rDNA (Figure 2.7) *G. belizeanus*, *G. caribaeus*, *G. carpenteri*, *G. carolinianus* Litaker, Vandersea, Faust, Kibler, Holland & Tester and *G. toxicus* strains of the same species clustered together, regardless of their geographic origin. Examining the strain distribution in light of their geographic origin for the SSU rDNA fragment (Figure 2.8), a similar trend was found. Strains of *G. carpenteri*, *G. caribaeus* and *G. toxicus* showed closer relationship to geographically removed strains than ones from their own location. This indicates that the rDNA fragments used in this study are likely not variable enough to reveal intraspecific, regional-level divergence. More investigation using tools such as microsatellites or RADseq are likely required to reveal regional differences.

CFP is endemic to the GBR in Australia, yet the particular species of *Gambierdiscus* that are most likely producing CTXs in this region are still not known. The report that *F. paulensis* from the Mediterranean produces possible analogs of CTXs [136] is interesting, as the study classifies the *F. cf. yasumotoi* found in Australia as *F. paulensis* (Table 1 in [185]). However, we suggest caution should be used in interpreting the LC-MS results on *F. paulensis* as presented [136], as: data to support the presence of MTX was not presented; no standard for gambieric acid A was used; and the peak at 6.29 min assigned to 54-deoxy CTX1B (panel b Fig. 9) was not highly similar to the 54-deoxy CTX1B standard peak (Fig. 9c). As 54-deoxy CTX1B is considered to be a fish metabolite [287], it appears novel that it would have been found in a species of *Gambierdiscus*. As the detection of CTX analogs using LC-MS is very difficult, and has not been conducted often for species of *Gambierdiscus*, further work clarifying the toxin profile of more species will shed light on these issues.

The identification of *Gambierdiscus* species producing CTXs in Australia has been hampered by several uncertainties and difficulties, and no identified strains that consistently produce known CTX analogs have yet been confirmed. *G. belizeanus* has been reported from Heron Island, GBR, [185]. Previously, CTX and MTX production in *G. belizeanus* has been indicated via bioassay [29, 90] although LC-MS analysis using known CTX analogs has not been conducted for this species as yet.

A previous study found a possible CTX analog in a *Gambierdiscus* strain from Queensland, however, the production of the compound was unstable, and the strain precedes the discovery of

species diversity in *Gambierdiscus* [94], and the identification of the strain is therefore uncertain. Unfortunately, those cultures are no longer available, and therefore the identity of the species cannot now be verified [94]. Based on the LC-MS/MS toxin profile of *G. lapillus*, it produces a number of unknown compounds, which may include undescribed congeners of CTX. These would require further work in order to characterize them. The toxicity measurements of *G. lapillus* by mouse bioassay were found to be equally unusual. The pathology of the mice was atypical, and not indicative of previously reported MTX or CTX intoxication (R. Munday, pers. comm.). Whether the abnormal pathology is due to a possible unknown CTX congener, or to another toxin produced by this species, remains unknown. Further fractionation of whole cell extracts into aqueous methanol and dichloromethane partitions, which are expected to contain MTX from CTX respectively [237], could be conducted to elucidate which class of toxin is likely causing the pathology. Whole cell extract MBAs in the literature were normalized to cells per MU by Litaker et al. (2010) (table 5 therein). Comparatively, HG4 and HG7 were much less toxic than all the reported samples. HG6 displayed a higher toxicity than the cultures established from bloom samples [32]. The strain's toxicity compared to the other reported strains from both the Atlantic and the Pacific was in the bottom quarter percentile. These studies precede the elucidation of the current *Gambierdiscus* species complex, so it remains unclear which species these results relate to. Markedly, the toxicity was found to differ greatly between the *G. lapillus* isolates, with the toxicity of HG6 being 150-fold higher than HG7, which was in turn more than 5-fold more toxic than HG4. The published intraspecies toxicity variation in *G. polynesiensis* M.Chinain & M.Faust via MBA was only 2-fold [29]. Potential reasons for differing toxin production in cultures may be physiological, such as phosphorous limitation in the media and different stages of the growth phase and growth rate [29, 251]. However, the HG4, HG6, and HG7 were grown under the same conditions and harvested simultaneously, which indicates an intrastrain variation in toxin production, rather than external influence on toxin production regulation.

2.7 Acknowledgements

We want to thank Dr. Nicolas Chomérat for his help finding a name for the species and his advice with Latin. Thank you to Dr. Rex Munday of AgResearch for conducting the toxicology investigation. Thanks to Dr. Gurjeet Kohli and Dr. Penelope Ajani, UTS, for advice and guidance in the lab. Thanks to Dr. Katherina Petrou, Dr. Daniel Nielsen, and Dr. Peter Ralph, UTS, for sample collection. Samples were taken under the permit issued by the Queensland Government, in relation to Marine Parks. The microbial imaging facility at UTS is acknowledged for the use of their epifluorescence microscope and Dr. Mike Johnson as well as Dr. Lynne Turnbull thanked for help. Janis Ortgies, Senckenberg am Meer, is thanked for help with the SEM preparations. Dr. Mireille Chinain (Institut Louis Malarde; French Polynesia) and Prof. Takeshi Yasumoto (Japan Food Research Laboratories, Tokyo) are acknowledged for providing the CTX and MTX standards,

respectively. We thank the Australian Research Council (FT120100704) for funding.

2.8 Conflict of interest disclosure

The authors report no conflict of interest in conducting this study.

2.9 Supplementary Material

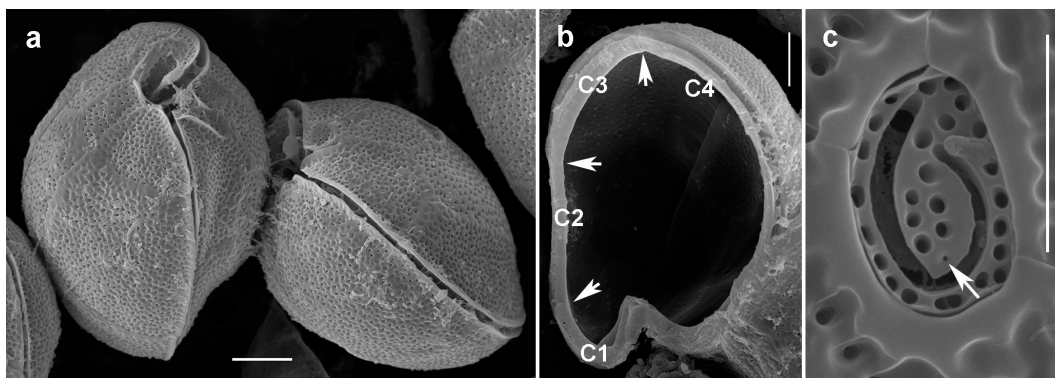


FIGURE 2.10 Supplementary scanning electron micrographs of *Gambierdiscus lapillus* (strain HG26). (a) Cell shape in lateral view. (b) Cingular plates (with the suture between C3 and C4 visible) with short arrows indicate the sutures. (c) Po plate with a small pore (arrow) next to the large depressions. Scale bars (a, b) 10 μm , (c) 5 μm .

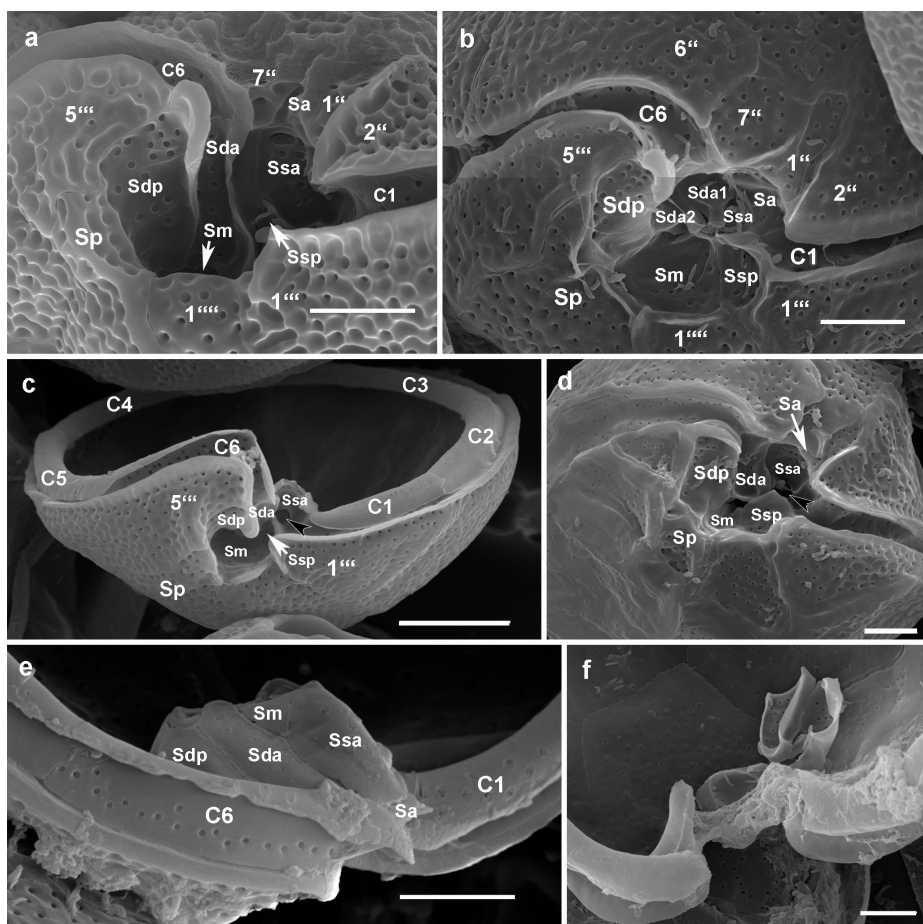


FIGURE 2.11 Supplementary scanning electron micrographs of *Gambierdiscus lapillus* showing sulcal and cingular details (a, b, d–f). Strain HG26 and (c) strain HG4. (a) Posterioventral view of a complete cell. (b) Ventral view of a young cell. (c) Ventral view of a split off hypotheca with cingular plates. (d) Ventral view. (e) Inside view of the apical part of the caved in sulcus at a split off hypotheca with cingulum. (f) Disassociated sulcal plates. Scale bars (a, b, d–f) 5 μm , (c) 10 μm .

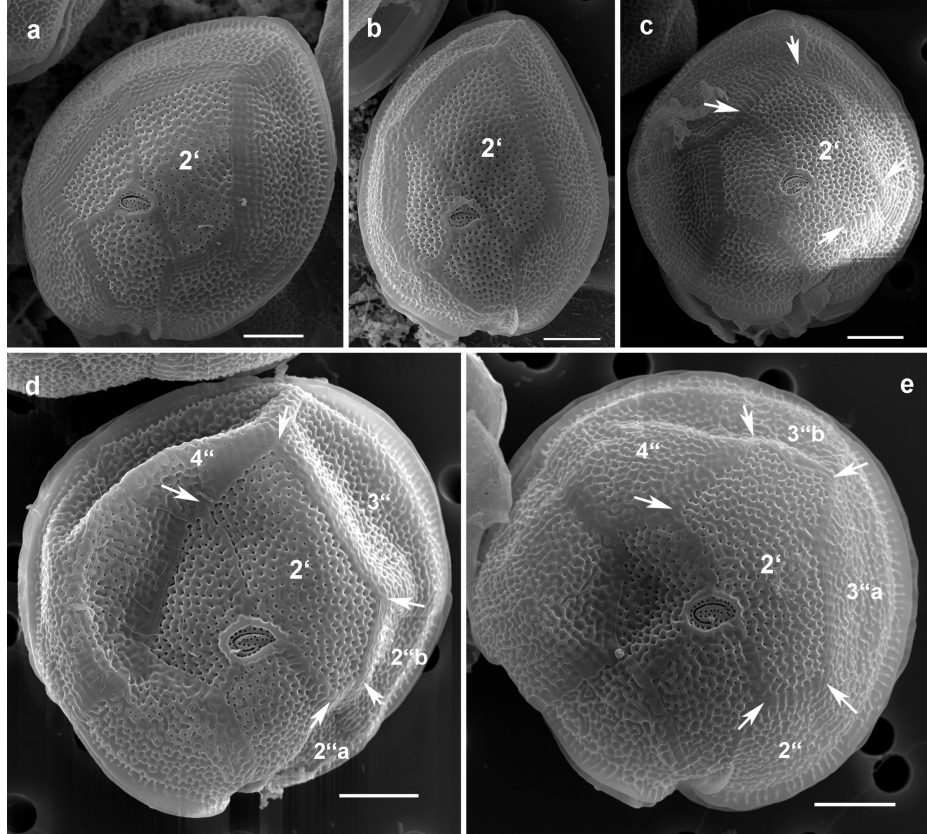


FIGURE 2.12 Supplementary scanning electron micrographs of *Gambierdiscus lapillus* (strain HG7) demonstrating the morphological variability in the second apical plate (2') and the cell shape. (a, b) Cells with teardrop-like shape and typical hatchet-shaped plate 2'. (c) Five-sided, nearly rectangular plate 2'. (d) Six-sided plate 2' and split second precingular plate (2''a and 2''b). (e) Six-sided plate 2' and split third precingular plate (3''a and 3''b). Arrows indicating plate corners.

CHAPTER 3

Development of a quantitative PCR assay for the detection and enumeration of a potentially ciguatoxin-producing dinoflagellate, *Gambierdiscus lapillus* (Gonyaulacales, Dinophyceae).

3.1 Authorship Declaration

By signing below I confirm that for the paper titled “Development of a quantitative PCR assay for the detection and enumeration of a potentially ciguatoxin-producing dinoflagellate, *Gambierdiscus lapillus* (Gonyaulacales, Dinophyceae)” and submitted to PLOS One, which is available on bioRxiv since February 2019, that:

Anna Liza Kretzschmar conceived and designed the experiment, developed and validated the qPCR assay, extracted DNA from environmental samples, performed qPCR experiments and analyzed the data, wrote the paper, prepared figures and tables, reviewed drafts of the paper.

Arjun Verma isolated the cultures of *Gambierdiscus lapillus*, extracted DNA from environmental samples, prepared figures and reviewed drafts of the paper.

Gurjeet Kohli collected the environmental samples, extracted DNA from environmental samples and reviewed drafts of the paper.

Shauna Murray conceived and designed the experiments, funded the research, and reviewed drafts of the paper.

Production Note:

Signature removed prior to publication.

Anna Liza Kretzschmar

Production Note:

Signature removed prior to publication.

Arjun Verma

Production Note:

Signature removed prior to publication.

Gurjeet Kohli

Production Note:

Signature removed prior to publication.

Shauna Murray

3.2 Abstract

Ciguatera fish poisoning is an illness contracted through the ingestion of seafood containing ciguatoxins. It is prevalent in tropical regions worldwide, including in Australia. Ciguatoxins are produced by some species of *Gambierdiscus*. Therefore, screening of *Gambierdiscus* species identification through quantitative PCR (qPCR), along with the determination of species toxicity, can be useful in monitoring potential ciguatera risk in these regions. In Australia, the identity, distribution and abundance of ciguatoxin producing *Gambierdiscus* spp. is largely unknown. In this study we developed a rapid qPCR assay to quantify the presence and abundance of *Gambierdiscus lapillus*, a likely ciguatoxic species. We assessed the specificity and efficiency of the qPCR assay. The assay was tested on 25 environmental samples from the Heron Island reef in the southern Great Barrier Reef, a ciguatera endemic region, in triplicate to determine the presence and patchiness of these species across samples from *Chnoospora* sp., *Padina* sp. and *Sargassum* macroalgal hosts.

3.3 Introduction

Benthic dinoflagellates of the genus *Gambierdiscus* Adachi & Fukuyo produce ciguatoxins (CTX), which can accumulate in humans via consumption of contaminated seafood and cause ciguatera fish poisoning (CFP) (Figure 3.1).

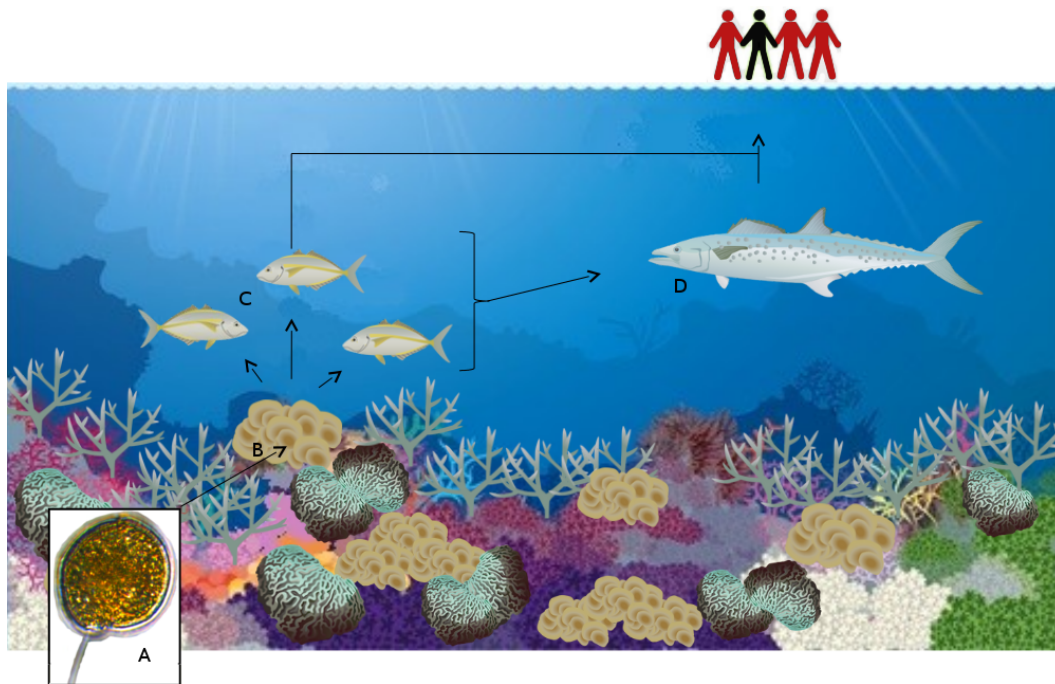


FIGURE 3.1 The mechanism of bioaccumulation of CTXs, with *Gambierdiscus* (for example *G. polynesiensis* (A)) at the base of the food web inhabiting the macroalgae *Padina* spp. (B) [279]. A herbivore, here a white trevally (*Pseudocaranx dentex*) (C) [127] consumes CTX from *G. polynesiensis* along with the macroalgae, which is then either passes directly to humans through consumption, or through an intermediary piscivorous vector such as Australian spotted mackerel (*Scomberomorus munroi*) (D) [238]. Image of *G. polynesiensis* (strain CG15) taken by A. L. Kretzschmar, 2016, Nikon Eclipse TS100 equipped with an Infinite Luminera 1 camera.

The symptoms of CFP are largely gastrointestinal and neurotoxic, however, in severe cases, further complications such as cardiovascular or severe neurological symptoms can appear [243]. In the most extreme cases, CFP can result in death [65]. Species of *Gambierdiscus* spp. are predominantly epiphytic, growing on macroalgae and other substrates such as coral detritus. Species of *Gambierdiscus* spp. can vary in the production of CTXs [30]. If a particular *Gambierdiscus* sp. is a CTX producer, and inhabit a palatable macroalgal substrate, the toxins bioaccumulate in herbivorous fish and filter feeders with the potential to travel up the food chain to cause CFP in humans [33, 91].

Gambierdiscus was first identified in 1977, with the type species *G. toxicus* Adachi & Fukuyo [2]. The genus remained monotypic for 18 years until the discovery of a second species *G. belizeanus* Faust [57]. To date, the genus comprises 14 described species and 6 ribo/species types [2, 31, 39, 57, 61, 62, 63, 128, 157, 158, 194, 218, 248, 282]. A major revision of the *Gambierdiscus* species taxonomy was undertaken by Litaker et al. (2009). Reports of *Gambierdiscus* spp. identified based on morphology alone, prior to this revision; need to be considered with caution as several new *Gambierdiscus* spp. were defined [92, 93, 94]. Further, intra-species variation and inter-species similarities can cause misidentification [22, 121, 128]. Hence, molecular genetic tools are important for determining the distribution and abundance of *Gambierdiscus* species and assess the risk of CFP in that region [121, 128].

Gambierdiscus spp. produce a suite of different polyketide compounds - CTX, maitotoxin (MTX), gambierone, gambieric acid and gambierol have been characterised to date [180, 181, 190, 226, 236]. While any of these can contribute to toxicity, only CTX has been clearly linked to CFP in humans [33, 91]. The toxin profile of many *Gambierdiscus* species is not well understood, and many different assays have been used to determine CTX toxicity [105]. Assays, such as mouse bioassays and neuroblastoma cell-line bioassays are good indicators of the toxicity of an organism, however species/strain specific toxin profiles needs to be elucidated with LC-MS/MS in order to characterise individual toxin congeners [49]. The toxin profile of *Gambierdiscus polynesiensis* Chinain & Faust is one of the only *Gambierdiscus* spp. whose production of CTX congeners (P-CTX-3B, P-CTX-3C, P-CTX-4A, P-CTX-4B and M-seco-CTX-3C) has been verified by LC-MS/MS in isolates from French Polynesia and the Cook Islands, and is thought to be the principal cause of CFP in the Pacific region [29, 217]. However recently, a *G. polynesiensis* strain isolated from the Kermadec Islands, Pacific Ocean, did not exhibit CTX toxicity detectable by LC-MS/MS [220]. An uncharacterised peak in the CTX phase of several strains of *Gambierdiscus lapillus* extracts was reported via LC-MS/MS, which did not match any available CTX standards (CTX-3B, CTX-3C, CTX-4A, CTX-4B) [128]. Further, Larsson et al. (2018) found that *G. lapillus* extracts showed CTX-like activity when investigated with a bioassay. Therefore, this species likely produces previously uncharacterised CTX congener(s), and its production of CTX compounds requires further investigation. Determining the toxin profile

of *Gambierdiscus* species requires toxin standards for comparative peak analysis. However, these are currently not commercially available. Therefore, progress in determining the toxins produced by species of *Gambierdiscus* has been comparatively slow, though bioassays provide a strong indicator for toxin production.

CFP was put forward as a "neglected tropical disease" by expert researchers in this area, supported by the Intergovernmental Oceanographic Commission's (IOC) Intergovernmental Panel on Harmful Algal Blooms (IPHAB), as part of the United Nations Educational, Scientific and Cultural Organization), and a global ciguatera strategy was developed [105]. One element of the IOC/IPHAB 'Global Ciguatera Strategy' is to investigate various species of the genus *Gambierdiscus*, determine which species produce CTXs through LC-MS/MS and other means, and develop efficient and reliable molecular monitoring tools for the species of interest [105]. Quantitative PCR (qPCR) is a useful molecular genetic screening tool, as it can give species-specific and quantitative results from DNA samples extracted from environmental samples [105].

Currently there is one qPCR assay to identify the presence of the genera *Gambierdiscus*/*Fukuyoa* [247]. Assays for species specific identification are available for 9 of the 14 described *Gambierdiscus* spp. and 3 out of 6 undescribed *Gambierdiscus* sp. types/ribotypes (Table 3.1). It is noteworthy that the qPCR assays described by Darius et al. (2017) rely on species identification based on the melt curve of the qPCR product, which requires any subsequent users of these assays to have a reference culture for positive identification rather than rely on a positive result being linked to the species investigated.

In the development of the qPCR assays for the quantification of microalgal species, several different methods have been used to quantify species [36]. Using a known cell number of the target species to construct standard curves for validating qPCR assays is a common strategy, however some genes, such as rRNA genes in dinoflagellates, can have gene copy numbers that vary significantly between strains. Hence comparing an assay developed with one strain as a standard might give irregular estimates of cell numbers when used for screening environmental samples [36]. An alternative method, using a synthetic oligonucleotide specifically designed for the assay tested, allows for a standard based on the amount of copies of a gene present rather than cell numbers. This has been successfully applied to *Alexandrium tamiyavanichii* [125], a sister genus to *Gambierdiscus*.

Assays are available for 2 of the 3 species of *Fukuyoa* (Table 3.1), which seceded from *Gambierdiscus* as their own genus in 2015 [73]. *Fukuyoa* spp. are of interest for monitoring purposes as they produce MTXs, however the involvement of MTXs in CFP has not been resolved yet [123].

TABLE 3.1 Published qPCR assays for *Gambierdiscus* and *Fukuyoa* spp.

Species	Method	Reference
<i>Gambierdiscus</i> spp.		
<i>G. australes</i>	TaqMan Probes & SYBR Green	[42, 192]
<i>G. belizeanus</i>	SYBR Green	[267]
<i>G. caribaeus</i>	SYBR Green	[267]
<i>G. carolinianus</i>	SYBR Green	[267]
<i>G. carpenteri</i>	SYBR Green	[267]
<i>G. jejuensis</i>	TaqMan Probes	[192]
<i>G. pacificus</i>	SYBR Green	[42]
<i>G. polynesiensis</i>	SYBR Green	[42]
<i>G. scabrosus</i>	TaqMan Probes	[192]
<i>G. toxicus</i>	SYBR Green	[42]
<i>Gambierdiscus</i> sp. ribotype 2	SYBR Green	[267]
<i>Gambierdiscus</i> sp. type 3	TaqMan Probes	[192]
<i>Fukuyoa</i> spp.		
<i>F. ruetzleri</i>	SYBR Green	[267]

In Australia, outbreaks of CFP occur annually in Queensland [208]. However, due to the complicated presentation of symptoms, the reporting rate is less than 20% [144]. Annually, there have been 7-69 reported cases between 2011 and 2015 (considering the report rate, > 35-345 cases, see [Table 3.2](#)), with 2 fatalities reported in the state [262]. Cases of CFP from Spanish Mackerel (*Scomberomorus commerson*) caught in NSW have been reported since 2014 [56], with five separate outbreaks affecting a total of 24 people since then [55]. Farrell et al. (2017) put forward a series of recommendations managing the emerging CFP risk in NSW.

TABLE 3.2 Cases of CFP reported to health authorities in Queensland, Australia, between 2011 and 2015, by Queensland Health [208].

Year	2011	2012	2013	2014	2015
Recorded CFP cases	18	7	25	69	11
Extrapolated CFP incidences	~90	~35	~125	~345	~55

Despite the prevalence of CFP in Australia, the characterization of *Gambierdiscus* species present in Australia is incomplete. A species that produces known CTX toxins has not been identified from Australia as yet. Larsson et al. (2018) have identified some candidate species, two of which show some CTX-like bioactivity [134]. Over 50% of Australia's vast coastline (total 66,000 km) is tropical or subtropical, and may be considered potential habitat for *Gambierdiscus* spp. [128]. Seven species of *Gambierdiscus* have been identified from the sub-tropical east Australian coastline namely, *G. belizeanus* [185], *G. carpenteri* [121, 250], *G. honu* (based on D8-D10 LSU sequence matching to a study by Richlen et al. [221]) [218], *G. lapillus* [128, 134], *G. toxicus* [80] and two yet undescribed species [134], as well as *F. paulensis* [185]. Using high throughput amplicon sequencing, *Gambierdiscus* was identified to the genus level in Broome, Western Australia [122], indicating that this is a coastline that should be examined further for CFP risk. qPCR primers that can be used for identification in Australia for potential monitoring purposes, have been developed for *G. belizeanus*, *G. carpenteri* and *F. yasumotoi* [192, 267].

The aim of this study was to develop a qPCR assay to exclusively amplify *G. lapillus*, as this has not yet been developed. The assay was then applied to environmental samples for the detection and enumeration of *G. lapillus* at Heron Island, GBR, in a region in which CFP cases are regularly reported. Hence this study adds to the suite of qPCR assays available to quantify organisms that contribute to CFP.

3.4 Materials and methods

3.4.1 Clonal strains and culturing conditions

Three strains of *G. lapillus* and one strain of *G. cf. silvae* were isolated from Heron Island, Australia, as previously described [128]. Two strains of *G. polynesiensis* were isolated from Rarotonga, Cook Islands (Table 3.3). Identification was performed using rDNA sequencing and phylogenetic inference, as previously described [128], and sequences have been submitted to GenBank (CG14: MH930987 for D1-D3 and MH915419 for D8-D10; CG15: MH930988 for D1-D3 and MH915420 for D8-D10). The cultures were maintained in 5x diluted F/2 media [94] at 27 °C, 60mol•m⁻² •-s light in 12hr light to dark cycles.

TABLE 3.3 List of *Gambierdiscus* clonal strains used for the qPCR assay.

Species	Collection site	Collection date	Latitude	Longitude	Strain name
<i>G. lapillus</i>	Heron Island, Australia	July 2014	23° 4420' S	151° 9140' E	HG4, HG7
<i>G. polyne-siensis</i>	Rarotonga, Cook Islands	November 2014	21° 2486' S	159° 7286' W	CG14, CG15
<i>G. holmesii</i>	Heron Island, Australia	July 2014	23° 4420' S	151° 9140' E	HG5

3.4.2 DNA extraction and species specific primer design

Genomic DNA was extracted from strains in Table 3.3 using a modified hexadecyltrimethylammonium bromide (CTAB) method [268]. The purity and concentration of the extracted DNA was measured using the Nanodrop (Nanodrop2000, Thermo Scientific), and the integrity of the DNA was visualised on 1% agarose gel. A unique primer set was designed for the small-subunit (SSU) rDNA region of *G. lapillus* based on sequences available in the GenBank reference database (accession numbers KU558929-33). The target sequences were aligned against sequences of all other *Gambierdiscus* spp. that were available on GenBank reference database, with the MUSCLE algorithm (maximum of 8 iterations) [53] used through the Geneious software, version 8.1.7 [113]. Unique sites were determined manually (Table 3.4, alignments available on request). Primers were synthesised by Integrated DNA Technologies (IA, USA). The primer set was tested systematically for secondary product formation for all 3 strains of *G. lapillus* (Table 3.3) via standard PCR in 25 µl mixture in PCR tubes. The mixture contained 0.6 µM forward and reverse primer, 0.4 µM BSA (Biolabs, Arundel, Australia), 2 - 20 ng DNA, 12.5 µl 2xEconoTaq (Lucigen Corporation, Middleton, WI, USA) and 7.5 µl PCR grade water. The PCR cycling comprised of an initial 10 min step at 94 °C, followed by 30 cycles of denaturing at 94 °C for 30 sec, annealing at 60 °C for 30 sec and extension at 72 °C for 1 min, finalised with 3 minutes of extension at 72 °C. Products were visualised on a 1% agarose gel.

TABLE 3.4 *G. lapillus* specific qPCR primer set for SSU rDNA designed in this study.

Primer name	Amplicon size	Synthesis direction of primer	Sequence (5'-3')
qGlapSSU2F	138bp	Forward	TTTTTGTCCCAGGAGGGTGA
qGlapSSU2R		Reverse	TGAGGCCAAACTCGAAAATC

3.4.3 Evaluation of primer specificity

To verify primer set specificity as listed in Table 3.4, DNA was extracted using CTAB buffer [297] from *G. australes* (CAWD149 and CG61), *G. belizeanus* (CCMP401), *G. carpenteri* (UTSMER9A3), *G. pacificus* (CCMP1650) and *G. cf. silvae* (HG5). *G. chelonae* (CAWD232) DNA was extracted using a PowerSoil™ DNA isolation kit (Mo Bio Inc., CA, USA). *G. scabrosus* (KW070922_1) DNA was extracted using DNeasy Plant Mini Kit (Quiagen, Tokyo, Japan) according to the manufacturer's protocol.

As DNA extracts for the purpose of specificity testing were supplied by four different researchers, the extraction methods varied. However, as these served as negative controls for primer specificity only, the difference in extraction methods would not be expected to impact any of the following cell enumeration methods.

For all extracted samples, the presence and integrity of genomic DNA was assessed on 1% agarose gel. The presence of PCR inhibitors was tested for each DNA extract by amplifying SSU region as per methodology in [128]. The primer set designed for *G. lapillus* was tested for cross-reactivity against all other *Gambierdiscus* spp. available via PCR (BioRadT100 Thermal Cycler (CA, USA)), appropriate positive and negative controls were applied. PCR amplicons were visually assessed on 1% agarose gel.

3.4.4 Evaluation of primer sensitivity

To test the primer sensitivity, qPCR assays were run with the specifications below. Initially the amplifications were screened for a single melt curve to show binding occurred at only one gene in the *G. lapillus* genome, then calibration curves to determine the range of detection were conducted.

The qPCR reaction mixture contained 10 μ l SYBR Select Master Mix (Thermo Fisher Scientific), 7 μ l MilliQ water, 0.5 μ M forward and reverse primers and 2 - 20 ng DNA template, for a final volume of 20 μ l. Cycling conditions consisted of 10 min at 95, then 40 cycles of 95 °C for 15 seconds and 60 °C for 30 seconds, followed by a temperature gradient for melt curve construction.

Standard curves were constructed to determine the efficiency of the assay, using a synthetic gene fragment approach, and also to use to quantify species presence, using calibration curves based on DNA extracted from known cell numbers.

The calibration curves for both methods were calculated (R^2 , PCR efficiency and regression line slope) and graphed in R version 3.2.3 [210], using R studio version 1.0.136 [231] and the ggplot2 package [276].

3.4.4.1 Gene based calibration curve

For the target amplicons of *G. lapillus*, a DNA fragment spanning the target sequence, the reverse and forward primer sites and an extra 50bp on either end was synthesised to a total length of 238bp (gBlocks®, IDT, IA, USA). The molecular weight and the amount of the synthesized gene

fragment was supplied by IDT, from which the exact number of copies of the gene fragment per micoliter can be calculated [36, 125]. Lyophilized gBlocks[®] was re-suspended in 1x TE (Tris 1M, EDTA 0.5 pH8) to a concentration of 1 ng/ μ l. The copy number of gene fragment was then calculated as 2.88×10^{10} for *G. lapillus*. The stock solution was serially diluted (10-fold) and dilutions between 10^3 and 10^8 were amplified by qPCR (on StepOnePlus System by Applied Biosystems (Thermo Fisher Scientific, Waltham, MA, USA) in triplicate.

3.4.4.2 Cell based calibration curve

Two strains of *G. lapillus* (HG4 and HG7) were used to construct cell based standard curves. Cells were counted under a Nikon Eclipse TS100 (Australia) microscope using a Sedgwick Rafter counting chamber. DNA was extracted with the FastDNA spin kit for soil by MP Biomedicals (CA, USA), as per the manufacturer's instructions. The gDNA extracts were 10-fold serially diluted. Dilutions ranging from 3880 to 0.04 cells and 5328 to 0.05 for HG4 and HG7 respectively. Samples were amplified via qPCR (on StepOnePlus System by Applied Biosystems (Thermo Fisher Scientific, Waltham, MA, USA) in triplicate.

3.4.5 Determination of gene copies per cell for *G. lapillus*

To determine the mean SSU rDNA copies per cell, a 10-fold dilution series of strains HG4 and HG7 were used, as described previously. These dilution series were based on DNA extracted from known numbers of cells, and then serially diluted (3880 to 0.04 cells and 5328 to 0.05 for HG4 and HG7 respectively).

The slope of the linear regression of SSU copies was used to determine copy number by correlating the qPCR detection of the gene based calibration curves and cell numbers. This slope of the linear regression was then used to determine the gene copy number per cell [125].

3.4.6 Screening environmental samples for *G. lapillus*

Around Heron Reef (Figure 3.2) 25 sites (within 1km of the shore) were sampled in October 2015, in spatial replicates (A, B, C) within a 2m radius. Representatives of three genera of macroalgae that commonly grow on this reef, *Chnoospora* spp, *Padina* sp. and *Saragassum* sp., were sampled for the presence of epiphytic *Gambierdiscus* spp. For each sample, about 200 g of macroalgae was collected from approximately 1 m deep water at low tide and briefly placed in plastic bags containing 200 to 300 ml of ambient seawater.

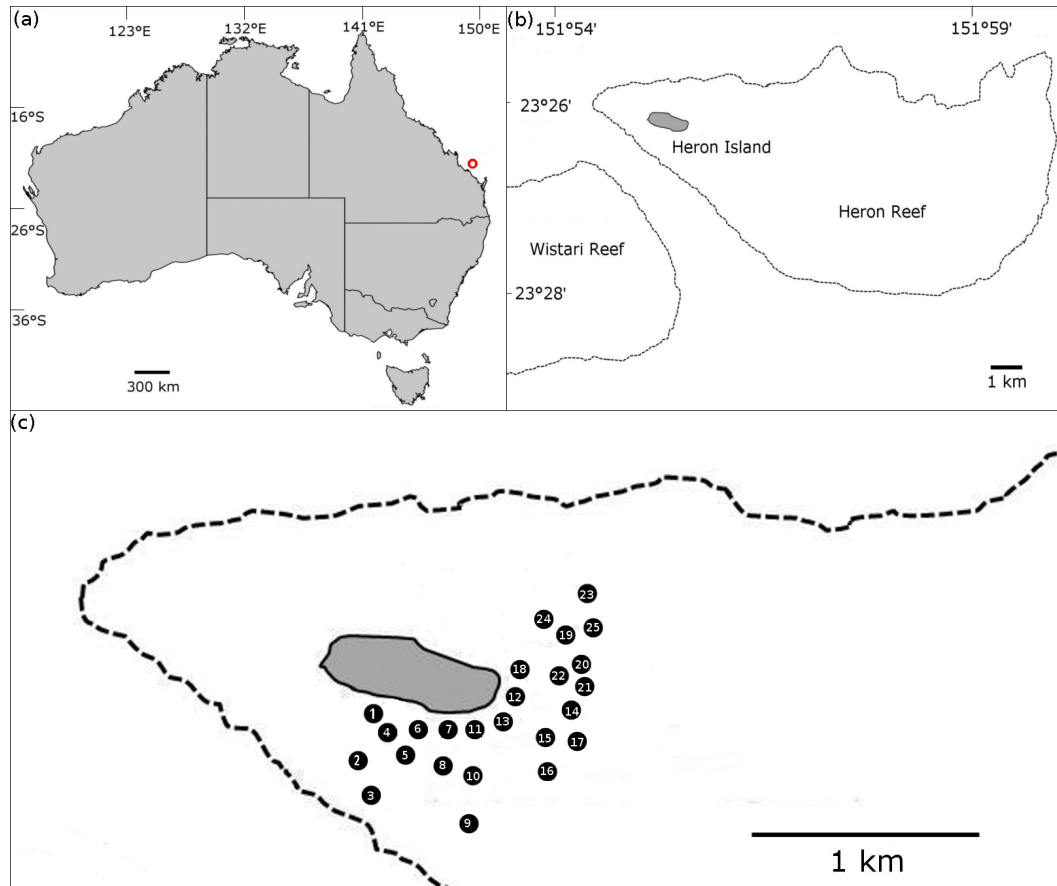


FIGURE 3.2 (A) Map of Australia, with the position of Heron Island (red circle); (B) Heron Island including surrounding reefs (dotted lines); (C) Approximate location of sampling sites around Heron Island.

They were shaken vigorously for 5 min to detach the epiphytic dinoflagellates from the macroalgae. This seawater was passed through $> 120 \mu\text{m}$ mesh filter to remove any remaining larger fauna and debris. The collected seawater was centrifuged at 500 rcf. The supernatant was discarded and the pellet was dissolved in 10 ml RNAlater (Ambion®, Austin, TX, USA) for preservation and stored at 4°C. Community DNA was extracted via modified CTAB method [268]. Samples were screened in triplicate for both *G. lapillus* on a StepOnePlus System by Applied Biosystems (Thermo Fisher Scientific, Waltham, MA, USA). The amplification threshold was compared to the cell based calibration curve using the *G. lapillus* type species, HG7. The cells.g⁻¹ wet weight macroalgae was calculated by determining the proportion of the total volume of DNA extract used in an individual qPCR reaction (i.e. 1 μl DNA per qPCR reaction from a 50 μL total DNA extraction volume = 0.02), quantifying the equivalent number of cells detected per qPCR reaction using the standard curve and multiplying this to determine total cells 200 g⁻¹ wet weight macroalgae sample. Normality of the data (cell numbers per macroalgal host) was tested using Shapiro test and a Bartlett test of homogeneity of variance a Kruskal-Wallis test was performed to determine significance, in R version 3.2.3 [210], using R studio version 1.0.136 [231].

3.5 Results

3.5.1 Evaluation of primer specificity

The qGlapSSU2F - qGlapSSU2R primer pair (Table 3.4) produced an amplified product of 138 bp for all five strains of *G. lapillus*, with a single peak at the same temperature for each strain in the melt curve. No signal for primer dimers or unspecific amplification was detected. While genomic DNA was visible for each strain on the agarose gel (gDNA band, Table 3.5) and the DNA could be amplified via PCR (SSU PRC amplification, Table 3.5), no cross-reaction was observed for genetically closely related species *G. belizeanus*, *G. cheloniae*, *G. pacificus* and *G. scabrosus*. Other species of *Gambierdiscus* from different clades, *G. australes*, *G. carpenteri*, *G. polynesiensis* and *G. cf. silvae* (Table 3.5) were also not amplified using this primer set [128, 248].

TABLE 3.5 Cross-reactivity of the qPCR primer set based on presence absence of PCR product visualised in agarose gel.

Template	Strain name	gDNA gel band	GlapSSU2F-GlapSSU2R	
<i>G. australes</i>	CAWD149	+	+	-
	CG61	+	-	
<i>G. belizeanus</i>	CCMP401	+	+	-
<i>G. carpenteri</i>	UTSMER9A3	+	+	-
<i>G. cheloniae</i>	CAWD232	+	+	-
<i>G. lapillus</i>	HG1	+	+	+
	HG4	+	+	+
	HG6	+	+	+
	HG7	+	+	+
	HG26	+	+	+
<i>G. pacificus</i>	CCMP1650	+	+	-
<i>G. polynesiensis</i>	CG14	+	+	-
	CG15	+	+	-
<i>G. scabrosus</i>	KW070922_1	+	+	-
<i>G. cf. silvae</i>	HG5	+	+	-
<i>O. cf. ovata</i>	HER27	+	+	-
<i>O. rhodesae</i>	HER26	+	+	-
<i>O. cf. siamensis</i>	HER24	+	+	-

3.5.2 Evaluation of primer sensitivity

The cell-based standard curves for *G. lapillus* (HG4 and HG7, [Figure 3.3a](#)) showed high linearity with R^2 approaching 1.00. The slope for the Ct vs. $\log_{10}$ cell for HG4 was -3.4, which corresponds to an efficiency 96.8 % ; and -3.51, which corresponds to an efficiency of 92.7 % for HG7 ([Figure 3.3](#)). The linear detection for both *G. lapillus* isolates covered five orders of magnitude. The lowest number of cells detected were 0.04 and 0.05 cells for HG4 and HG7 respectively ([Figure 3.3a](#)).

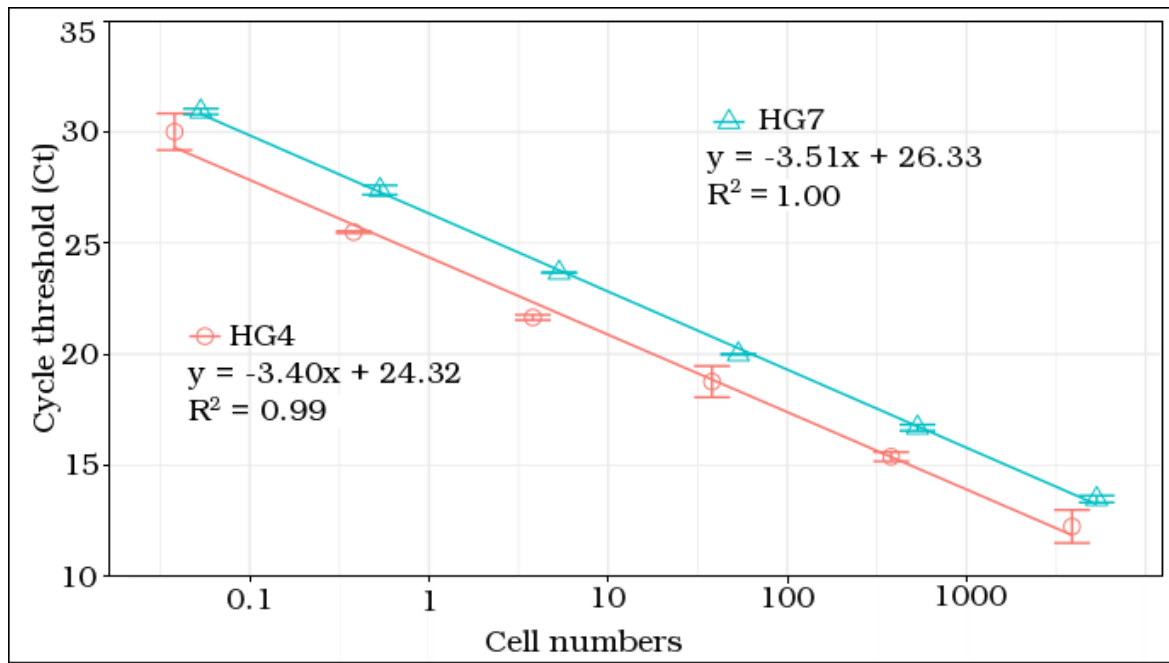


FIGURE 3.3 qPCR cell based standard curves of *G. lapillus* strains HG4 (circle) and HG7 (triangle). Error bars represent the deviation of technical replicates during reactions; x-axis is log scale.

The gene based (gBlocks) standard curve for *G. lapillus* covered linear detection over five orders of magnitude, with a slope of -3.42, and a PCR efficiency of 96 % (Figure 3.4). The detection limit tested was less than 10^5 gene copy numbers. The Ct for the lowest gene copy number tested was less than 25, so it is likely that the sensitivity is lower than 10^5 gene copy numbers (Figure 3.4).

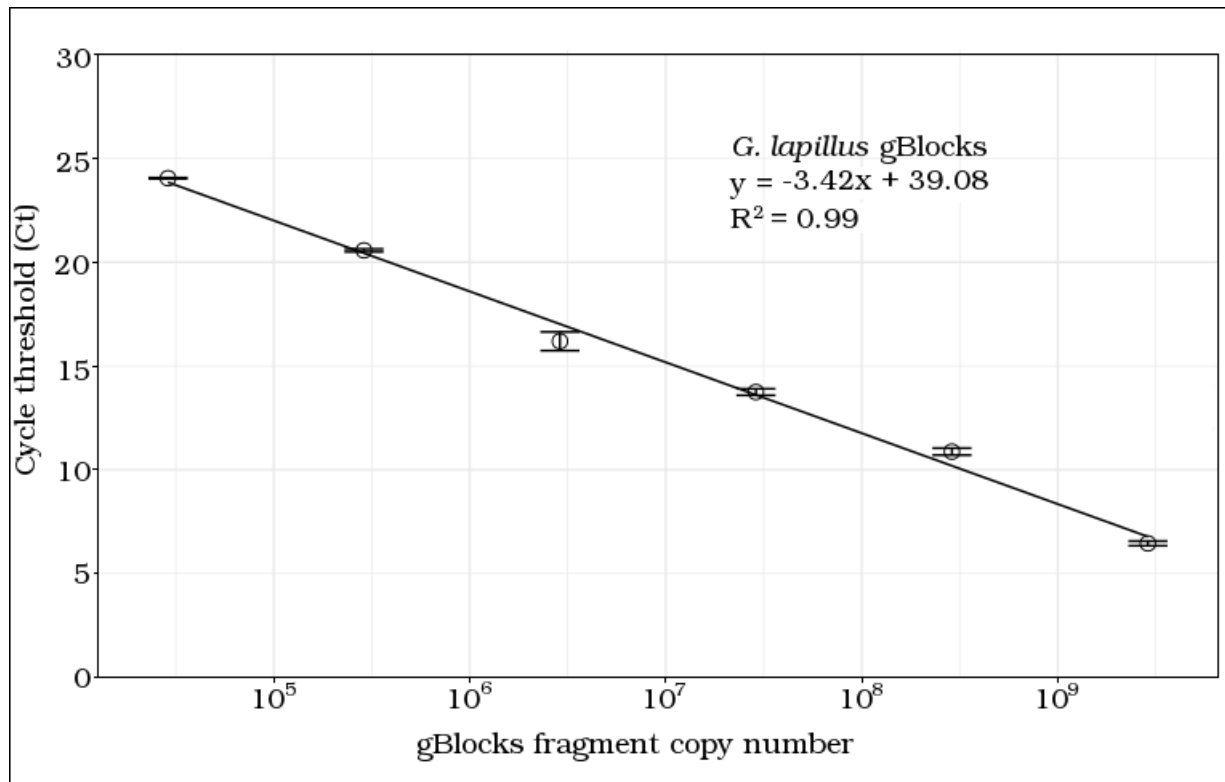


FIGURE 3.4 qPCR gene based standard curves of *G. lapillus*. Error bars represent the deviation of technical replicates during reactions; x-axis is log scale.

3.5.3 Quantification of SSU rDNA copy number per cell of *G. lapillus*

The detectable SSU copies for *G. lapillus* were 2.24×10^4 and 5.85×10^3 copies per cell for HG4 and HG7 respectively.

3.5.4 Screening environmental samples for *G. lapillus* abundance

To evaluate the adequacy of the *G. lapillus* qPCR assay for environmental screening, the assay was applied to environmental community DNA extracts collected around Heron Island (Figure 3.2). A relatively low cell abundance was detectable for *G. lapillus*. Ct values for *G. lapillus* detection in environmental samples were calibrated to the HG7 standard curve and calculated as cells.g^{-1} wet weight macroalgae (Table 3.6). *G. lapillus* was detected across 24 of the 25 sampling sites. At sites at which *G. lapillus* was present, it showed a patchy distribution, being present at two of the three spatial replicates in the majority of samples (17 of 25 sample sites), followed by all three spatial replicates testing positive (6 out of 25 sites) and at one site only one of the spatial replicates was positive (Figure 3.5).

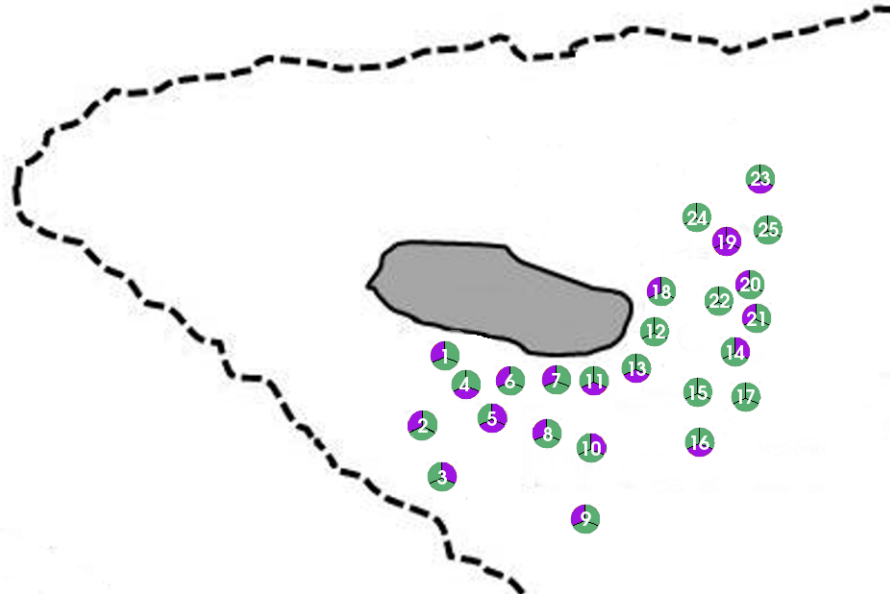


FIGURE 3.5 *G. lapillus* presence at the macroalgal sampling sites around Heron Island. The spatial replicates for each site are set up as shown in (A); the sites in (B) linked to numbering in [Figure 3.2](#) where positive (green) and negative (purple) as per [Table 3.6](#).

G. lapillus was detected at 71 out of the 75 spatial replicates, specifically at 24/32, 22/33 and 8/10 samples from *Chnoospora* sp., *Padina* sp. and *Saragassum* sp. as substrate respectively ([Table 3.6](#)). Patchiness was also found in the abundance as well as the distribution of *G. lapillus*, from 0.24 cells.g⁻¹ wet weight macroalgae to 49.51 cells.g⁻¹ wet weight macroalgae, with a mean of 5.84 cells.g⁻¹ wet weight macroalgae. For example (4A - *Chnoospora* sp.) and (4B - *Padina* sp.) hosted comparable cell numbers (1.12 cells and 1.65 cells.g⁻¹ wet weight algae respectively) while no *G. lapillus* cells were detected on (4C - *Padina* sp.). Only at one of 25 sampling sites, no *G. lapillus* presence was detected across all three spatial replicates (19A, B, C). At all other sites, the presence of *G. lapillus* varied between spatial replicates but did not significantly differ between macroalgal host or location (chi-squared = 2.1453, p-value = 0.3421) ([Figure 3.6](#)).

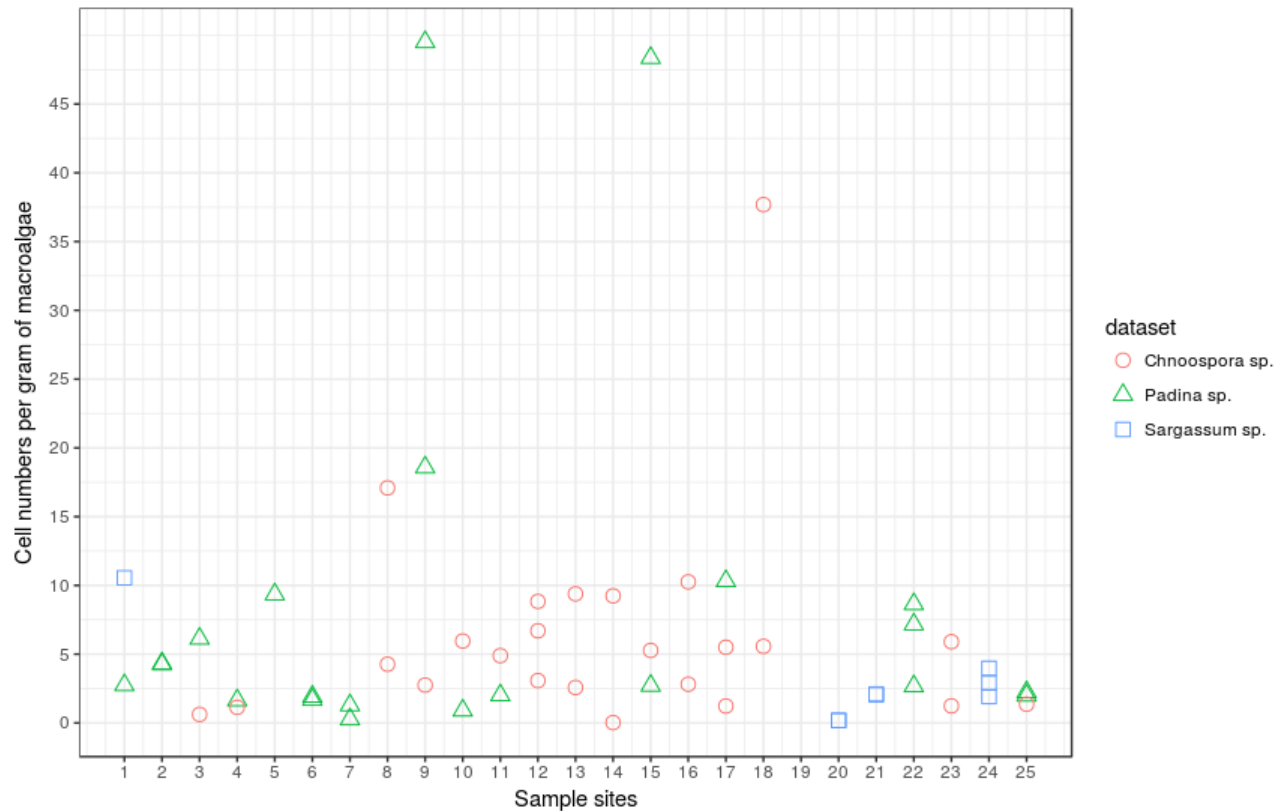


FIGURE 3.6 Detection of *G. lapillus* per spatial replicate at each macroalgal sampling site. Cell numbers were normalised to the HG7 standard curve (Figure 3.3A). Figure also shows spatial replicates per macroalgal substrate where *Chnoospora* sp. samples are represented by circles, *Padina* sp. by squares and *Sargassum* by crosses (Table 3.6).

3.6 Discussion

The aim of the study was to design and validate a species-specific qPCR assay to quantify *G. lapillus* a species that may produce CTX- like toxicity in the Australian GBR region. Species-specific qPCR primers with high specificity and sensitivity were developed and the SSU copy number for two strains were determined, and were found to differ from one another considerably, as one strain had more than four times the number of genomic rDNA copies. We also established that this primer set was effective in measuring the abundance and distribution of *G. lapillus* at the Heron Island reef. The cross-reactivity of primers designed in this study showed high specificity for both *G. lapillus* while not amplifying when tested against other *Gambierdiscus* spp. The species tested for cross-reactivity were chosen because they represented species that are genetically most similar to each target species for the SSU region (as per Fig. 2 in [128]). Standard curves were constructed for two strains of *G. lapillus* for which the primers showed high linearity and amplification efficiency (Figure 3.3). Hence, this primer set is an accurate and reproducible molecular tool to enumerate the target species exclusively from environmental community DNA extracts. More importantly, this assay does not require the operator to rely on melt curves to identify species, or to have access to *G. lapillus* DNA extracts as a positive control. Due to the potential CTX production of *G. lapillus* [128, 134] the presence and distribution of this species is of interest in Australia where the causative organism(s) for CFP is yet to be established.

As CFP risk is linked to the abundance of *Gambierdiscus* species producing CTXs [15, 105], it was important to establish a quantitative assay for detection. We validated a synthetic gene fragment standard curve of the target region (gBlocks[®]) and compared this to cell standard curves to establish an 'absolute' qPCR assay [84, 192]. Further, we determined the copy SSU rDNA number for two strains of *G. lapillus* (HG4 and HG7). The copy number for *G. lapillus* (5,855.3 to 22,430.3 rDNA copies per cell) were comparable to the copy numbers determined by Vandersea et al. (2012), which ranged from 690 rDNA copies for *G. belizeanus* to 21,498 copies for *G. caribaeus*. In comparison, the cell copy numbers determined by Nishimura et al. (2016) ranged from 532,000 copies for *G. scabrosus* and 2,261,000 for *G. sp. type 3*. While the difference in rDNA copy numbers may be due to inter-species differences, or even intra-species as per the *G. lapillus* results, Nishimura et al. (2016) argue that the difference could be underestimation of rDNA copy numbers due to 'ghost' cells (cells that look viable during cell counts under the microscope, but which are not living and therefore do not contain amplifiable DNA) [84, 192]. The difference observed in the SSU copy numbers between the two strains of *G. lapillus* could similarly be due to ghost cells. Though the cell counts and extractions were conducted by the same operator and at the same time, a repetition of the cell counts before dilution is recommended for future studies. Further to that, variation in DNA extraction efficiency could also contribute to the difference in observed SSU copy numbers between the *G. lapillus* strains [73]. The results

presented highlight the importance of carefully verifying qPCR assays based on rRNA genes using multiple local strains, as their target gene copy numbers might vary, but also the necessity and possible issues that can arise while constructing “absolute” standard curves. Tentatively, a difference of this magnitude in SSU copy numbers may lead to considerably different abundance estimates of *G. lapillus*. As the variation between the two strains tested is within the observed variation reported by Nishimura et al. (2016) from single cell qPCR experiments for rDNA copy number elucidation, the difference reported here is likely representative of biological intra-strain variation rather than methodological artifacts. A 5-fold difference in toxicity between the same HG4 and HG7 strains for *G. lapillus* was also reported by Kretzschmar et al. (2017), and there was a noticeable difference in growth rate between the two strains observed (but not quantified) in this study. The mounting evidence of intra-strain variability in toxicity, detectable rDNA copy numbers and potentially growth rate could have severe implications for qPCR based cell enumeration of environmental samples when attempting to extrapolate CFP risk and requires further investigation.

The qPCR assay was successfully tested on environmental DNA extracts from around Heron Island, and gave some insight into *G. lapillus* distribution and abundance. The qPCR assay detected *G. lapillus* at all of the sites tested (Figure 3.5). Within the spatial replicates, the distribution of *G. lapillus* was patchy, as 24 of the 25 sites included at least one replicate with no *G. lapillus* present (Figure 3.5). Patchiness in the distribution of *Gambierdiscus* species has previously been found in a study of 7 *Bryothamnion* macroalgae spaced 5 to 10 cm apart, in which 5 to 70 cells g⁻¹ algae were found [257].

There was no significant difference in the presence/absence of *G. lapillus* cells observed as per the macroalgal host, *Padina* sp. or *Sargassum* sp.

Motile behaviour has been observed previously in the field at various time points [17, 286]. Parsons et al. (2011) reported *Gambierdiscus* sp. behaviour as facultative epiphytes during lab scale experiments, as cells showed attachment as well as motile stages over time in the presence of different macroalgae [197]. Taylor & Gustavson (1983) reported that *Gambierdiscus* cells were captured in plankton tows by de Silva in 1956 but reported as *Goniodoma* [257]. Motility could be a factor for the patchy distribution observed in the spatial replicates. Across spatial replicates where *G. lapillus* was detected, cell densities were consistent (Figure 3.6). The average cell density of *G. lapillus* 5.84 cells.g⁻¹ wet weight macroalgae, which is comparable to the cell densities recorded by Nishimura et al. (2016) in their environmental screening (Table 4 in [192]).

As many authors have pointed out (e.g. [19, 38, 105, 158, 161, 197, 258]), there are several difficulties in determining precise quantification of *Gambierdiscus* species on macroalgae in order to assess potential CFP risk. Due to the difference in habitable surface area between samples taken from structurally diverse macroalgae, including those sampled in this study (*Chnoospora*, *Padina* sp. and *Sargassum* sp.), the potential habitable space is difficult to compare. Further, in order to

assess CFP risk in a given area, the properties of the macroalgae with *Gambierdiscus* spp. epiphytes need to be considered. If the macroalgae is structurally or chemically defended against herbivory, any CTX produced by the epiphytes is unlikely to enter the food chain and cause CFP [38]. Due to the difficulty in quantifying *Gambierdiscus* spp. on a particular substrate, Tester et al. (2014) proposed have the use of an artificial substrate (commonly available black fibreglass screen of a known surface area) and a standardised sampling method [258]. Molecular analysis, such as species specific qPCR, based on this standardised sampling method would be directly comparable across sampling sites and times. Adopting this approach for future monitoring studies is recommended.

3.7 Conclusion

The qPCR assay developed in this study is an accurate molecular tools to detect and enumerate the presence of *G. lapillus* in environmental samples. The assay was shown to be highly sensitive and accurately detected 0.05 to over 4000 cells for *G. lapillus*. Although the toxin profile of *G. lapillus* has not been completely defined, it may produce uncharacterised CTXs congeners [128, 134] and therefore is a part of the ciguateric web in Australia. The assay was applied to spatial replicates from 25 sites around Heron Island on the GBR, which found that *G. lapillus* was commonly present, but had a patchy spatial distribution and abundance. The development and validation of a quantitative monitoring tool presented here for *G. lapillus* is in line with Element 1 of the Global Ciguatera Strategy [105].

3.8 Acknowledgements

We are grateful to Dr. Masao Adachi and Dr. Tomohiro Nishimura for supplying *G. scabrosus*, Michaela Larsson for supplying *G. carpenteri* DNA as well DNA as Dr. Kirsty Smith and Dr. Lesley Rhodes for supplying *G. cheloniae* DNA. DNA from all three species was used for cross reactivity assessment. A.L.K. was supported by a UTS doctoral scholarship funded by the University of Technology Sydney. S.M and G.K's involvement was supported by the Australian Research Council grant FT120100704.

3.9 Supplementary Material

TABLE 3.6 Supplementary screening of macroalgal samples for *G. lapillus* and cell density estimates via qPCR. Cell numbers were modeled on the type strain HG7. N/D denotes not detected; N/A denotes not attempted due to loss of sample.

Sample ID	Spatial replicate	Macroalgal substrate	<i>G. lapillus</i> cell number
1	A	<i>Padina</i> sp.	N/D
1	B	<i>Sargassum</i> sp.	10.55
1	C	<i>Padina</i> sp.	2.75
2	A	<i>Padina</i> sp.	N/D
2	B	<i>Padina</i> sp.	4.33
2	C	<i>Padina</i> sp.	4.27
3	A	<i>Padina</i> sp.	6.13
3	B	<i>Chnoospora</i> sp.	0.62
3	C	<i>Padina</i> sp.	N/D
4	A	<i>Chnoospora</i> sp.	1.12
4	B	<i>Padina</i> sp.	1.65
4	C	<i>Padina</i> sp.	N/D
5	A	<i>Padina</i> sp.	9.35
5	B	<i>Padina</i> sp.	N/D
5	C	<i>Padina</i> sp.	N/D
6	A	<i>Chnoospora</i> sp.	N/D
6	B	<i>Padina</i> sp.	1.69
6	C	<i>Padina</i> sp.	1.92
7	A	<i>Padina</i> sp.	N/D
7	B	<i>Padina</i> sp.	0.26
7	C	<i>Padina</i> sp.	1.29
8	A	<i>Chnoospora</i> sp.	N/D
8	B	<i>Chnoospora</i> sp.	17.09
8	C	<i>Chnoospora</i> sp.	4.27
9	A	<i>Chnoospora</i> sp.	N/D
9	B	<i>Padina</i> sp.	49.51
9	C	<i>Padina</i> sp.	18.58

10	A	<i>Padina</i> sp.	0.91
10	B	<i>Padina</i> sp.	N/D
10	C	<i>Chnoospora</i> sp.	5.95
11	A	<i>Padina</i> sp.	2.01
11	B	<i>Chnoospora</i> sp.	4.89
11	C	<i>Chnoospora</i> sp.	N/D
12	A	<i>Chnoospora</i> sp.	6.70
12	B	<i>Chnoospora</i> sp.	8.83
12	C	<i>Chnoospora</i> sp.	3.08
13	A	<i>Chnoospora</i> sp.	2.58
13	B	<i>Chnoospora</i> sp.	9.39
13	C	<i>Chnoospora</i> sp.	N/D
14	A	<i>Chnoospora</i> sp.	0.02
14	B	<i>Chnoospora</i> sp.	N/D
14	C	<i>Chnoospora</i> sp.	9.24
15	A	<i>Chnoospora</i> sp.	5.27
15	B	<i>Padina</i> sp.	48.46
15	C	<i>Padina</i> sp.	2.71
16	A	<i>Chnoospora</i> sp.	2.81
16	B	<i>Chnoospora</i> sp.	10.26
16	C	<i>Chnoospora</i> sp.	N/D
17	A	<i>Chnoospora</i> sp.	5.50
17	B	<i>Chnoospora</i> sp.	1.23
17	C	<i>Padina</i> sp.	10.32
18	A	<i>Chnoospora</i> sp.	N/D
18	B	<i>Chnoospora</i> sp.	37.68
18	C	<i>Chnoospora</i> sp.	5.57
19	A	<i>Padina</i> sp.	N/D
19	B	<i>Padina</i> sp.	N/D

19	C	<i>Padina</i> sp.	N/D
20	A	<i>Sargassum</i> sp.	N/D
20	B	<i>Sargassum</i> sp.	0.19
20	C	<i>Sargassum</i> sp.	0.18
21	A	<i>Sargassum</i> sp.	N/D
21	B	<i>Sargassum</i> sp.	2.11
21	C	<i>Sargassum</i> sp.	2.05
22	A	<i>Padina</i> sp.	7.17
22	B	<i>Padina</i> sp.	2.67
22	C	<i>Padina</i> sp.	8.64
23	A	<i>Chnoospora</i> sp.	1.24
23	B	<i>Chnoospora</i> sp.	5.90
23	C	<i>Chnoospora</i> sp.	N/D
24	A	<i>Sargassum</i> sp.	1.91
24	B	<i>Sargassum</i> sp.	2.90
24	C	<i>Sargassum</i> sp.	3.97
25	A	<i>Padina</i> sp.	2.24
25	B	<i>Chnoospora</i> sp.	1.36
25	C	<i>Padina</i> sp.	2.00

CHAPTER 4

Using comparative transcriptomics for the pan-transcriptome of *Gambierdiscus*.

Anna Liza Kretzschmar¹, Shauna Murray², Aaron E. Darling¹, Tim Kahlke²

¹The ithree institute, University of Technology Sydney, Ultimo, 2007 NSW, Australia

²Climate Change Cluster (C3), University of Technology Sydney, Ultimo, 2007 NSW, Australia

4.1 Abstract

The dinoflagellate species of the genus *Gambierdiscus* are reported to produce ciguatoxins (CTXs), the causative agent of ciguatera fish poisoning, a debilitating seafood-borne illness. Although the specific toxins produced by *Gambierdiscus* remain largely unknown, it has been verified using LC-MS/MS that multiple strains of *G. polynesiensis* produce CTX analogs. Bioassays have implicated several more species of *Gambierdiscus* with CTX-like bioactivity. An understanding for the evolution of *Gambierdiscus* and their toxins requires the availability of reference genomes. However, *Gambierdiscus* species possess very large (32-35 Gbps) and highly repetitive genome and complex genome architecture making genome sequencing of these microalgae almost impossible, a common feature shared among the dinoflagellates. To overcome the challenges of genome sequencing, we generated *de novo* RNA assemblies of four species, *G. carpenteri*, *G. holmesii*, *G. lapillus* and *G. polynesiensis* from [Section 5](#) (as [Section 5](#) has been submitted for publication, the methodology here will reference that chapter despite preceding it in the thesis), for a comparative study with the previously sequenced *G. australes*, to determine transcripts that are shared among the investigated strains. Here we present a *Gambierdiscus* pan-transcriptome, exploring the commonalities and uniquely expressed protein coding sequences for each species, which can be used for future *de novo* sequencing efforts for *Gambierdiscus*.

4.2 Introduction

A challenging aspect of protist *de novo* sequencing projects lies in the lack of annotated closely related references for comparison, without which it is difficult to assess the adequacy and completeness of sequencing, library processing and assembly methods employed. This issue is particularly relevant for dinoflagellates, whose expansive and complex genomes and genetic machinery tend to be a barrier to genomic sequencing. As an alternative to wrangling with dinoflagellate genomes, transcriptomes can be used to explore their genetics (see [Section 1.3](#) and [Section 5.3.2](#)). Dinoflagellates appear to have an uncharacterized genetic regulatory mechanism which leaves protein synthesis regulation to the post-transcriptional stage. For this reason, mRNA may give an insight into the genomic content of the organism. An indication of these regulatory mechanisms comes from a number of direct observations. [85] cultured the dinoflagellates *Prorocentrum minimum* and *Alexandrium monilatum* under stress conditions by severely limiting nitrogen as well as phosphorous availability. The cultures showed significant biochemical changes (e.g. altered growth rate, particulate organic carbon and particulate carbohydrates content) between the control and stress conditions at time of harvest, yet change in transcriptome expression was minimal, between 0.1 to 1 % depending on stressor and species used [85]. While the difference in biochemical changes was not captured by mRNA profiling of the cultures, the study did not include a protein expression observation to verify a difference in expression despite a static pool of mRNA availability [85].

Due to the relative difficulty in maintaining dinoflagellates in culture and extracting their RNA, the number of available marine dinoflagellate transcriptomes was limited until the Marine microbial eukaryote transcriptome sequencing project (MMETSP) data release. When searching for *Gambierdiscus* on NCBI's SRA database, 5 relevant projects were found in addition to the MMETSP results (searched on November 10, 2018). These sequencing projects covered two strains of *G. polynesiensis*, as well as for *G. australes* and *G. excentricus*. The fifth project focused on the bacterial associations of *G. caribaeus* and *G. carolinianus*. Broadening the search to the order gonyaulacales yielded a further 19 projects, including another on bacterial associates as well as 3 projects on *Azadinium* and *Crypthecodinium*, which are arguably not part of the gonyaulacales (Section 5.5.3.4). Searching for members of the phylum dinoflagellata yielded a further 84 projects. Despite their ecological relevance for nutrient cycling, dimethylsulfoniopropionate production, coral symbiosis and neurotoxin production (for a review see [188]), a paucity of sequencing data, even with the MMETSP dataset, is evident. This is further confounded by a large proportion of dinoflagellate transcripts sharing little or no similarity to known proteins or domains in public databases [254]. When compared to NCBI's non-redundant database, the proportion of contigs with no known match was 60 % for *Azadinium spinosum* [171], over 50 % for *G. australes* & *G. belizeanus* [119], 57.9 % for *G. excentricus* [117], 63 % for *G. polynesiensis* [117, 198], and 55 - 57 % for *Karenia brevis* [232]. In summary, dinoflagellate genomic work is challenging due to their unusual genetics and limited similarity to previously characterised organisms.

The concept of a reference genome or transcriptome allows for direct comparison of sequencing data to a high quality standard. However, it has become apparent that a reference may not be sufficient in capturing the genetic diversity inherent to living organisms. This was first conceived by [260] from the observation that genomes in bacteria revealed a large transitory subset of genetic content. They concluded that a single strain-based reference would be inadequate for capturing a large proportion of the species' genetic diversity and coined the concept of pan-transcriptomics [260, 270]. This is defined by a genomic reference for a species composed of a core genome common to all strains, an optional soft-core common to most strains, and an accessory (alternatively called unique) genome which is strain specific. Since then the pan-genome, and a similarly conceived pan-transcriptome, concept has been adopted for eukaryotes, with the realization that the accessory genomic content exists when multiple strains of a species are sequenced (e.g. [109, 150, 201, 202, 214, 249]). The pan-genome analyses concept has been applied at different taxonomic levels, from intra-genus to super kingdom [86, 109, 124, 133, 270].

Five transcriptomes of *Gambierdiscus* were compared in this study with the aim of providing a pan-transcriptome for *Gambierdiscus de novo* transcriptome sequencing, which can be expanded and refined in future studies. The taxa originated from two locations in Australia (Merimbula, NSW, and Heron Island, QLD) and Rarotonga in the Cook Islands (Table 4.1). All of the five species have shown MTX-like bioactivity, while *G. carpenteri* did not register for CTX-like activity in a bioassay [134]). The LC-MS/MA based toxin profiles indicated all species apart from *G. carpenteri* as MTX

producers, while only *G. polynesiensis* had a confirmed CTX production profile ([Table 4.1](#)). This study revealed a set of core protein coding sequences shared by all taxa as well as a subset of species-specific, unique portions of the transcriptome. This can be utilized as a tool for future *de novo* sequencing efforts, and form a basis for comparison of expression between toxic and non-toxic species.

4.3 Methods

Scripts used for this project are available on Github under [hydrahamster/pan-tran](#). Venn diagrams were created with InteractiVenn [88]. Bar graphs were generated in Origin Pro (OriginLab, Northampton, MA).

4.3.1 Transcriptome acquisition

Species of *Gambierdiscus* used in this chapter are summarized in [Table 4.1](#). Unless otherwise noted, toxicity and toxin profile reports are specific to the strains used as inter-species variation in toxin production has been recently reported [134, 220]. The *G. polynesiensis* toxin profile was elucidated by Tim Harwood at the Cawthron Institute (Nelson, NZ) with the same methodology as for *G. lapillus* used in [Section 2](#). RNA-seq libraries were assembled as per the transcriptome assembly as per [Section 5.4.2](#) and [Section 5.4.2.1](#) without diginorm. A publicly available *G. caribaeus* transcriptome [206] was not included in this study, because the quality assessment through BUSCOv2 showed a high level of fragmentation. It was suspected that the inclusion of this transcriptome would have a detrimental effect on this study.

TABLE 4.1 *Gambierdiscus* species transcriptomes used in this study along with their toxicity, toxin profile, accession numbers and source. Where possible, information is strain specific & otherwise denoted with *

Species	<i>G. australes</i>	<i>G. carpenteri</i>	<i>G. lapillus</i>	<i>G. polyne-siensis</i>	<i>G. holmesii</i>
Strain	CAWD149	UTSMER9A	HG4	CG15	HG5
Transcriptome source	MMETSP	Section 5	Section 5	Section 5	Section 5
Accession ID	MMETSP0766	SRR6821720	SRR6821722	SRR6821723	SRR6821721
Isolation location and year	Rarotonga, Cook Islands (2007)	Merimbula, Australia (2014)	Heron Island, Australia (2014)	Rarotonga, Cook Islands (2014)	Heron Island, Australia (2014)
Toxin profile (LC-MS/MS)	CTX -ve; MTX +ve	CTX -ve; MTX -ve	CTX -ve; MTX +ve	CTX +ve; MTX +ve	CTX -ve; MTX +ve
Toxicity via bioassay	CTX +ve; MTX N/A	CTX -ve; MTX +ve	CTX +ve*; MTX +ve*	CTX +ve*; MTX +ve*	CTX +ve*; MTX +ve*
References	[114, 178, 219]	[134]	[128, 134]	[29]	[128, 134]

4.3.2 Spliced leader search

The dinoflagellate spliced leader (dinoSL) sequences reported by [293] were used to build a HMMER library. The transcriptome assemblies were searched with the dinoSL hmmer library to investigate for spliced leader presence. All clusters were searched for membership of one or more contigs with a dinoSL.

4.3.3 Homolog clustering

Cd-hit [66] was used to cluster highly similar transcripts to reduce redundancy with the flags -T 10 -M 5000 -G 0 -c 1.00 -aS 1.00 -aL 0.005 as shown by a study with similar aims by [27]. Transdecoder was used to predict coding regions on the clustered nucleotide sequences [78]. Protein clusters were annotated with Interproscan v5.27 with local lookup server [209]. Protein clusters were processed to include the species of origin instead of the TRINITY tag and concatenated for input to get_homologues [272]. The -t 0 flag was used for get_homologues to acquire all possible clusters even with only one species representative, and -G for the OMCL algorithm. The core transcriptome was defined as clusters with members from all five species. Clusters containing four of the five species were considered as the softcore. The unique transcriptome consisted of clusters with only a single species representative.

4.3.4 Comparison of gene ontology annotations

The transcriptomes as well as the pan-transcriptome clusters were matched with their Interpro annotations and Gene Ontology (GO) terms were queried with GOSUM against the basic GO database with the go-basic GO reference [6, 35, 111]. The GO annotations are refined into three principal components: 1) molecular function which is defined as a biochemical process or a macromolecule directly interacting with other molecules, 2) Cellular components which present a location within the cell where a molecular function takes place, and 3) Biological process which is defined as a molecular machinery participating in the execution of the cell's genetic programming (e.g. cell division). GO basic is structured in a hierarchical manner, with parent terms (e.g. level 1) provide a broad overview of functionality, while child terms (e.g. level 2 in relation to level 1) are more refined subsets of their parent terms. For a general overview of the functions present in each transcriptome, level 1 GO terms were queried. A more in depth query of the functions present in each transcriptome was conducted with a GO search of the child terms at level 2.

4.3.5 Ketosynthase domain search

The transcriptome assemblies were queried for the ketosynthase (KS) active domain of the polyketide synthase (PKS) enzyme using hmmer [52] with libraries developed for this project. The contigs which were identified to contain an active domain were then searched for within the protein coding clusters to identify how the active domains clustered; and the assemblies were searched to compare KS abundance between species.

4.4 Results

4.4.1 Overview of the transcriptomes

The progression of clustering and annotation results per transcriptome can be found in [Table 4.2](#). A total of 287,546 clusters were found across all five species ([Figure 4.1](#)).

TABLE 4.2 Progression of clusters found in each *Gambierdiscus* transcriptome during processing.

Species	<i>G. aus- trales</i>	<i>G. carpenteri</i>	<i>G. lapillus</i>	<i>G. polyne- siensis</i>	<i>G. holmesii</i>
Contigs	102,863	263,829	148,972	270,315	191,224
Spliced leader contigs	304	683	232	1,570	1,524
Nucleotide clus- ters (cd-hit)	102,861	263,743	148,966	270,265	191,205
Predicted cod- ing regions (Transdecoder)	63,299	180,568	111,862	176,290	132,688
Contigs anno- tated (Interpro Scan)	131,970	334,737	225,324	225,324	254,844
Core- transcriptome clusters	13,750	13,750	13,750	13,750	13,750
Softcore- transcriptome clusters	2,372	16,058	16,297	16,557	16,636
Unique clusters	35,356	61,494	32,341	60,769	41,350

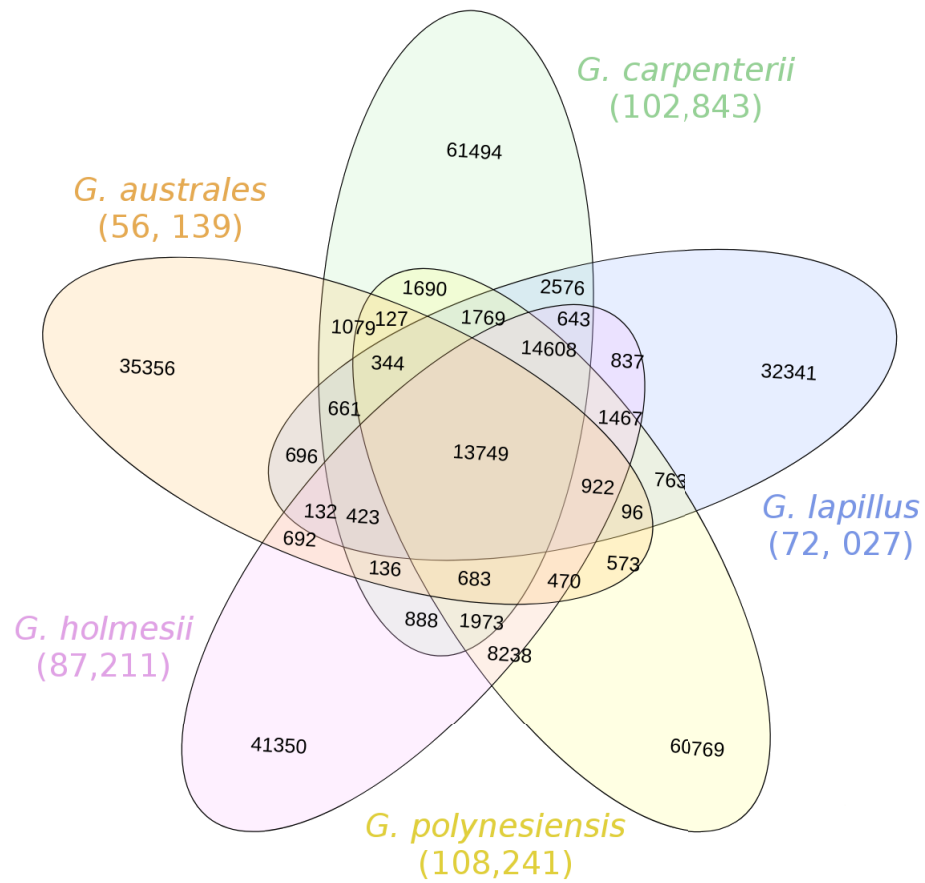


FIGURE 4.1 Venn diagram of nucleotide clusters distribution between the five *Gambierdiscus* species.

4.4.1.1 Level 1 GO annotations between *Gambierdiscus* species

With the exception of *G. australes* being consistently underrepresented, the number of GO annotations found between the species were similar. In the molecular functions category, most annotated protein coding sequences matched to catalytic activity, followed by binding and transporter activity (Figure 4.2). Molecular carrier activity was only registered for *G. australes* and *G. carpenteri* with 1 annotation each.

Within the cellular components category, the most protein sequence annotations matched to cell parts, followed by protein-containing complexes and organelle parts (Figure 4.3). In the cell junction category, *G. carpenteri* and *G. polynesiensis* had one protein coding sequence each.

For biological process based annotations, the most protein coding sequences for each species fell within cellular processes, followed by metabolic processes (Figure 4.4). Biological processes and localization both had the third most representative protein coding sequences, depending on which species was examined. The least represented category was for growth, with only one annotation for *G. holmesii* and *G. polynesiensis*.

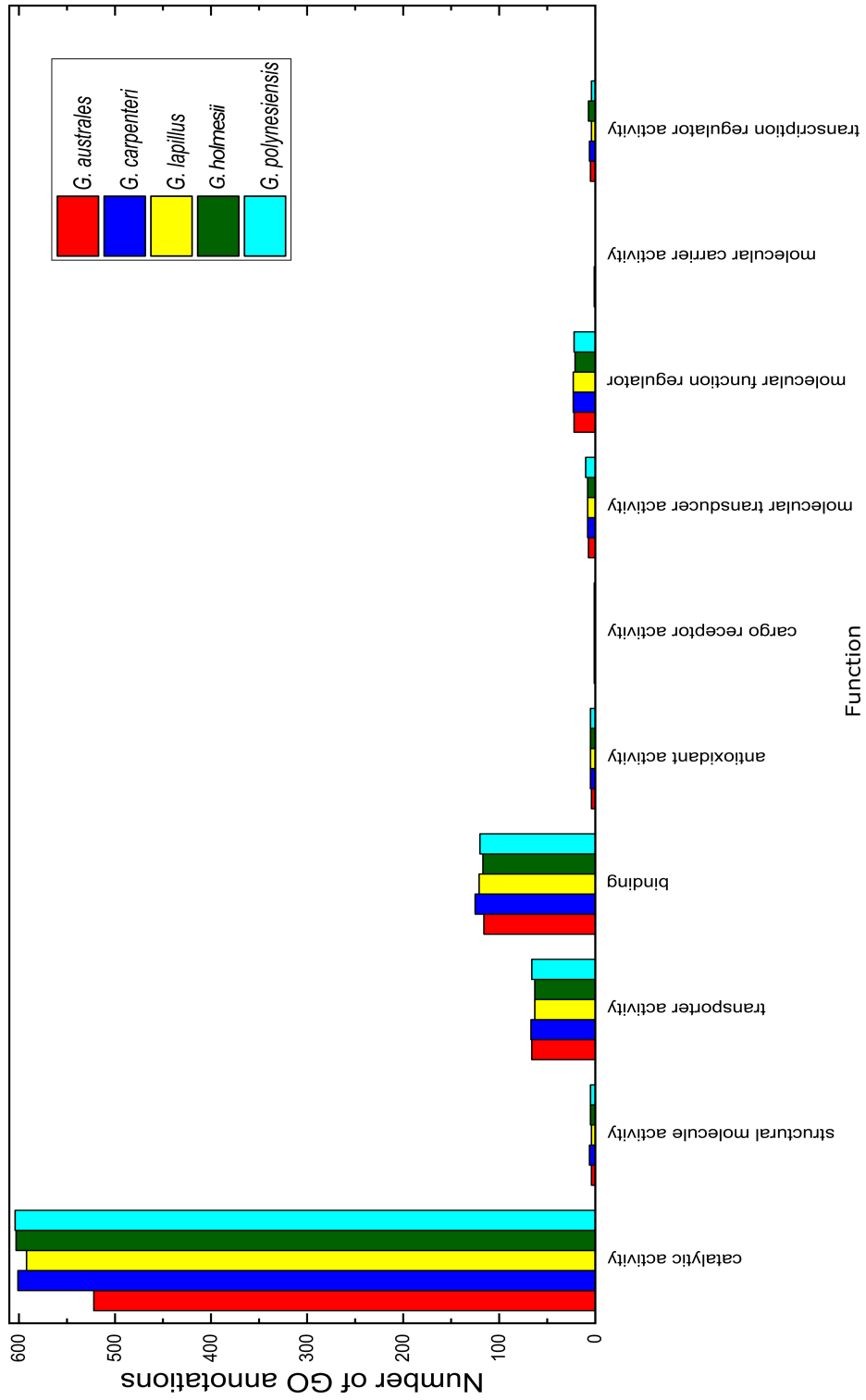


FIGURE 4.2 Summary of GO annotations between *Gambierdiscus* species that mapped to molecular functions at GOSUM level 1.

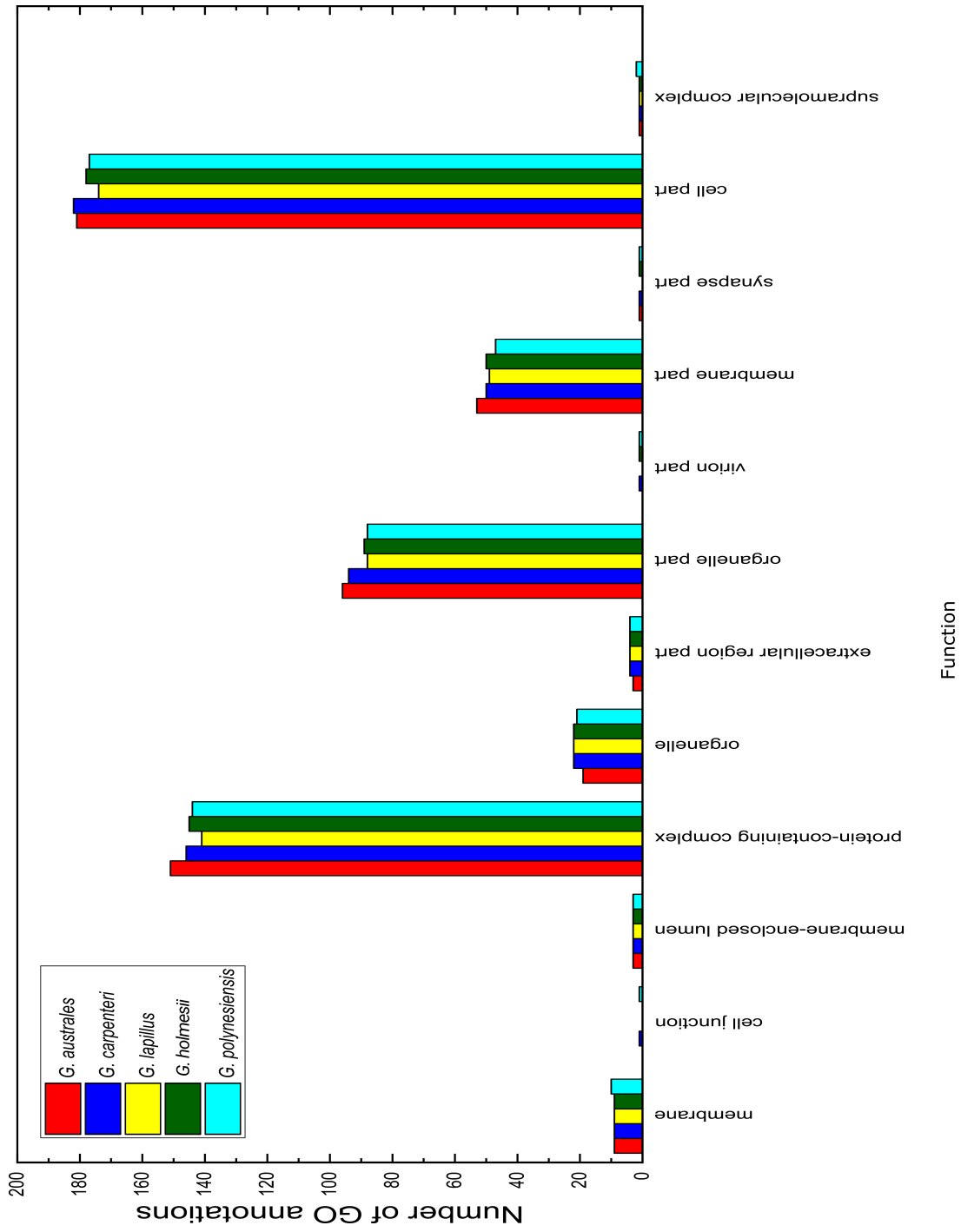


FIGURE 4.3 Summary of GO annotations for cellular components between *Gambierdiscus* species at GOSUM level 1.

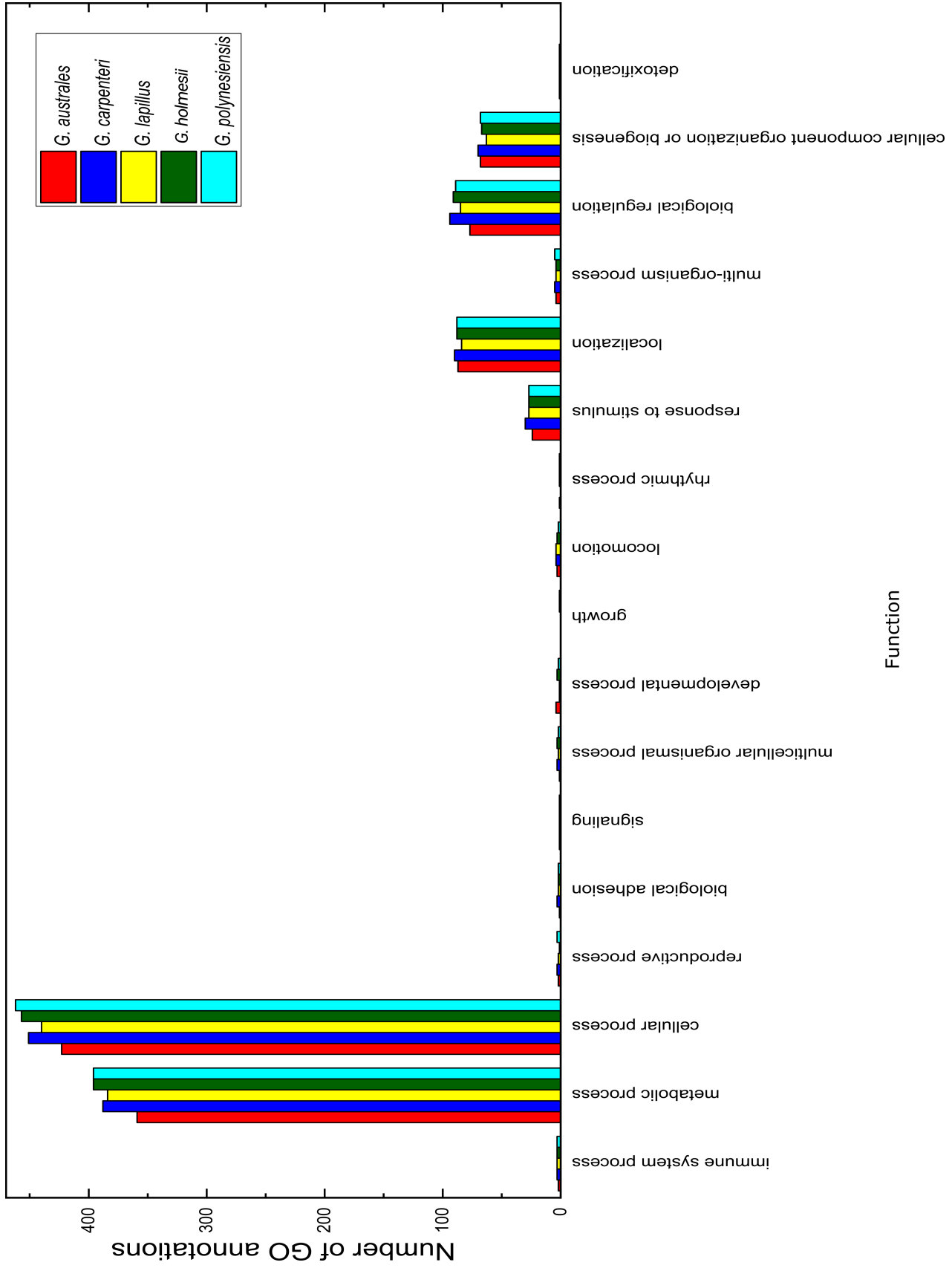


FIGURE 4.4 Summary of GO annotations in the biological processes category, between *Gambierdiscus* species at GOSUM level 1.

4.4.1.2 Level 2 GO annotations between *Gambierdiscus* species

While at level 2 of GO terms the inter-species variations in protein coding sequences across the molecular functions (Figure 4.7), cellular components (Figure 4.6) and biological processes (Figure 4.5) become apparent, *G. australes* is consistently underrepresented or absent across all three processes. Conversely, within the molecular function *G. australes* was the only species with protein coding sequences for nucleocytoplasmic carrier activity and general transcription initiation factor activity (Figure 4.7); as well as for anatomical structure morphogenesis and movement within environment as part of symbiotic interaction in the biological processes section (Figure 4.5). *G. holmesii* had a much higher representation of GO terms matching sperm-egg recognition. The most common molecular function (Figure 4.7) mapped to transferase activities, followed by hydrolase activity and oxidoreductase activity. For cellular components (Figure 4.6) the highest number of GO term annotations matched to intracellular parts, then intracellular organelle parts and membrane protein complexes. Within the biological processes category, organic substance metabolic processes, cellular metabolic processes and primary metabolic processes contained the most protein coding sequence matches (Figure 4.5).

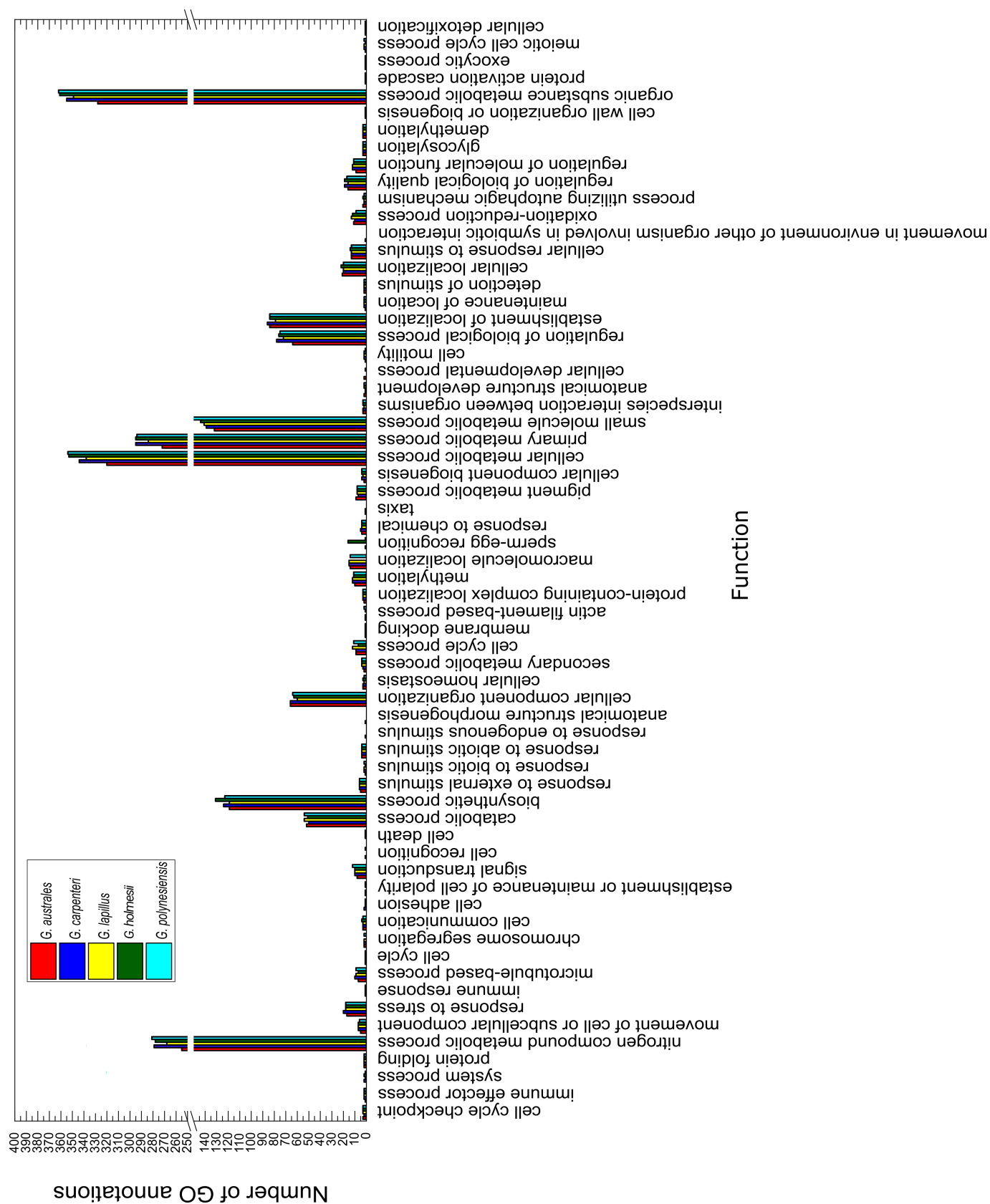


FIGURE 4.5 Summary of GO annotations for biological processes between *Gambierdiscus* species at GOSUM level 2.

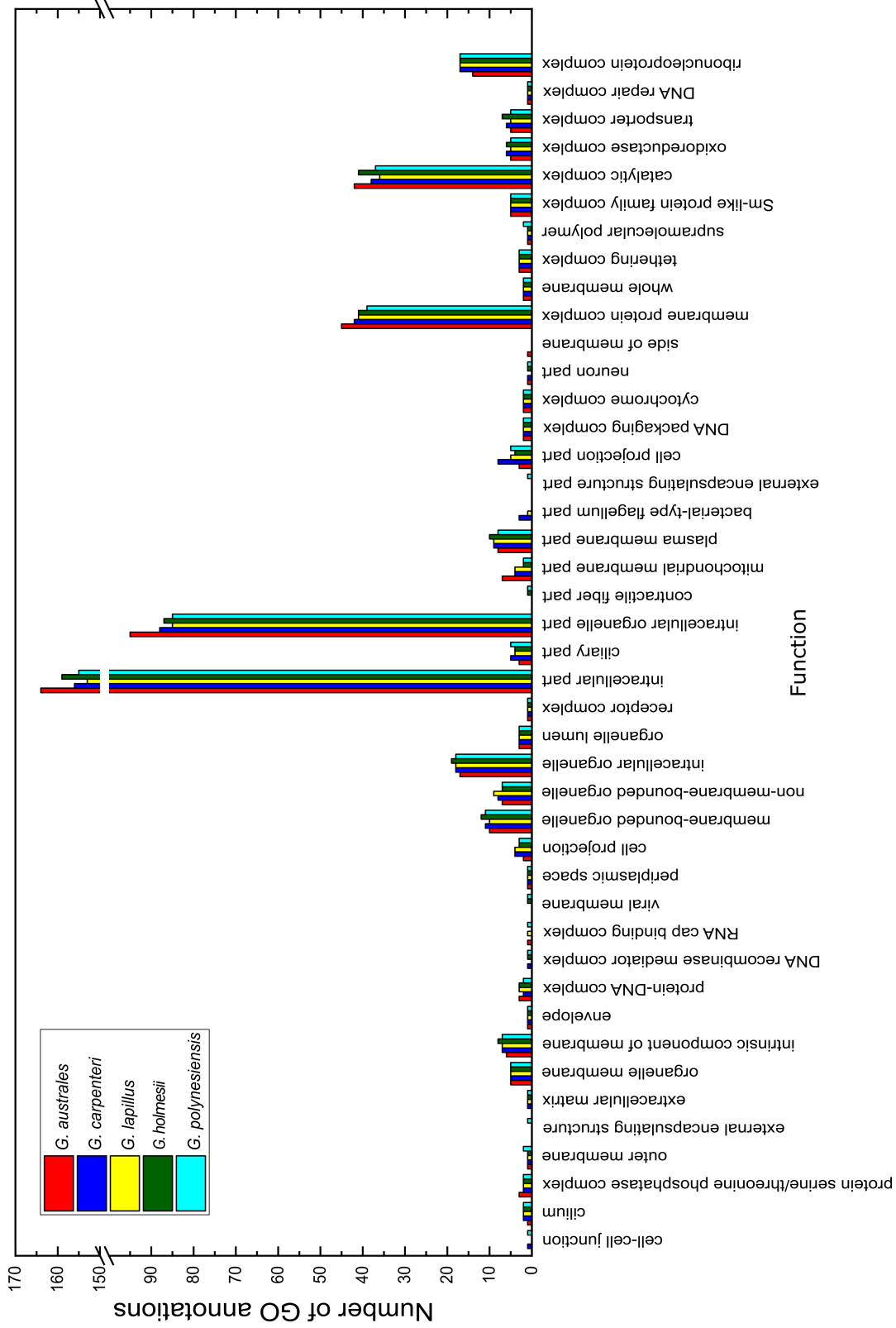


FIGURE 4.6 Summary of GO annotations matching to cellular components between *Gambierdiscus* species at GOSUM level 2.

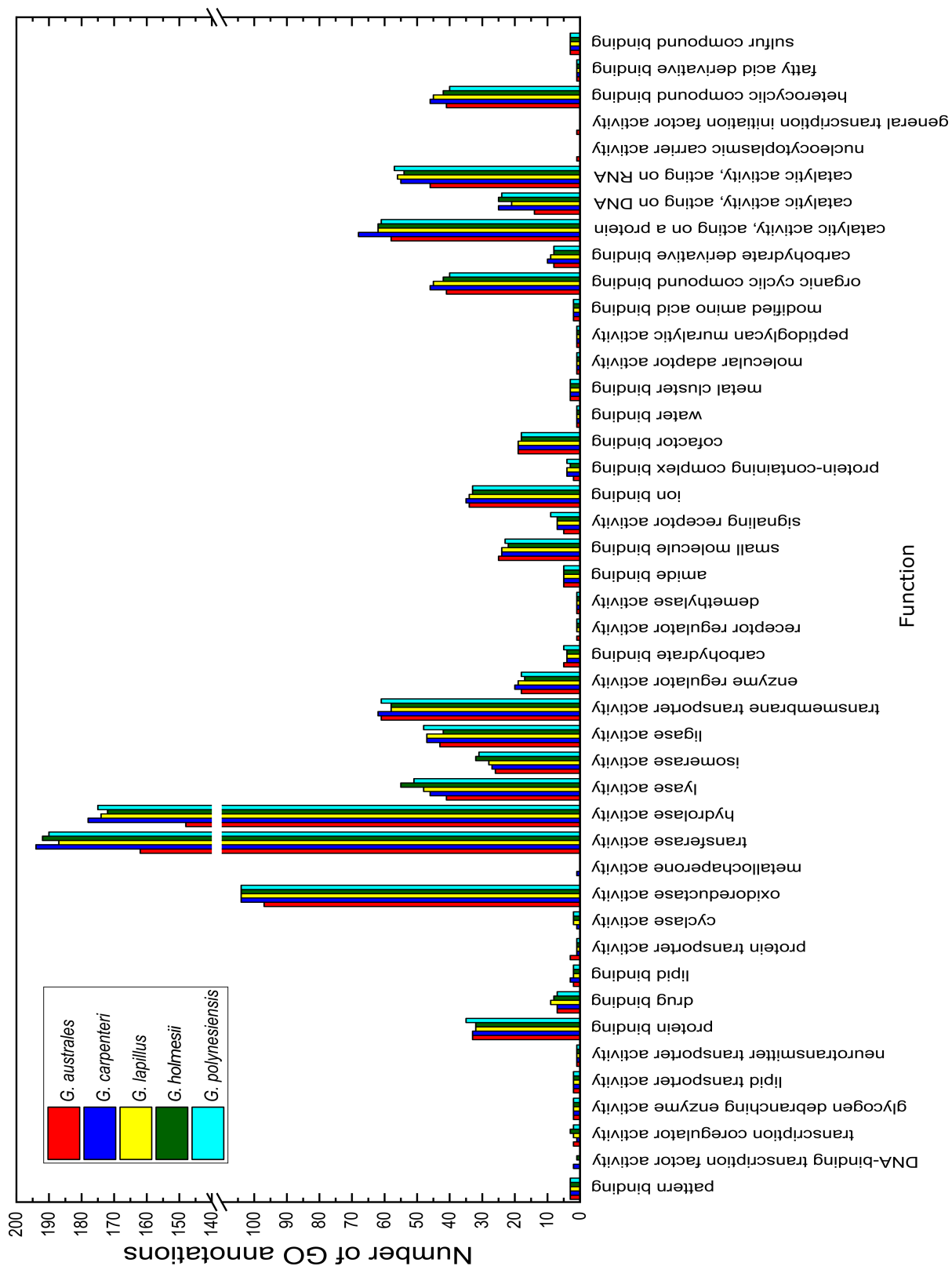


FIGURE 4.7 Summary of GO annotations for molecular functions between *Gambiardiscus* species at GOSUM level 2.

4.4.2 Pan-transcriptomics of *Gambierdiscus*

4.4.2.1 Core transcriptome

A set of core protein coding regions common to all five species of *Gambierdiscus* were found. This set consisted of 13,750 amino acid clusters (Table 4.2) of which 45 % were annotated with GO terms. The largest core cluster contained 180 protein coding sequences with 23, 45, 32, 31 and 49 from *G. australes*, *G. carpenteri*, *G. lapillus*, *G. polynesiensis* and *G. holmesii* respectively, and could not be annotated. Twelve of the core clusters contained 100 or more protein coding sequences, of which another two could not be matched to any known annotation. The predicted protein coding regions for the other nine clusters, in descending order of sequence numbers contained, are: an enzyme with catalytic activity involved in metabolic process, a calcium binding transmembrane transport channel, a protein involved in calcium binding, a protein binding enzyme, a domain for unspecified protein binding, an enzyme with O-glucosyl hydrolase activity involved in carbohydrate metabolic process, membrane bound ion transporter with cation channel activity & ionotropic glutamate receptor activity, a transmembrane transporter with voltage-gated calcium channel activity, and calcium ion binding transmembrane ion transporter. A total of 3,943 core clusters contained 10 or more protein coding sequences, i.e. 71.32 % of the total core clusters contained less than 10 protein coding sequences. The majority of clusters fell within metabolic processes, cellular components and catalytic activity.

4.4.2.2 Softcore transcriptome

A softcore, with clusters containing protein coding sequences from four of the five species queried, was identified. The softcore consisted of an additional 16,980 clusters to the core (Table 4.2), of which 48 % were annotated. The most prolific cluster in the softcore contained 163 protein coding sequences, where *G. carpenteri*, *G. lapillus*, *G. polynesiensis* and *G. holmesii* represented with 50, 42, 41 & 30 sequences respectively. This cluster could not be annotated. A further 5 clusters contained more than 100 sequences, four of which could be annotated. Of the six clusters with over 100 entries, none had representative sequences from *G. australes*, and this species was absent from 86 % of the softcore clusters. The annotations of these clusters matched to: a protein involved in selective protein binding, a protein involved in actin binding, a protein involved in calcium binding, and a protein with cysteine-type peptidase activity. In the softcore, 14,035 clusters contained 10 or more protein coding sequences.

4.4.2.3 The unique portion of the transcriptome

Clusters with only single species representatives, i.e. unique clusters to the five *Gambierdiscus* species examined, was the largest portion of the pan-transcriptome with 231,310 clusters. Only 15.23 % of these clusters could be annotated. The unique transcriptome from *G. australes*, *G.*

carpenteri *G. lapillus*, *G. polynesiensis* and *G. holmesii* contained 35,356, 62,494, 32,341, 60,796 & 41,350 clusters respectively (Table 4.2). The highest number of protein coding sequences in a unique cluster was 37, found in two clusters from *G. carpenteri*. One of these was annotated for RNA and metal ion binding activity. Of the unique clusters, 83.1 % contained only one sequence and 97.8 % of clusters contained 5 protein coding sequences or less.

4.4.3 Level 1 GO annotations for pan-transcriptomes

The number of annotated clusters across the biological processes, cellular components and molecular function groups was very similar between the core, softcore and unique clusters, with the unique clusters consistently highest across all three groups.

Clusters falling within the cellular components category predominantly matched to cellular parts, followed by protein-containing complexes and organelle parts (Figure 4.8). For biological processes, the prevalent GO annotations matched to cellular components, metabolic processes and localization (Figure 4.9). GO annotations for molecular functions most commonly matched to catalytic activity then binding and transporter activities (Figure 4.10).

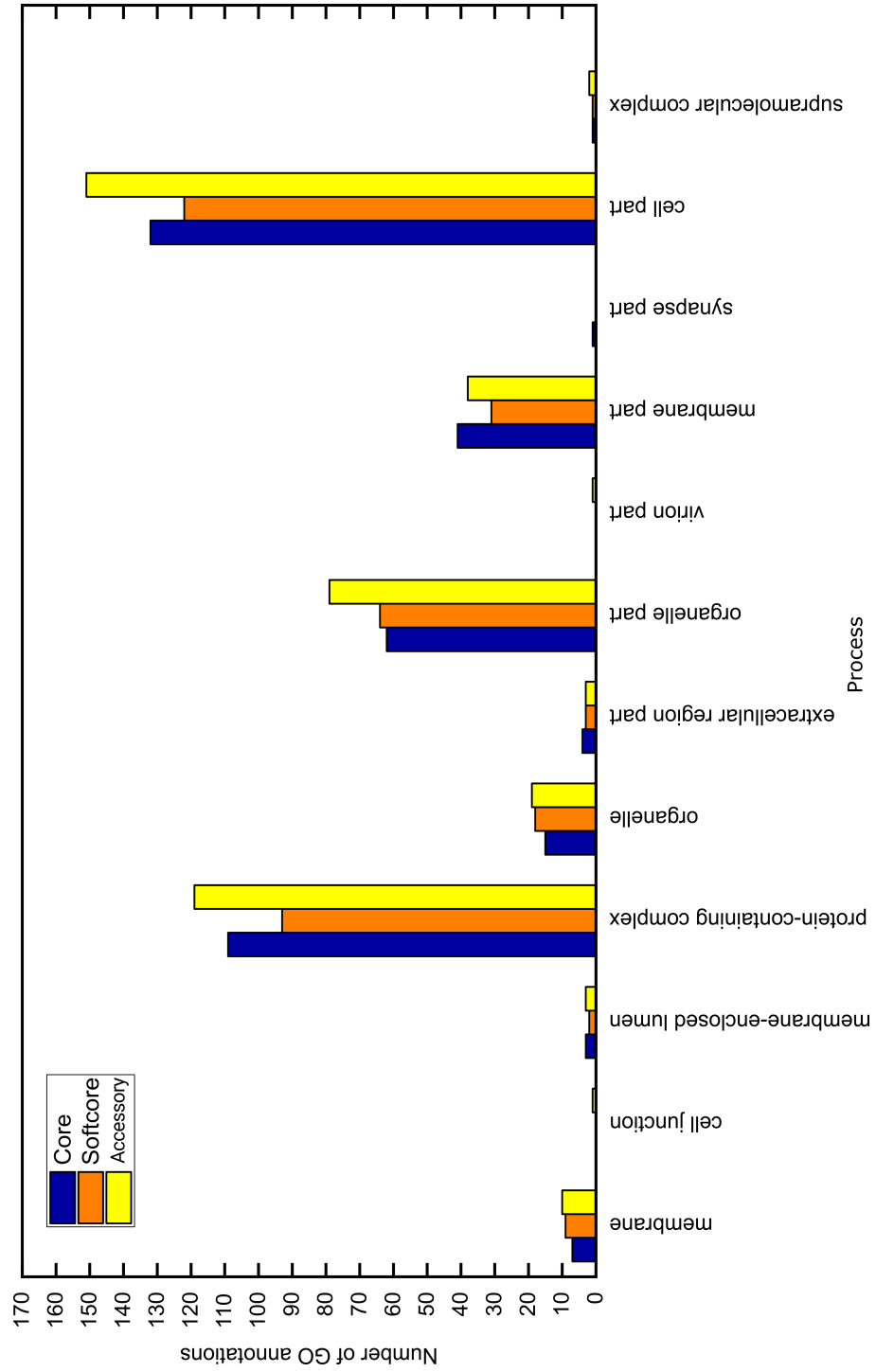


FIGURE 4.8 Summary of pan-transcriptome clusters matching to cellular component GO annotations at GOSUM level 1.

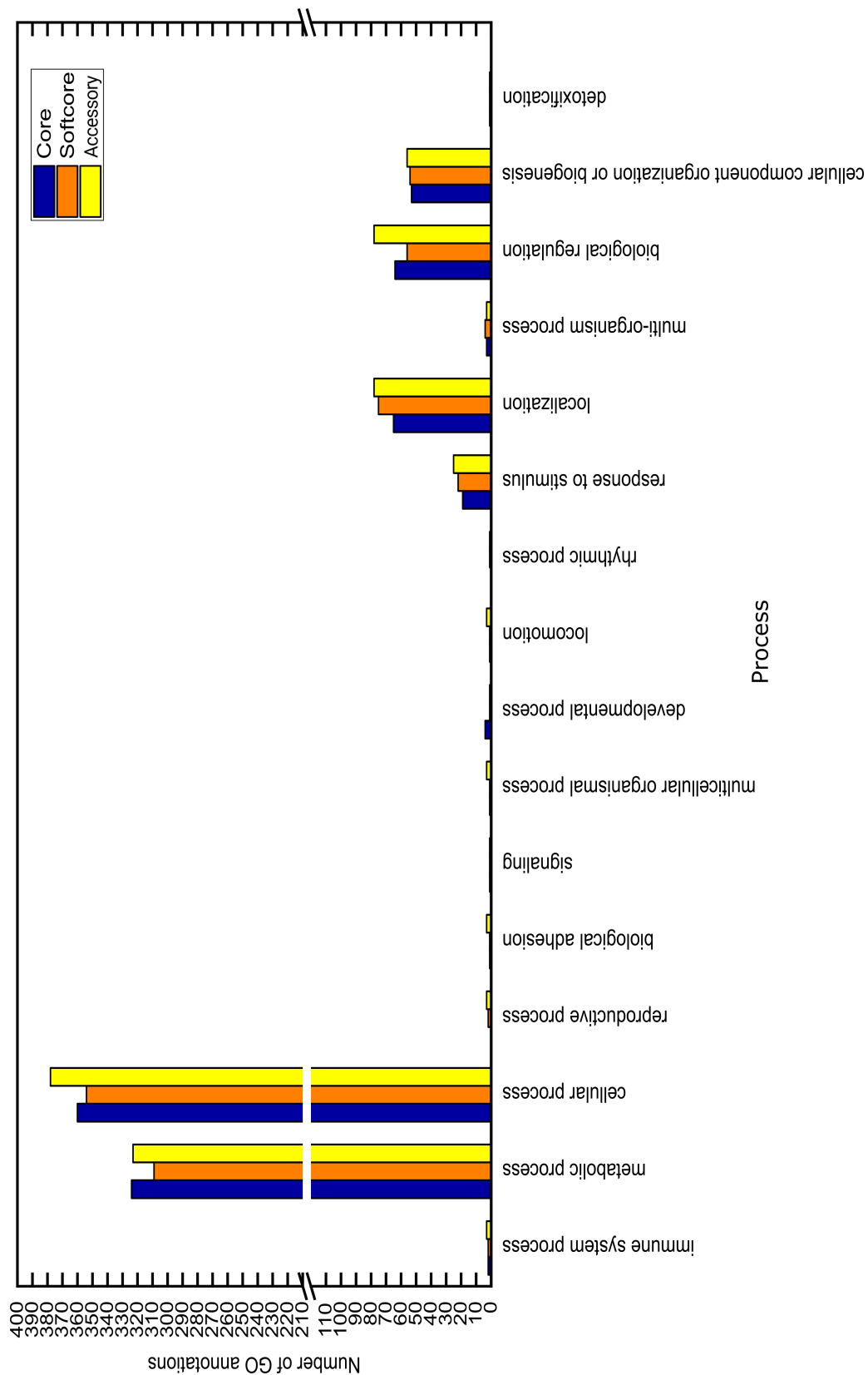


FIGURE 4.9 Summary of biological processes GO annotations of pan-transcriptome at GOSUM level 1.

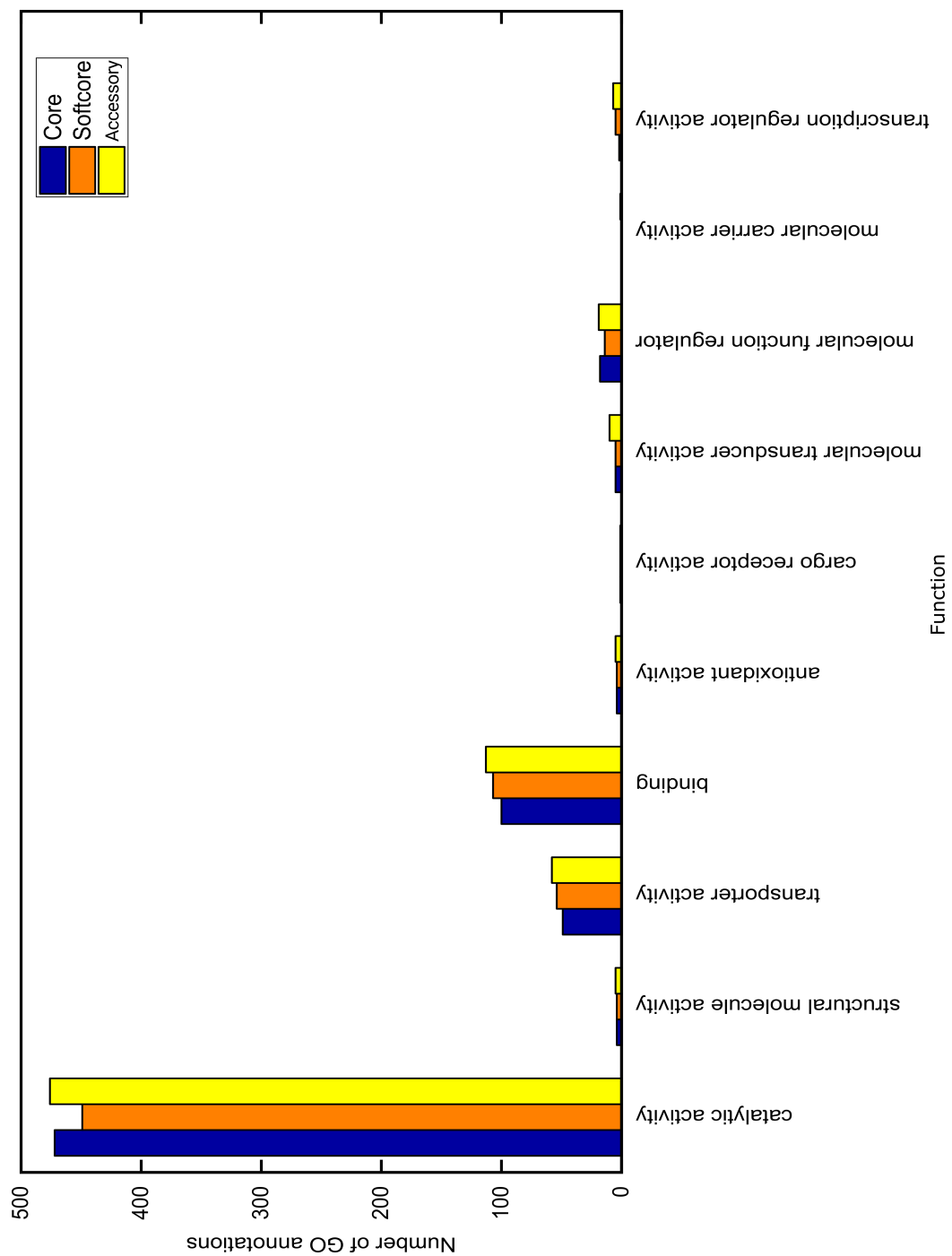


FIGURE 4.10 Summary of level 1 molecular function GO annotations of pan-transcriptome.

4.4.4 Level 2 GO annotations for pan-transcriptomes

Differences between the annotations of the different pan-transcriptome clusters for the biological processes, cellular components and molecular function were more distinctive at level 2. In most cases, the core and unique transcriptomes were evenly matched in number of clusters represented across the three categories, while the soft-core was less prevalent. Within the cellular components, most annotations matched to intracellular parts, followed by intracellular organelle parts then membrane protein complexes ([Figure 4.11](#)).

Within the molecular function, the most common GO annotations for the accessory clusters matched to transferase, hydrolase and oxidoreductase activities while only the unique clusters had annotations matching to DNA binding transcription factor activity, metallochaperone activity and water binding ([Figure 4.12](#)).

The biological processes annotations most commonly matched to organic substance metabolic processes, cellular metabolic processes and primary metabolic processes, in descending order ([Figure 4.13](#)). Unique clusters were the only representatives for system processes, immune response, cell adhesion, cell death, sperm-egg recognition, cell motility and a protein activation cascade.

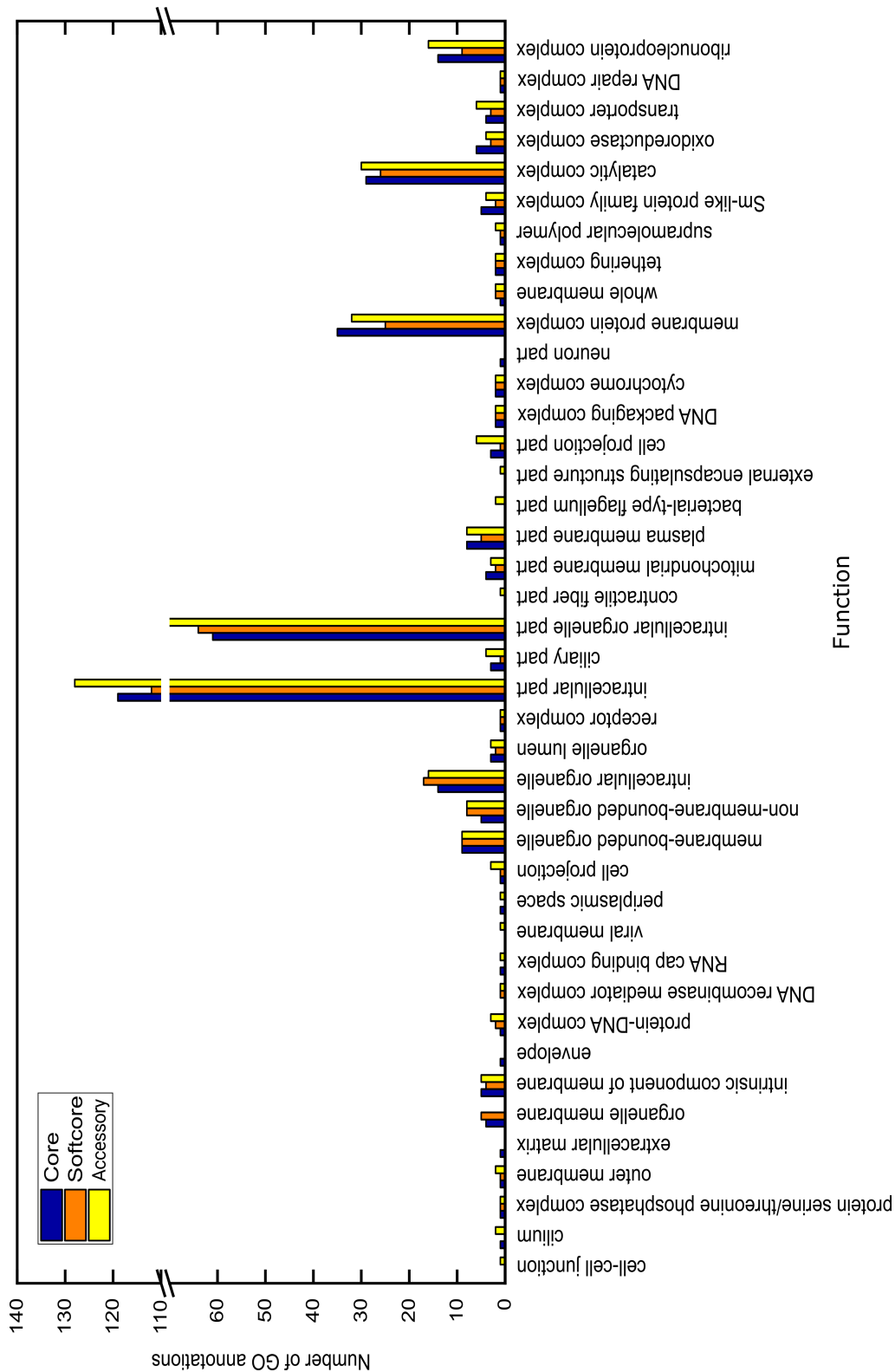


FIGURE 4.11 Summary of cellular component category GO annotations for the pan-transcriptome at level 2.

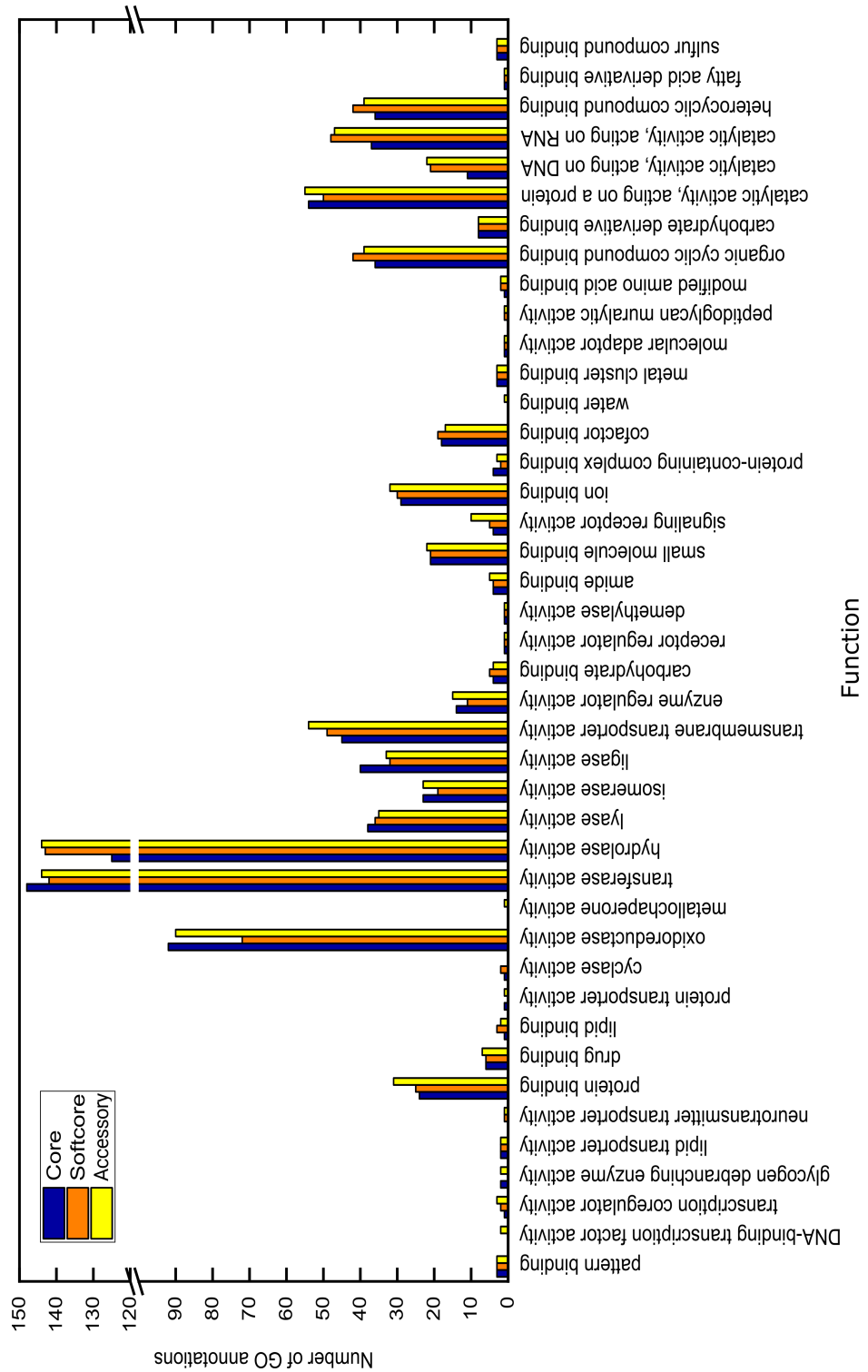


FIGURE 4.12 Summary of level 2 molecular function GO annotations for pan-transcriptome clusters.

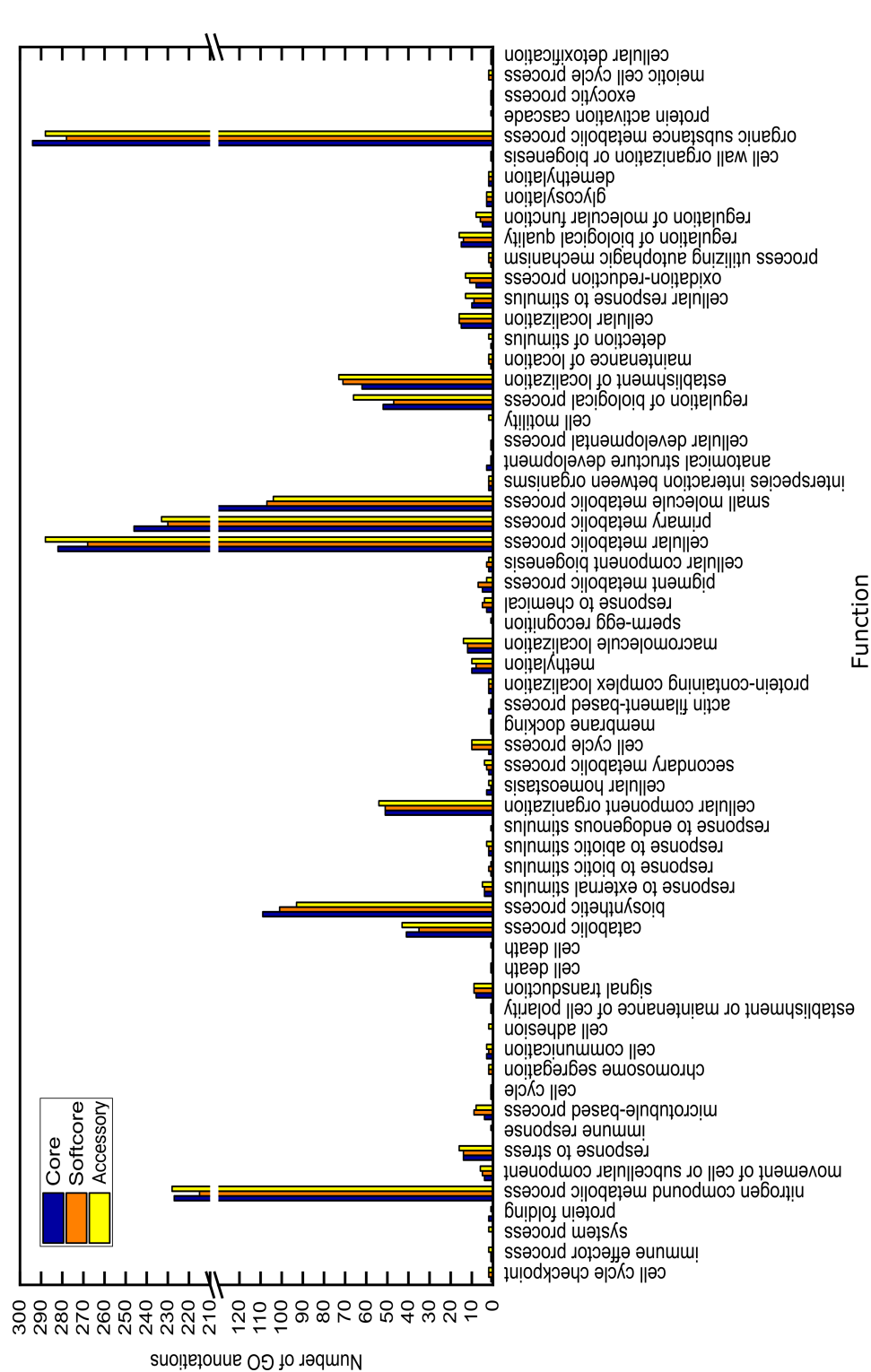


FIGURE 4.13 Summary of biological processes annotations for the pan-transcriptome clusters at level 2.

4.4.5 Ketosynthase active domain search

A total of 850 contigs were identified with KS domains which grouped into 314 clusters ([Figure 4.14](#)). Nine clusters contained more than 10 protein coding sequences, the most prolific of which contained 130 sequences from all *Gambierdiscus* species. Only two of these nine clusters did not contain all the species examined.

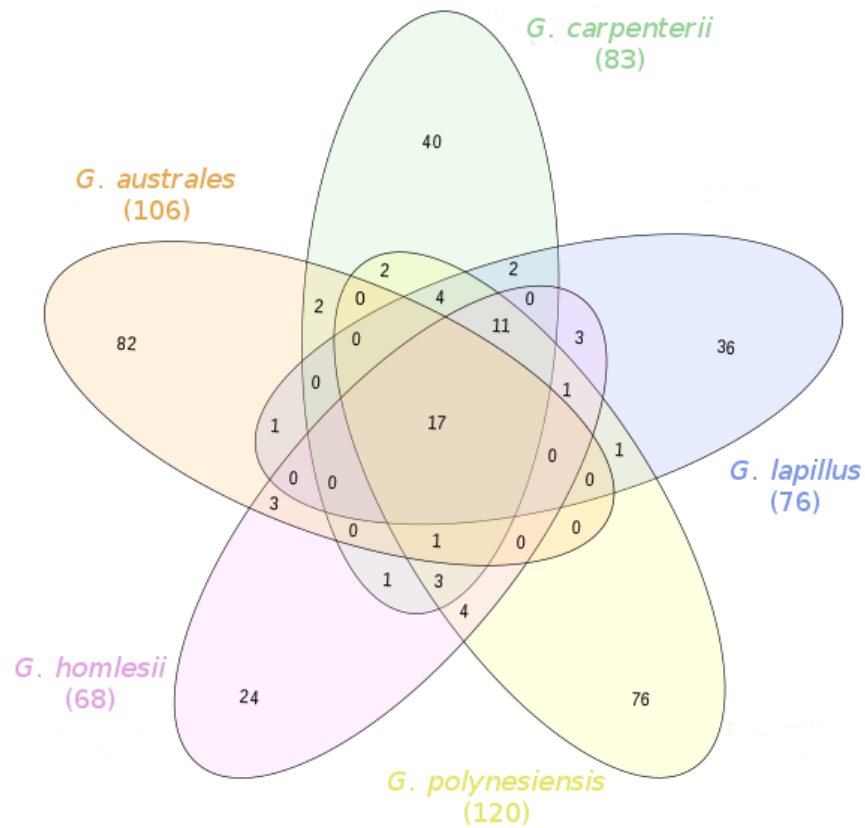


FIGURE 4.14 Venn diagram of ketosynthase containing clusters distributed across the five *Gambierdiscus* species. Core, accessory and inter-species ketosynthase cluster cross over depicted.

Fifty-seven of the 314 clusters contained contigs from multiple species, so 81.8 % of KS clusters were species specific while 78.7 % contained only a single KS coding protein sequence (Figure 4.14). The non-ciguatoxic *G. carpenteri* was absent from 73.6 % of the clusters. Of the clusters with *G. carpenteri* absent, there were no clusters which contained representatives from all of the other four species. However, one cluster contained *G. lapillus*, *G. polynesiensis* and *G. holmesii* with equally represented protein coding sequence numbers. Four contigs contained *G. polynesiensis* and *G. holmesii* only, one of which had a higher number of protein coding sequences from *G. polynesiensis* than *G. holmesii*. *G. polynesiensis* was the only representative species in 71 clusters, of which three clusters contained two and one cluster contained three protein coding sequences. *G. holmesii* was representative as the only species in 23 clusters, one of which contained three and the others contained a single protein coding sequence(s). *G. australes*, *G. carpenteri* and *G. lapillus* were the solo representatives of 81, 39 & 35 KS clusters respectively.

4.5 Discussion

Comparing five *Gambierdiscus* species revealed a core, soft-core and unique fraction in the transcriptomes. Further, differences between species with different toxin production characteristics were observed. The number of predicted peptides found in this study is high compared to the pan-transcriptome study of four prymnesiophyte algae conducted by [124]. The predicted peptides in the [124] study ranged from around 25,000 to 56,000 peptides, where this study found a range from about 63,000 to 270,000 (Table 4.2). The lowest number of peptides was predicted in *G. australes*, similar to the findings of [124] and both originating from the MMETSP [114, 124]. *G. holmesii* and *G. lapillus* protein coding sequences numbered 132,688 and 111,862 respectively, while the highest numbers were found in *G. carpenteri* and *G. polynesiensis* at 180,568 and 176,290 respectively (see Table 5.1). The abundance of dinoSL was quite low (Table 4.2) compared to the abundance observed by [294] in *Amphidinium carterae* (for relevance of spliced leader sequences in dinoflagellates see Section 1.3). Similar to this study, a low abundance of spliced leaders has been observed in other dinoflagellates [9, 77]. The function of the spliced leader, and hence whether this variation in observation between high and low abundance is species specific or due to assembly method employed, is as yet unknown. Interestingly, *G. holmesii* and *G. polynesiensis* had the most contigs with dinoSLs. These were sequenced with different read lengths and collected from different geographic locations, but are phylogenetically in the same *Gambierdiscus* sub-clade (for phylogenetic placement of *G. holmesii* and *G. polynesiensis* see Section 8.2.2).

4.5.1 Core, softcore and unique clusters

The number of contigs and protein coding sequences from this study was markedly less for *G. australes*, from the MMETSP dataset, in comparison to the transcriptome assemblies generated in

Section 5. To accommodate for the low number of contigs from *G. australes*, the softcore spanned 4 of the 5 taxa. Noticeably, *G. australes* was absent from 86 % of the softcore dataset which indicates that a large proportion of the softcore is likely part of the *Gambierdiscus* pan-transcriptome core which was not captured in the *G. australes* sequencing. This is an example highlighting the value of a reference pan-transcriptome.

There was no distinct difference between the species GO annotations, again with the exception of *G. australes*. This is as expected, as the species operate under similar nutritional modes and within similar temperature ranges [115, 196] and were isolated from tropical or sub-tropical conditions, apart from *G. carpenteri* which was isolated from the temperate Merimbula region [134].

No difference was observed in the biological processes, cellular components and molecular function GO annotation groups at levels 1 and 2 between the core, softcore and accessory clusters. This is somewhat less expected as it would be reasonable to predict a functional difference for genes unique to each species. Possible reasons could be that this observation only captures protein coding sequences with GO annotations which were only 15.23 % of the 231,310 accessory clusters. This indicates that no functional match for over 196,000 of the unique clusters could be found and no indication what their function could not be extrapolated. As dinoflagellate sequencing and genome exploration is still comparably sparse, it follows that a lack of annotated references in the sequencing databases translates to low annotation rates. The need for annotations and a comparable reference transcriptome becomes apparent when investigating genetic elements as uncharacterized as these.

4.5.2 Ketosynthase domain detection

Between all five species, 850 contigs with KS domains were identified. These resolved into 314 clusters, of which 17 were shared among all species. The majority (81.8 %) resolved as unique clusters per species, ranging from 24 clusters for *G. holmesii* to 82 clusters for *G. australes*. The toxin profile for *G. polynesiensis* and the toxicity assay conducted by [134] for *G. holmesii* indicate that these two species are the most ciguatoxic of the taxa included in this study. While *G. lapillus* displayed lower levels of ciguatoxicity in bioassays, the strain included here (HG4) was the least toxic of the *G. lapillus* strains tested in [Section 2.5.3](#). Four KS clusters included sequences from *G. holmesii* and *G. polynesiensis* only, one of which contained four contigs from *G. polynesiensis* and one from *G. holmesii*. A further 22 and 75 clusters only contained *G. holmesii* and *G. polynesiensis* sequences respectively. These clusters could be avenues of pursuit for future investigation into KS domains involved in CTX synthesis.

4.5.3 Areas for possible improvement in this study

This chapter presents a novel approach to analyzing *Gambierdiscus* transcriptomes and possible avenues for investigating commonalities and differences between toxic and non-toxic species.

However, there are several aspects that can be improved upon with future studies.

4.5.3.1 Coverage of taxa

As with bacterial pan-transcriptomics, eukaryotic studies reveal a correlation between the number of species and strains included for refining the core transcriptome and expanding the unique fraction of the transcriptome [109, 124, 260]. For *Gambierdiscus*, including several strains of a species is essential for the discovery of a non-ciguatoxic *G. polynesiensis* strain [220]. Variability in morphology and toxicity has also been observed in *G. lapillus* [128]. While this study sought to cover taxa from the three main *Gambierdiscus* clades, an increase in species and strain coverage is highly likely to impact the resolution of the core and unique portions especially as the sequencing coverage of the *G. australes* transcriptome is of lower quality than the other four species (Table 4.2). Thus, due to limited species and the singular strain per species coverage, this represents an exploratory study for establishing a pan-transcriptome for *Gambierdiscus* which should be improved upon.

4.5.3.2 PKS active domain search

PKS complexes consist of a number of active domains that synthesize and manipulate the polyketide backbone. The KS domain is but one of several essential domains for a functional polyketide. The search for active domains should be extended to include other domains for comparison in the search for the ciguatoxin production pathways.

4.6 Conclusion

Protists represent a large section of the tree of life and are involved in vital geochemical cycling, symbiotic and toxic relationships in their environment. Yet due to the convoluted nature of their genomes, querying the genetic content of these organisms is fraught with obstacles. This study presents a pan-transcriptome for the genus *Gambierdiscus*, some species of which are involved in CTX production. The approach opens up an alternative avenue for investigation of the differences and similarities of toxic *Gambierdiscus* species in general, and specifically in regard to the toxin production pathway(s). This study provides a starting point for *Gambierdiscus* pan-transcriptomic exploration and a rudimentary reference for future sequencing efforts. It is recommended that this dataset is expanded to encompass both more *Gambierdiscus* species and strains to crystallize the genus' pan-transcriptome in more detail.

CHAPTER 5

Trial by phylogenetics - Evaluating the Multi-Species Coalescent for evolutionary inference on taxa with high levels of paralogy (Gonyaulacales, Dinophyceae).

5.1 Authorship Declaration

By signing below, I confirm that for the paper titled “Trial by phylogenetics - Evaluating the Multi-Species Coalescent for evolutionary inference on taxa with high levels of paralogy (Gonyaulacales, Dinophyceae).”, that:

Anna Liza Kretzschmar conceived and designed the experiment, extracted RNA from samples, used student funding for sequencing fees, wrote the bioinformatic scripts to process the dataset, wrote the paper, prepared figures and tables, and reviewed drafts of the paper.

Arjun Verma isolated and cultured dinoflagellate strains used for analysis in this study, extracted and sequenced RNA from *Ostreopsis* spp. and *Coolia malayensis*, used student funding for sequencing costs, and reviewed drafts of the paper.

Shauna Murray conceived the idea of using transcriptomics for a gonyaulacales phylogeny, and reviewed drafts of the paper.

Tim Kahlke annotated transcriptomes, and reviewed drafts of the paper.

Mathieu Fourment conducted the marginal likelihood calculations, and reviewed drafts of the paper.

Aaron E. Darling conceived the methodology of the manuscript, taught ALK bash and python scripting, reviewed scripts, provided access to the GVL environment, and reviewed drafts of the paper.

Production Note:

Signature removed prior to publication.

Anna Liza Kretzschmar

Production Note:

Signature removed prior to publication.

Arjun Verma

Production Note:

Signature removed prior to publication.

Shauna Murray

Production Note:

Signature removed prior to publication.

Tim Kahlke

Production Note:

Signature removed prior to publication.

Mathieu Fourment

Production Note:

Signature removed prior to publication.

Aaron E. Darling

5.2 Abstract

From publicly available next-gen sequencing datasets of non-model organisms, such as marine protists, arise opportunities to explore their evolutionary relationships. In this study we explored the effects that dataset and model selection have on the phylogenetic inference of the Gonyaulacales, single celled marine algae of the phylum Dinoflagellata with genomes that show extensive paralogy. We developed a method for identifying and extracting single copy genes from RNA-seq libraries and compared phylogenies inferred from these single copy genes with those inferred from commonly used genetic markers and phylogenetic methods. Comparison of two datasets and three different phylogenetic models showed that exclusive use of ribosomal DNA sequences, maximum likelihood and gene concatenation showed very different results to that obtained with the multi-species coalescent. The multi-species coalescent has recently been recognized as being robust to the inclusion of paralogs, including hidden paralogs present in single copy gene sets (pseudoorthologs). Comparisons of model fit strongly favored the multi-species coalescent for these data, over a concatenated alignment (single tree) model. Our findings suggest that the multi-species coalescent (inferred either via Maximum Likelihood or Bayesian Inference) should be considered for future phylogenetic studies of organisms where accurate selection of orthologs is difficult.

5.3 Introduction

Historically, the availability of genetic data has been the limiting factor in phylogenetic inference of evolutionary relationships. Now, the breadth of publicly available data sets generated by high throughput sequencing techniques allows for an increasingly detailed investigation into the evolutionary relationships between organisms. The quest to untangle an organism's phylogeny is often challenging but can inform a broad range of further studies, for example epidemiology, toxicology and ecological interactions, e.g. [26, 140, 168, 189, 244].

Factors impacting phylogenetic studies range from the computational methods and availability of compute infrastructure, the methods and models applied to the data as well as the accuracy of the initial genetic data set itself. Furthermore, the practitioners themselves need to have a solid understanding of the methods, including their shortcomings.

An example of the breadth of publicly available data is the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP), which provides transcriptome sequences of over 650 marine eukaryotic microbes [114]. The MMETSP project focuses on a group of understudied organisms which are abundant and play vital roles in the marine environment, from geochemical cycling, to predation, to symbiosis [71, 72]. This data set offers an excellent opportunity to explore the evolutionary relationships between these taxa through phylogenetics.

Central to phylogenetic inference is the existence of characters (such as nucleotides) derived from a common ancestor, which is called homology [60]. There are several types of homology,

each differing in how the characters diverged, and determining the mechanisms through which characters have evolved is essential for choosing the correct inference model. Orthology refers to the case where the divergence of two gene copies has followed a speciation event [59]. Paralogs are two gene copies whose divergence is initiated by gene duplication [59]. Xenologs are genes which, having previously diverged from a common ancestor, have since undergone transfer between organisms through a horizontal gene transfer mechanism [40]. The distinction between these cases is usually considered essential in identifying gene candidates that are informative for species evolution inference, as the selection of orthologs ensures the inclusion of a signal that is based on the speciation of the taxa examined, while selection of paralogs confounds that signal by including information that does not pertain to the speciation of the taxa [51]. As gene duplication and subsequent loss commonly occur over the course of evolution and speciation, the identification of genes that have orthologous relationships is more difficult than may seem apparent from the definition [67]. Importantly, identifying single copy genes does not ensure the selection of orthologous gene copies, as the gene candidates could well be from paralogous lineages where a secondary copy has been lost between the taxa. The case where paralogous homology of genes is masked by gene loss, is termed pseudoorthology [126].

Once candidate genes have been identified, there are further issues that can arise and impact the veracity of the phylogenetic inference. Two common, well characterized types of errors are random (sampling error) and systematic errors. The former arises from the data, as individual gene histories may differ to the species tree. With a small number of genes, this error can reduce the confidence (through node support values) of the topology, and in extreme cases can entirely skew the inference away from resolving a good approximation of a species tree. Increasing the number of genes directly reduces the impact this error has on the analysis [87, 199].

Conversely, systematic errors arise due to the misspecification of the model used for the inference, leading to an incorrect species tree topology. In this case, an increase in data set size can exacerbate systematic errors rather than reduce them as would happen with random errors (see Box 1). In the presence of this type of systematic error, the resulting inference can be positively misleading, with high clade support values for the incorrect tree topology, obfuscating the presence of the error [108, 129, 225].

In summary, common problems in carrying out a species tree inference arise from:

- 1) Selection of paralogs (including pseudoorthologs). If genes with different evolutionary histories are selected, and if this violates the phylogenetic model, the inferred tree may not accurately reflect the history of any of the individual genes or that of the species;
- 2) Concatenation of genes. Can be a statistically inconsistent estimator of the species tree due to incomplete lineage sorting (ILS) and concatenation acts as an imperfect estimator of species tree topology [225];
- 3) Inference of model adequacy from bootstrap values. [129] demonstrated high bootstrap support under maximum likelihood (ML) inference for incorrect species trees with concatenated

gene sets as input [129]. As high bootstrap values are often used as an indicator for robust species topology resolution, this fallacy is particularly problematic if the reader/operator is unfamiliar with the statistical phenomenon.

In this study, we explored the application of data analysis techniques which attempted to mitigate several of the pitfalls in species tree inference, beyond what has previously been applied in the study of protist phylogenetics. The sequence data was prepared using a workflow that assembled RNA-seq data sets, identified and extracted single copy genes across input taxa, and aligned selected genes ready for Bayesian inference (BI) phylogenetics. Next, we evaluated the impact of model and data selection on the resulting phylogenetic inference. Finally, we applied the methodology to a group of organisms notorious for their extensive paralogy - the Gonyaulacales (phylum: Dinoflagellata) (see box 2 for further information on the dinoflagellates). We present a phylogenetic inference of the Gonyaulacales generated under the multi-species coalescent (MSC) and compare the topology to inferences with commonly used methodologies.

5.3.1 Summary: Statistical nomenclature & errors this study seeks to address

For in depth explanations see [283].

- **Potential statistical error types:**

- 1) random. Sampling-based error which decreases and approaches zero as the size of the data set approaches infinity.
- 2) systematic. Arises from incorrect model assumptions or problems with the model itself. Error type persists and increases as data set size approaches infinity. If strong, can override true phylogenetic signal.

- **Incomplete lineage sorting (ILS):** discordance of gene evolutionary history with the species evolutionary history causing the phylogenetic species tree to be incorrectly inferred. Difference in the topology of a gene tree compared to the species evolution can arise from the divergence of those orthologs prior to the species divergence, where in effect the ancestral populations contain two or more already diverged copies of the gene across one or more species divergence points. Another mechanism is the introduction of a copy of the gene which is not based on ancestral inheritance (xenology), such as horizontal gene transfer or hybridization.
- **Long branch attraction (LBA):** placement of two heavily divergent but non-monophyletic sequences with each other. The model is unable to extract the correct evolutionary signal due to the number of mutations that have occurred, so places the two taxa together. Also called the Felsenstein zone.

5.3.2 Summary: Who/what are the Gonyaulacales?

The Gonyaulacales are an order within the super-phylum Alveolata and sub-phylum Dinoflagellata, which are an ancient eukaryotic lineage [173]. They play a role in several important ecological processes in aquatic environments where they cover a diverse array of niches such as symbionts, parasites and autotrophs. Some taxa can cause harmful algal blooms through proliferation (by restricting light and nutrient availability to other organisms) and/or neurotoxin production (e.g. causing paralytic shellfish poisoning, ciguatera fish poisoning) [188]. Dinoflagellates possess large genomes (estimated size range 1.5 to 185 Gbp), with extensive paralogy and repetitive short sequences [25]. In particular paralogy has proven problematic for efforts investigating the genetic content and structure of the dinoflagellates, as this feature has prevented the assembly of genomes apart from draft genomes from symbiodiniacean taxa which possess some of the smaller genomes [132, 153, 241]. Gene duplication, loss, and cDNA recycling is rife within these organisms, therefore they have likely undergone complementary gene deletion events [186, 240, 246]. For a review on the genetic features of dinoflagellates see [188]. While the evolutionary relationship of most orders within the dinoflagellates has been inferred with consistently high support values, one order has often escaped elucidation - the Gonyaulacales. As neurotoxin production, which can accumulate up the food chain, is prevalent in this order, the evolution of the order is of interest to provide a frame of reference for future investigations into how the toxins have evolved [98, 183, 234, 239, 291].

5.4 Materials and methods

5.4.1 Culture conditions

Cultures were isolated from locations as per suppl. [Table 5.2](#) and clonal cultures established by micropipetting single cells through sterile seawater as described in in [128]. Clonal cultures were maintained in 5x diluted F/2 medium [94] and maintained at temperatures indicated in indicated in suppl. [Table 5.2](#).

5.4.2 RNA isolation, library preparation and Sequencing

Gambierdiscus spp. and *Thecadinium kofoidii* were harvested during late exponential growth phase by filtration onto 5 μ m SMWP Millipore membrane filters (Merck, DE) and washed off with sterile seawater. Cells were pelleted via centrifugation for 10 minutes at 350 rcf. The supernatant was decanted and 2ml of TRI Reagent (Sigma-Aldrich, subsidiary of Merck, DE) was added to the pellet and vortexed till resuspended. Samples were split in two and transferred to 1.5ml eppendorf tubes. Cellular thecae were ruptured by three rounds of freeze-thaw, with tubes transferred between liquid Nitrogen and 95 °C. RNA was extracted as per the protocol for TRI Reagent [223]. RNA eluate was purified with the RNeasy RNA clean up kit RNeasy Mini Kit (Qiagen, NL) as per protocol. DNA was digested with TurboDNase (Life Technologies, subsidiary of Thermo Fischer scientific, AU). RNA was quantified with a Nanodrop 2000 (Thermo Scientific, Australia) and frozen at -80 °C until sequencing. The quality of samples was assessed via an Agilent 2100 Bioanalyzer at the Ramaciotti Center (UNSW, AU) and the libraries were prepared using TruSeq RNA Sample prep kit v2 (Illumina, USA). Paired-end sequencing was performed with a NextSeq 500 High Output run at the Ramaciotti Center (UNSW, AU) with 75bp read length for *G. holmesii* and *G. lapillus*; and 150bp read length for *G. carpenterii*, *G. polynesiensis* and *T.kofoidii*.

5.4.2.1 Publicly available transcriptome libraries

From NCBI, the *Gambierdiscus excentricus* VGO790 transcriptome was downloaded via the accession ID SRR3348983 [117], while *Coolia malayensis*, *Ostreopsis ovata*, *Ostreopsis rhodesae* and *Ostreopsis siamensis* transcriptomes were downloaded via the accession IDs SRR9044102, SRR9046040, SRR9047231 and SRR9038703 respectively [269]. Accession numbers are provided in table 5.1. RNA-seq libraries for all remaining transcriptomes were generated by, and downloaded from, the MMETSP [114].

5.4.3 Transcriptome processing scripts

The workflow was separated into two parts. See section [Section 5.4.6.2](#) for script details.

5.4.3.1 Transcriptome assembly

Individual RNA sequencing libraries were processed through FastQC [5] for quality metrics, sequences were trimmed with Trimmomatic (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:5 MINLEN:25) [16] and assembled with Trinity v2.4.0 (default settings for paired end libraries) [79]. Assemblies were then processed with BUSCOv2 with the protist specific library [242]. The RNA libraries with 150bp reads generated as part of this study were also subjected to Digital Normalization [23] prior to assembly, to reduce data set size by removing highly similar sequences, which were then used for downstream analysis.

5.4.3.2 Construction of multiple sequence alignments

The BUSCOv2 output from all transcriptomes from the previous step formed the input for identification of single copy genes and construction of multiple alignments. Any genes that BUSCOv2 identified as single copy and were present in at least 75% of the transcriptomes were indexed, the corresponding contig extracted from the assemblies, aligned with hmmer3.1b2 [52] and unaligned regions trimmed. If several candidate sequences were processed for the same organism, a warning message in the terminal window alerted the user before proceeding. The output for this section was used as a basis for single copy gene phylogenetic inferences in subsequent sections.

5.4.4 Assembly analysis

Contigs from assemblies were clustered with CD-HIT with the flags -T 10 -M 5000 -G 0 -c 1.00 -aS 1.00 -aL 0.005 [66]. Protein coding regions within the clusters were predicted with Transdecoder [78]. Amino acid clusters were clustered again with CD-HIT with the flags as previously except -c 0.98. Protein sequences were analyzed with interproscan v5.27 with local lookup server [209].

5.4.5 Phylogenetic inferences

5.4.5.1 Ribosomal DNA based inference

Ribosomal DNA (rDNA) sequences for the small subunit (SSU) region as well as the D1-D3 large subunit (LSU) region were acquired from NCBI [37] and the SILVA rRNA database project [207], accession IDs in suppl. [Table 5.4](#). Individual genes were aligned using MUSCLE [53] for a maximum of 8 iterations and then were concatenated in Geneious v11.3 [113]. ML phylogenies were inferred using RaxML [252] with the model GTRGAMMA and with 100 bootstrap replicates.

5.4.5.2 Concatenated single copy gene based inference

Amino acid substitution model selection was carried out with ProtTest3 with the Bayesian Information Criterion as well as the log likelihood [43, 76]. The best-fit model for the data set identified by both criteria was VT followed by LG, however neither are available in BEAST2 so the third best model, WAG, was chosen for analysis.

Maximum likelihood with concatenated sequences. ML inference was run as described in the previous [Section 5.4.5.1](#), with the PROT, GAMMA and WAG flags.

Bayesian inference with concatenated sequences. BI was run in BEAST2 with the Gamma site model with 4 discrete categories under the WAG substitution model [275]. A local random clock was used under the birth-death model 3,000,000 million chains.

Bayesian probability under the MSC. BI of the species tree was carried out under the *BEAST2 model in BEAST2 [20]. The analysis was performed with the WAG amino acid substitution model [275] and with a Gamma distribution with four rate categories. A random local clock was employed [50]. Posterior distributions of parameters were approximated after 300,000,000 generations of MCMC, subsampled every 5,000 generations with a burn-in of 15%. The inference was run four times to evaluate convergence of parameters, then log and tree files (without burn-in) were merged.

5.4.6 Stepping stone analysis

We estimated the marginal likelihood of the data under the coalescent (i.e. concatenated alignment) and the MSC (*BEAST) models to compare their fit. We used the stepping stone algorithm by [281] along a path of 30 power posteriors. The β values are set equal to the quantiles of the beta distribution with shape parameter $\alpha = 0.3$ and $\beta = 1$, as recommended by [281].

5.4.6.1 Generation of figures

Tanglegrams were generated with Dendroscope v3.5.9 [102]; images were edited in GIMP [69] to improve readability.

5.4.6.2 Implementation

The analysis workflow in section “Transcriptome processing scripts” was constructed as a Nextflow workflow [47] and is available on Github at https://github.com/hydrahamster/gonya_phylo. Packages within the scripts are written in bash, Python 2.7 [255] and pandas [167]. Source code for the scripts is provided under an open source license. The scripts (1) assemble RNA-seq data sets, (2) identify and extract single copy genes across input taxa with extensive paralogy, and (3) align selected genes in preparation for phylogenetic analysis. The data sets were processed on a Genomics Virtual Lab (GVL) [4] instance in the NeCTAR cloud. Phylogenetic analyses were carried out on the University of Technology Sydney’s High-performance computing cluster (HPCC)

and were accelerated using BEAGLE [7] on the GPU. GPU processing units were either Nvidia Tesla K80 or a Tesla P100.

5.5 Results

Assemblies, annotation files, BUSCOv2 output, single-copy gene alignments and single copy gene MSC BI trace files generated in this study are available on Zenodo doi: [10.5281/zenodo.2576201](https://doi.org/10.5281/zenodo.2576201)

5.5.1 Transcriptomes overview

RNA-seq libraries generated in this study are available in the NCBI sequence read archive (SRA) under the project ID SRP134273. Sequencing of transcriptomes for *Gambierdiscus* spp. and *T.kofoidii* generated data sets ranging in size from 143,155,667 to 233,822,334 reads, resulting in 97,634 to 191,224 assembled contigs ([Table 5.1](#)). Clusters with gene ontology (GO) annotations made up 30.9% to 34.8% of the total clusters.

TABLE 5.1 Summary of transcriptome sequencing and assembly statistics.

<i>Sequences:</i>	<i>G. carpen- teri</i>	<i>G. lapillus</i>	<i>G. polyne- siensis</i>	<i>G. holmesii</i>	<i>T. cf. ko- foidii</i>
Sequencing					
SRA accession	SRR6821720	SRR6821722	SRR6821723	SRR6821721	SRR6821724
Raw sequencing reads	186,422,744	145,366,966	217,031,342	143,155,667	233,822,334
Assembly					
Contigs #	105,464	148,972	114,622	191,224	97,634
Average length (bp)	607	1,139	633	953	581
Maximum length (bp)	7,448	12,370	6,608	8,198	7,922
Transcript clustering & annotation					
# clusters	139,699	92,418	139,487	107,766	116,468
Contigs with GO annotations	44,167	32,140	43,098	34,201	37,656

5.5.2 Single copy gene search with BUSCOv2

Assemblies were searched with BUSCOv2 for 234 candidate single copy genes and homologs to these single copy genes were extracted. The single copy genes acquired through the BUSCO HMMER libraries curated for protists are reported in suppl. [Table 5.3](#), as well as accession numbers and identifiers for each transcriptome. The alignments are available on Zenodo, with the BUSCO gene IDs included in the alignment name.

5.5.3 Phylogenetic inferences

Support for branches was interpreted as follows, for ML and BI, respectively: 100%/1.0 was considered fully supported, above 90%/0.9 was very well supported, 80%/0.8 and above was interpreted as relatively well supported and above 50%/0.5 was considered weakly supported. Below 50%/0.5 was considered unsupported. As *Azadinium spinosum*, *Dinophysis acuminata* and *Karenia brevis* are members of different orders (Dinophyceae ordo incertae sedis, Dinophysiales & Gymnodiniales respectively) and are consistently placed outside of the Gonyaulacales in phylogenetic analyses, their placement as an outgroup was considered a given for this study. Therefore, the branch separating these taxa from others was used to root ML trees in subsequent analyses where rooting was required for tree layout in visual comparisons.

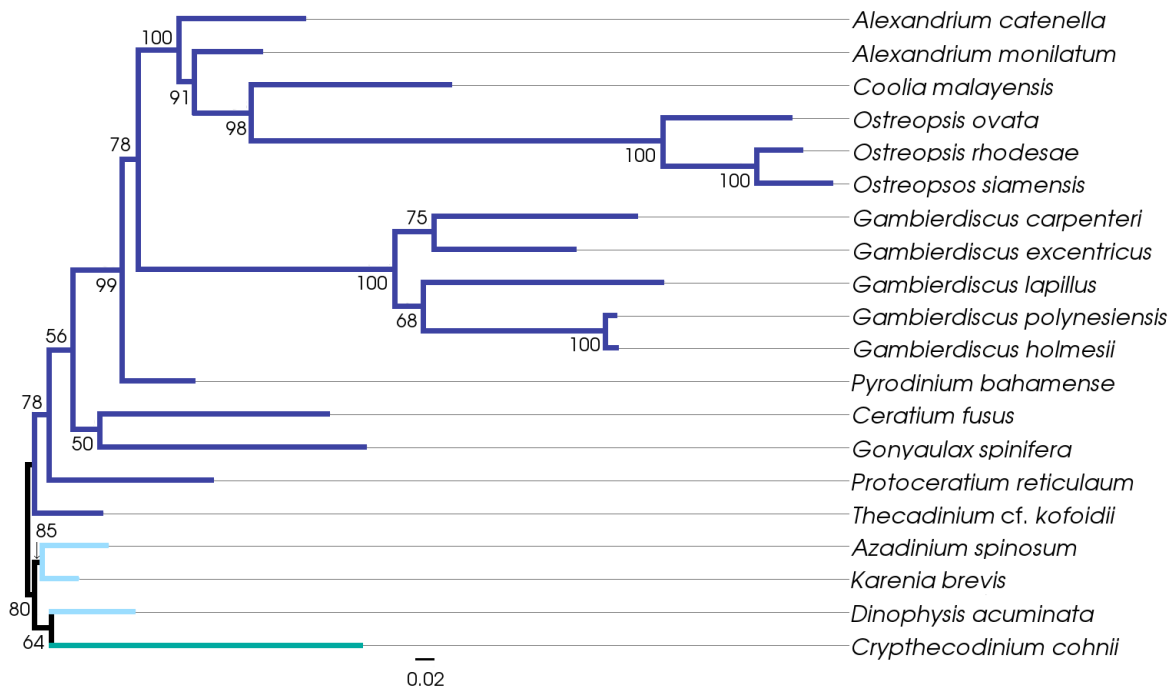


FIGURE 5.1 Maximum likelihood phylogenetic inference of ribosomal DNA genes. Concatenation of small subunit rDNA and D1-D3 region large subunit rDNA. Accession numbers for concatenated genes in suppl. Table 5.4. Gonyaulacales (n=16) in purple, outgroups (n=3) in light blue and taxa *incertae sedis* (n=1) in teal. The topology was rerooted on the branch separating outgroup taxa with the Gonyaulacales. The scale represents the expected number of substitutions per site.

5.5.3.1 rDNA based phylogeny

All nodes were supported, with a range of certainty (Figure 5.1). Species within the genera *Gambierdiscus* and *Ostreopsis* resolved with their sister species with full support. Within the *Gambierdiscus* clade, nodes were either weakly supported or fully supported. The two species of *Alexandrium* resolved as well supported closest relatives, but did not form an individual clade. Deeper nodes were supported but with less certainty than the nodes near the tips. Two distinct clades were observed from the topology: One including *Alexandrium*, *Coolia* and *Ostreopsis*; another with only *Gambierdiscus*. Sister to these clades, in descending order, was *Pyrodinium*, *Ceratium* and *Gonyaulax*, *Protoceratium* and *Thecadinium*. The outgroup were relatively well supported and included *Cryptothecodinium*. Support for deeper nodes varied from weak to well supported.

5.5.3.2 Concatenated single copy gene based phylogeny inferred with ML

All nodes except one within the *Gambierdiscus* species cluster were relatively well supported (Figure 5.2). Species of the genera *Alexandrium*, *Gambierdiscus* and *Ostreopsis* clustered as individual clades with their sister species. The topology showed three distinct, well supported clades: One encompassing *Alexandrium*, *Coolia* and *Ostreopsis*; another which only contained *Gambierdiscus*; and one which includes *Pyrodinium*, *Gonyaulax* and *Protoceratium*. Sister to these

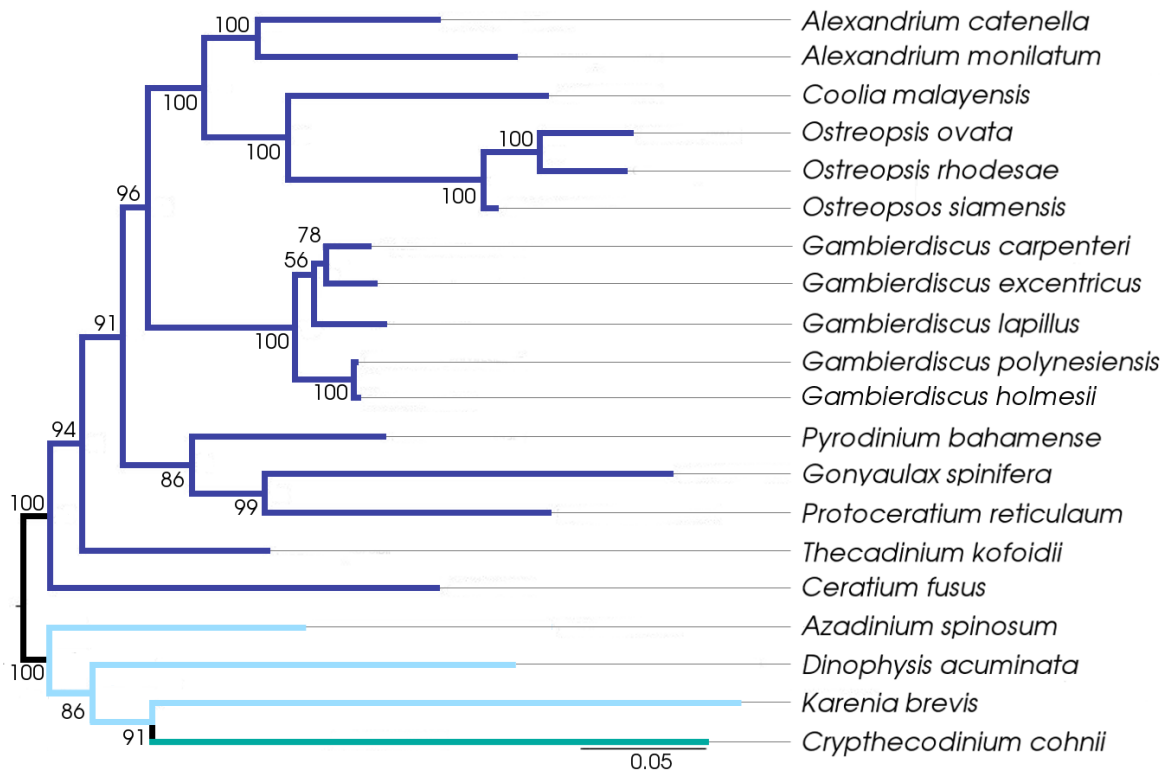


FIGURE 5.2 Maximum likelihood phylogenetic inference of concatenated single copy gene set (62 single copy genes from 20 taxa). Gonyaulacales (#16) in purple, outgroups (#3) in light blue and taxa *incertae sedis* (#1) in teal. Topology was rerooted on the branch separating the outgroup taxa from the Gonyaulacales. The scale represents the expected number of substitutions per site.

clades is *Thecadinium*, followed by *Ceratium*. The split of the outgroup was fully supported, while the internal nodes were very well supported. *Cryptothecodinium* was placed within the outgroup, sister to *Karenia*. Other deeper nodes were well supported.

5.5.3.3 Concatenated single copy gene based phylogeny inferred with BI

All nodes resolved with full support, except one node within the genus *Gambierdiscus* which was very well supported as well as an internal node within the outgroup clade (Figure 5.3). The species in the genera *Alexandrium*, *Gambierdiscus* and *Ostreopsis* were monophyletic with full support. The overall topology of the Gonyaulacales was resolved as three clades with *Thecadinium* and then *Ceratium* as ancestral lineages. *Alexandrium*, *Coolia* and *Ostreopsis* clustered together, followed by *Gambierdiscus* on their own in a sister clade. The third clade encompassed *Gonyaulax*, *Protoceratium* and *Pyrodinium*.

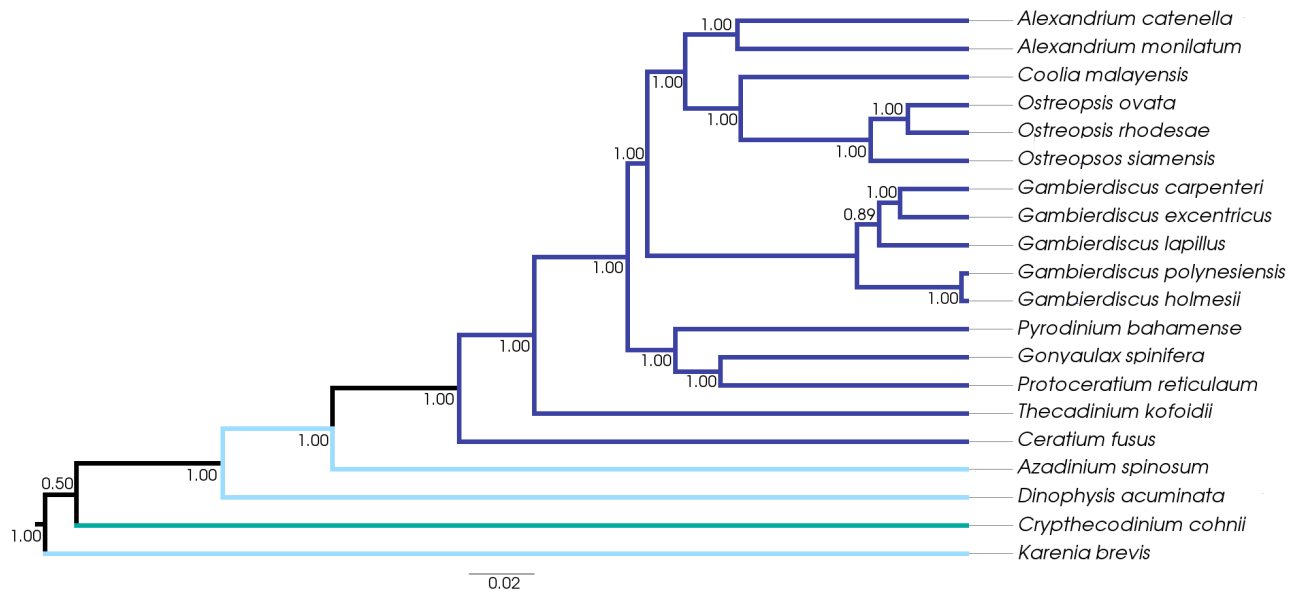


FIGURE 5.3 Bayesian phylogenetic inference of concatenated single copy gene set (62 single copy genes from 20 taxa). Gonyaulacales (#16) in purple, outgroups (#3) in light blue and taxa *incertae sedis* (#1) in teal. The scale represents the expected number of substitutions per site.

5.5.3.4 Single copy gene based phylogeny under MSC

Species of *Alexandrium*, *Ostreopsis* and *Gambierdiscus* were either well or fully supported within their genus clades (Figure 5.4). The topology within the Gonyaulacales resolved into three clades: one fully supported encompassing *Alexandrium*, *Coolia* and *Ostreopsis*; a well supported clade with *Gambierdiscus* and *Pyrodinium*; and a weakly supported clade including *Ceratium*, *Gonyaulax*, *Protoceratium* and *Thecadinium*. The outgroup taxa clustered together with high support. *Cryptothecodinium* was placed as a sister taxon to the outgroup. Other deeper nodes were well or fully supported.

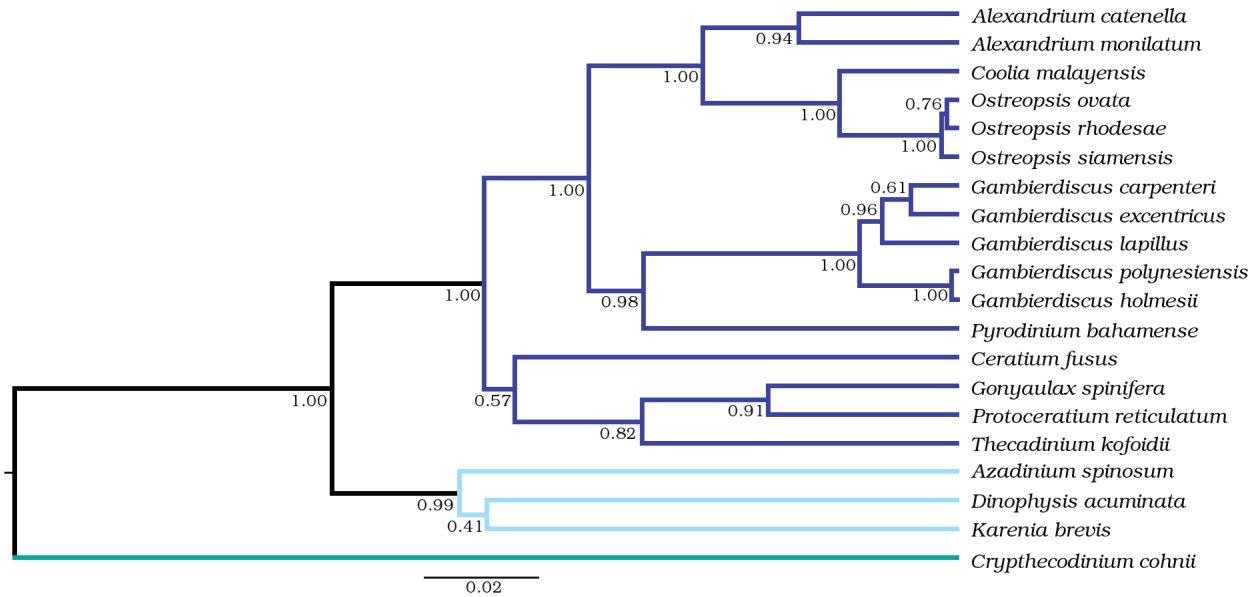


FIGURE 5.4 Bayesian phylogenetic inference of a Gonyaulacales species tree under the MSC model with 62 single copy genes from 20 taxa. Gonyaulacales (#16) in purple, outgroups (#3) in light blue and taxa *incertae sedis* (#1) in teal. The scale represents the expected number of substitutions per site.

5.6 Discussion

Advances in sequencing technologies have seen an increase in high throughput sequencing initiatives such as MMETSP, which revealed the genomic diversity of a relatively uncharacterized group of marine microbial eukaryotes [114]. However, the methodologies used for investigating the evolutionary relationships using this type of genome-scale data remain an obstacle, as the choice of input data and method employed influences the outcome of the inference. In particular, the effects of paralogs and pseudoorthologs (hidden paralogy) are particularly problematic as they can lead to incorrect inference with classic phylogenetic methods. To address this, a synopsis on a method for single copy gene extraction, and synthesis of phylogenetic inference model availability and selection is presented in this study - as well as possible shortcomings of the parameters and methods selected.

Dinoflagellates are notorious for their large genomes with suspected whole or partial genome duplication and potential cDNA retro-insertion into the genome [12, 101, 152, 246, 266]. This can lead to unusually high gene copy numbers and extensive paralogy. With this in mind, the Gonyaulacales (an order within the dinoflagellates, see [Section 5.3.2](#)) represented a good case study for examining the impacts of paralogy on phylogenetic inference. This study presents the first species tree for the Gonyaulacales that has been inferred with a method robust to paralogy, including hidden paralogy.

The phylogenetic inference for Gonyaulacales that resulted from the workflow we developed, which incorporates several of the most recent innovations in analytical methodology, resolved within-genus relationships well and showed high posterior probability support throughout the

species tree (Figure 5.4). The inferred species tree topology followed a broad revised taxonomic classification of the Gonyaulacales based on morphological characteristics [96] and was used as a point of comparison to results from other commonly employed methods in later sections. The scripts which form the basis of this study are publicly available through github and the single copy gene alignments used to infer the species trees, as well as the XML input and log files for the *BEAST2 runs, are available on zenodo (doi: 10.5281/zenodo.2576201). Our study was designed to be transparent and reproducible for those with basic programming skills.

5.6.1 Considerations for data set selection and pre-processing

5.6.1.1 Quantity of taxa in phylogenetic inference

Two phenomena that can confound the veracity of conclusions drawn from phylogenetic inference are ILS and LBA. The impact of ILS on phylogenetic inference has been explored through simulated data sets with a known species tree. When species have recently diverged, sampling more individuals per species can improve resolution of the species tree. However, when the species divergences are older, as is the case here, using more gene loci per species yields greater resolving power than sampling more individuals per species [164].

LBA can arise if some species have disproportionately high substitution rates, leading to the presence of long and short branches in the phylogenetic tree [160]. The risk of LBA artefacts can be reduced by denser taxon sampling to break up long branches and ensuring that the models specified are appropriate [87]. The Gonyaulacales data set in this study included a single representative species per genus, with the exception of *Alexandrium*, *Gambierdiscus* and *Ostreopsis*. This resulted in some genera on long branches (eg. Figure 5.4: *Ceratium fusus* & *Pyrodinium bahamense*) indicative of a proportionally large number of genetic changes to their closest relative. This tree shape was consistent with sparse taxon coverage and can lead to LBA artefacts [87]. To investigate the presence of ILS and as a topological comparison to the BI and ML inferences, a neighbor-joining (NJ) inference was run as well (Phylip with Protdist JTT matrix and neighbor packages [58]). The rationale for evaluating this method was that NJ can recover an accurate species topology despite ILS in cases where ML would fail [169]. However NJ is more susceptible to LBA than ML or BI methods. The resulting topology was so anomalous, with out and in-groups clustering together as well as negative length branch lengths, that we chose to exclude it from further discussion. Both BI and ML are more robust to the effects of LBA than NJ, where BI tends to outperform ML especially if the latter is performed conjunction with concatenation [129, 225].

5.6.1.2 Quality of transcriptome assemblies

Publicly available data sets may have been generated with a variety of different methods, and their resulting quality can be highly variable, so an initial quality assessment step is essential. In the time since the MMETSP data sets were made available, several studies have utilized a broader range

of taxa to explore evolutionary stories involving the Gonyaulacales. However, these have relied on the assemblies supplied as part of the project. The stringency for quality trimming of RNA-seq libraries prior to assembly plays a role in determining the number of unique contigs recovered and the subsequent assembly quality of transcriptomes. Regarding the transcriptome assembly method, [110] evaluated the publicly available assemblies from MMETSP using BUSCO scores, compared to processing and re-assembly with Trinity [110]. [110] demonstrated that while the raw data available from the MMETSP project is an excellent resource, the assemblies available as part of the project are of a lower quality than what can be achieved with current methods [110]. Another factor in assembly quality is RNA-seq data processing prior to assembly, especially trimming. High stringency is usually favored, however [163] found that this can be detrimental to the assembly and the quality cut off scores used in the present study were based on those recommendations [163]. In short, the trimming and assembly pipeline used for the assemblies available as part of MMETSP is no longer state-of-the-art and this is reflected in the quality comparison conducted by [110]. To address this problem, we developed a workflow implementation of current best-practice transcriptome assembly methods as part of this study.

Assembly parameters. Trinity was chosen as the assembler for this study based on the findings of [95], in which Trinity was one of the top performing assemblers for *de novo* transcriptomes as tested with *Arabidopsis thaliana*. Further, Trinity performed well for identifying isoforms of genes and excelled at assembling highly expressed genes [95]. Conversely, [27] found that Trinity, CLC Bio and IDBA-Tran assemblies all contain errors introduced by the assembly algorithms. Using a combination of all three assemblers yielded a final assembly closer to biological reality than any individual assembler, when no reference genome is available [27]. As our present study used Trinity exclusively, it may be subject to the type of errors found by [27] which could affect downstream analysis.

5.6.1.3 Selection of paralogs to infer species evolution

Inclusion of genes which diverged through a process other than speciation events, such as paralogs, violates the assumptions of most commonly used phylogenetic models, which typically assume all genes analysed have an orthologous relationship. Our study sought to mitigate the issues arising from paralogs in two ways: (1) by identifying and using single copy genes and (2) using a phylogenetic inference method that is robust to the presence of pseudoorthologs (hidden paralogs). Single copy genes were identified via the curated BUSCO gene collection and software. As BUSCO uses lineage specific profile HMM libraries designed to target single copy genes, and the output distinguishes between single copy genes and duplications, it presents a method for reliably screening for single copy genes for phylogenomics [274].

The first study to address the issue of paralogy for species tree inference within the dinoflagellates was published in 2017 [205]. That study excluded paralogs by using single copy genes as input for the phylogenetic inference, similar to the first part of the approach we have used. A second

study by [254] expanded on the dataset by [205] by applying the same methodology to a somewhat larger group of species. Since it includes the order Gonyaulacales and represents a comprehensive baseline for both studies, we compare our phylogenies to that of [205] and mention points of difference in the [254] phylogeny when relevant.

The phylogenies inferred by [205] and by our study resulted in markedly different topologies. Specifically, the placement of two sister taxa (Figure 5.5) are noteworthy: [205] placed *Alexandrium* spp. as the closest genus to *Gambierdiscus*, while our study placed *Pyrodinium* as the sister to *Gambierdiscus*. Interestingly, one of the few points of difference between the [205] and the [254] inference topologies was that the latter placed *Pyrodinium* as the sister genus to *Gambierdiscus* too. Similarly, in [205] the *Azadinium* is part of the Gonyaulacales, while this study firmly places this genus as an outgroup with *Dinophysis* spp. and *Karenia* spp. Given that some *Gambierdiscus* spp. as well as *Azadinium* spp. produce toxins that cause severe fish and shellfish poisoning it is of high importance for the analysis of toxin evolution to infer the phylogenetic relationships of these and closely related taxa [198]. Potential factors that may have impacted the present study and explain the differences between the two phylogenies are discussed in detail below. Additionally, factors that may have impacted the phylogeny published by [205] which could also explain the observed differences are (i) older assembly methods used in the MMETSP data set, (ii) the concatenation of genes for the alignment as well as (iii) the use of a ML estimation method. In our opinion, especially the use of concatenated alignments in conjunction with ML inference methods (as discussed previously) makes the study published by [205] susceptible to the effects of pseudoorthologs (e.g. hidden paralogs). Unfortunately, a more rigorous comparison between the two approaches was not possible as the methodology for the identification of single-copy genes was neither reported by [205] nor available on request.

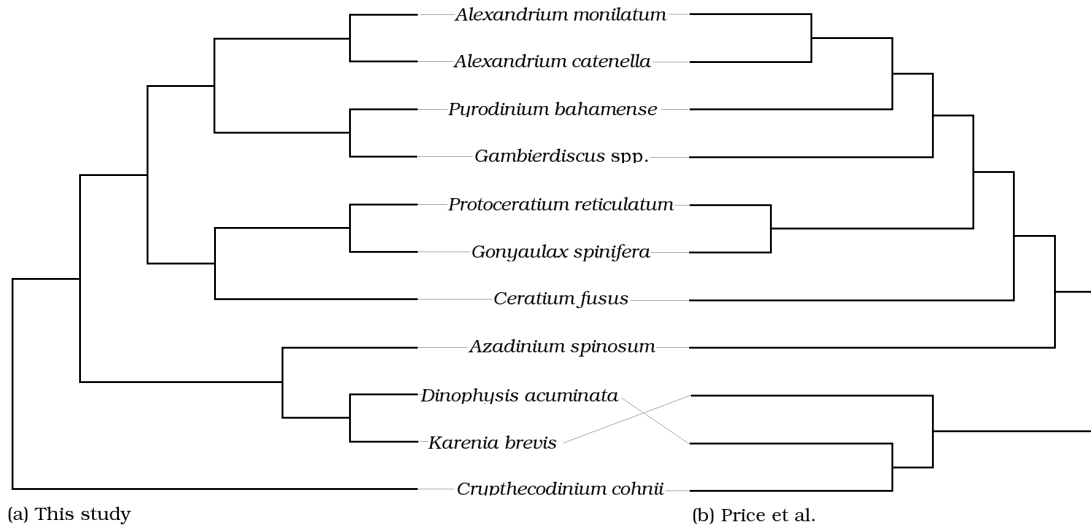


FIGURE 5.5 Tanglegram of the single copy gene topologies presented in (a) this study under MSC; and (b) concatenated by [205]. Taxa not common to either study are not shown due to the reduced topologies from the original studies, closest PP or BS to branch split were included.

5.6.1.4 Model selection for inference

The issue of model choice is an important one, as the choice of model can heavily influence the resulting topology. Mis-specification of the model, or individual parameters, can lead to a well supported but erroneous result. While models are a simplistic approximation of the underlying biological drivers of evolutionary processes, getting as close an approximation as possible is essential [21]. However under- and over-parameterization have been shown to impact topology and PPs to varying degrees, in and outside the Felsenstein zone [139]. Marginal likelihood comparison penalizes for over-parameterization and can be used to compare the fit of one model compared to another for a given data set [280]. To compare how well concatenation vs. MSC fits the single copy gene data set used in this study, stepping stone comparison was conducted using the model-selection package in BEAST2 [20].

5.6.2 Comparison to commonly employed models and data sets

5.6.2.1 Phylogenetic inference using ribosomal genes

Using LSU or SSU rDNA regions for phylogenetics is common practice, at times supplemented with a small number of other genes [98, 183, 234, 239, 291]. It is important to acknowledge that these represent the evolutionary history of highly conserved genes, which does not necessarily represent the species evolution and assumptions of their congruence is statistically inadequate [46]. Yet, because rDNA sequencing is easy and inexpensive it continues to be employed for the Gonyaulacales even if it does not yield comprehensive results.

Comparing the topology from a rDNA ML inference with the single gene copy MSC phylogeny presented here ([Figure 5.6](#)) shows that most clades in both topologies were completely or very well supported. Within the genera *Gambierdiscus* and *Ostreopsis*, the species resolution differed between the two data sets. In several cases, the placement of sister taxa was incongruous between the two analyses. For example, the rDNA concatenation data set places *Ceratium* & *Gonyaulax* as well as *Alexandrium* and *Gambierdiscus* as sister taxa, while the single copy gene data set under MSC places *Gonyaulax* with *Protoceratium* and *Gambierdiscus* with *Pyrodinium*. This is an example of how using rDNA segments as a proxy for species evolution produces different results than an analysis of single-copy protein coding genes.

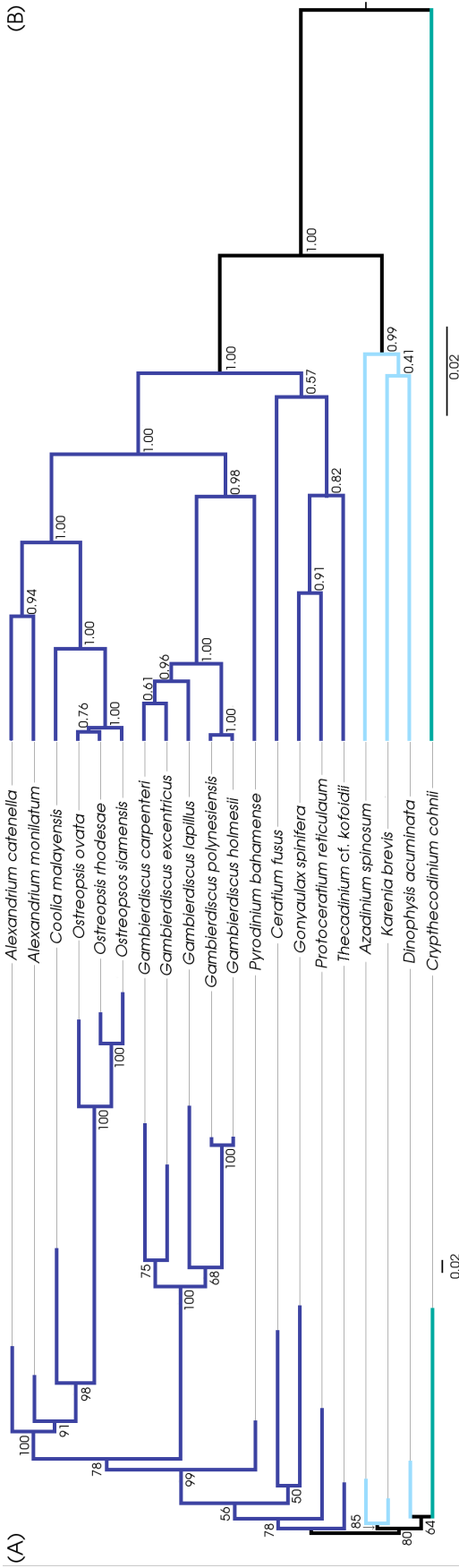
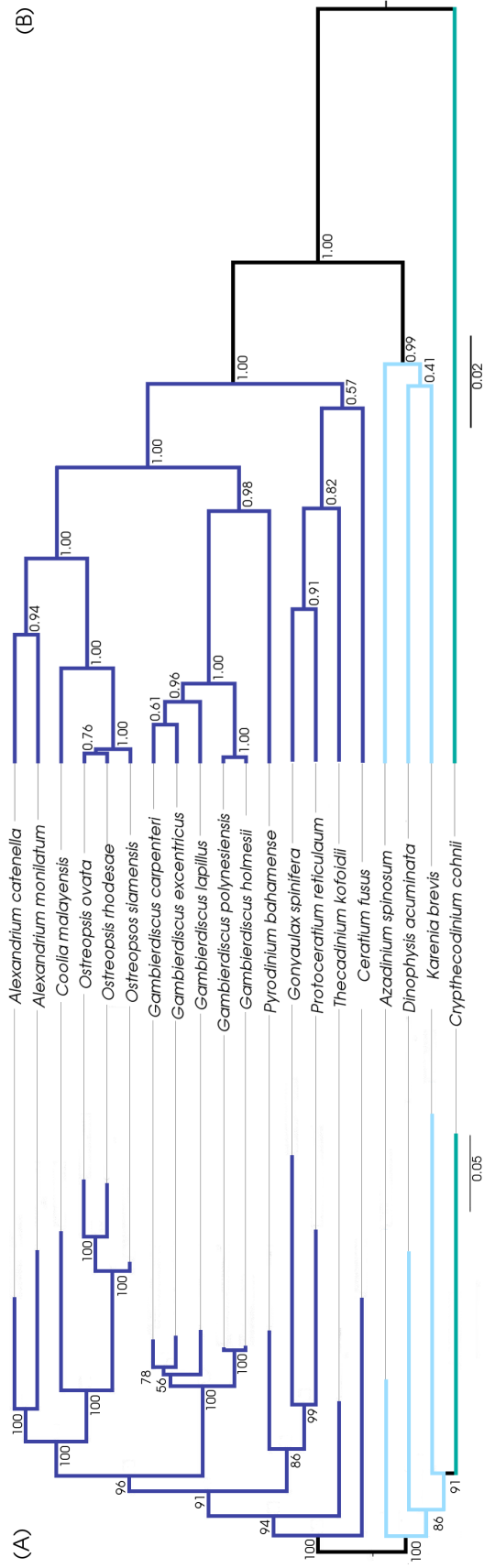


FIGURE 5.6 Tanglegam showing the topological differences in phylogenies from (A) concatenated rDNA genes (SSU and D1-D3 LSU) inferred with ML; and (B) MSC inference with 58 single copy genes. Gonyaulacales (#16) in purple, outgroups (#3) in light blue and taxa *incertae sedis* (#1) in teal.

5.6.2.2 Concatenating selected genes and using ML methods for species inference

Concatenation of alignments coupled with ML inference is a commonly used method as it is less computationally demanding than BI methods. However as demonstrated by [129] and [225], this approach is error prone. Concatenation assumes uniform evolutionary history across genes, with a small amount of variation possible - however this still averages the evolutionary rate for all the input genes which doesn't allow for divergent gene histories [225]. The combination of concatenation and ML for phylogenetic inference can result in high bootstrap values for incorrectly resolved clades, over inflating confidence in erroneous topologies [46]. The application of concatenation in combination with ML is common practice in phylogenetic studies for gonyaulacoids [98, 183, 234, 239, 291]. We investigated whether the use of a technique explicitly designed to handle multiple genes to estimate species trees would yield different results than concatenation and ML.

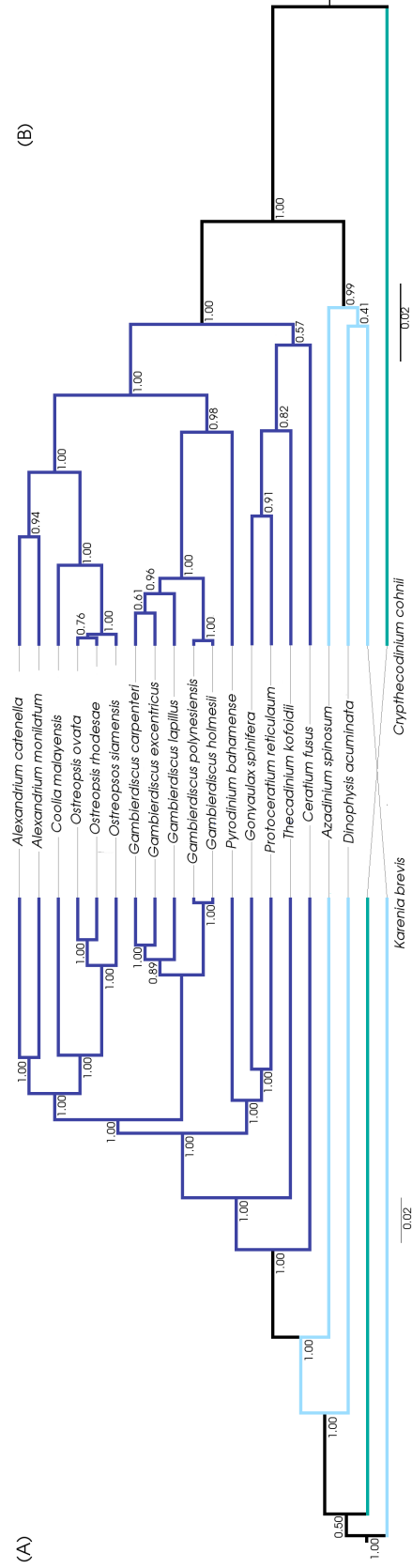
A comparison between a BI inference under MSC and concatenated ML inference on the same single copy gene data set showed differences in topology ([Figure 5.7](#)). The species resolution within the genera *Alexandrium*, *Gambierdiscus* and *Ostreopsis* matched between the two inference methods. The major difference was in the *Pyrodinium* placement, where the BI MSC approach places the genus sister to *Gambierdiscus* while the concatenated ML approach places it with *Gonyaulax* and *Protoceratium*. Further, the deeper branches of the phylogenies differ. The BI MSC method clusters *Ceratium*, *Gonyaulax*, *Protoceratium* and *Thecadinium* as a clade, while the concatenated ML approach clusters *Gonyaulax*, *Protoceratium* and *Pyrodinium* as a clade to which *Thecadinium* and then *Ceratium* feature as ancestral genera.



5.6.2.3 Concatenating selected genes and using BI methods for species inference

Even within a BI framework concatenation can introduce a number of errors. Under simulated data sets, even under the coalescent methods, the species tree topology is inaccurate when concatenation is used [129]. Further to that, the PP values tend to be overestimated for concatenation [256]. Theoretically for the Gonyaulacales, and taxa prone to paralogy and convoluted evolutionary histories, the MSC is a preferable approach to concatenation as MSC is more robust to ILS and LBA artifacts as well as pseudoorthologs [51, 160].

To isolate the effects of phylogenetic model from those of the statistical framework (ML vs BI), the single copy gene data set was run with BI both under MSC and with concatenation (Figure 5.8). We then used a statistical framework to compare the two model approaches to verify the veracity of model adequacy through stepping stone sampling. Stepping stone is a method for estimating marginal likelihoods of phylogenetic models, enabling model comparison and selection of the model with the better fit [10, 281]. The marginal likelihood of the MSC model (-160538.6) was over 10,000 log units higher than that of the concatenated single copy gene model (-170866.6), favoring the MSC approach significantly. The large difference in marginal likelihood between the models could be in part due to the inclusion of pseudoorthologs in the dataset, against which MSC models are more robust than the concatenation approach [51, 225]. The resolution of *Alexandrium*, *Coolia* and *Ostreopsis* was identical between the two methods. Further, the species resolution within the genera *Gambierdiscus* and *Ostreopsis* was also identical (Figure 5.8). Differences were found in the topology, in that *Pyrodinium* clustered with *Gambierdiscus* in the MSC analysis, while for concatenation this genus clusters with *Gonyalax* and *Protoceratium*. The *Pyrodinium* placement also differed to the study by [205] (Figure 5.5), where the genus was more closely related to *Alexandrium* rather than *Gonyaulax* and *Protoceratium* in the BI topology. Further, in the MSC analysis *Ceratium*, *Gonyalax*, *Protoceratium* and *Thecadinium* formed their own clade while with concatenation, *Ceratium* and *Thecadinium* were ancestral genera to the rest of the Gonyaulacales. There was a marked difference in the internal branch arrangement, which resulted in different taxa clustering, between the concatenation and MSC methods. The concatenated approach closely mirrored the ML arrangement of taxa, apart from *Cryptothecodinium* placement. Both inferences were topologically distinct to the MSC approach.



5.6.3 Areas for possible improvement of this study

In the previous section we identified potential problems with common approaches to species inference in the literature, and in particular for the Gonyaulacales. We then sought to evaluate the effects of different methodological approaches on analytical results in the Gonyaulacales. There are several important limitations to our study.

5.6.3.1 Contamination of other taxa

The 650+ RNA extract submission to MMETSP was from a large number of investigators and low level contamination is inherent in the project's data set [114]. As the cultures tested in all the studies contributing to this data set were not axenic, contamination could be bacterial or eukaryotic in nature. While any contaminating bacterial genes in our data would likely be heavily diverged and therefore obvious, eukaryotic contamination may be more subtle.

5.6.3.2 No representative genome for comparison

Without an available reference genome, it is difficult to evaluate the accuracy of the transcriptome assembly and whether the genes selected are single copies, or misassemblies of paralogs.

5.6.3.3 Different methods for RNA-seq

Three different approaches for RNA-seq library generation were employed for the libraries used in this study, the MMETSP taxa were sequenced on HiSeq platform with 50nt reads; while all other taxa were sequenced on the NextSeq platform with 75nt or 150nt reads. The different sequencing methods may each influence the single copy gene coverage and transcriptome assembly accuracy, leading to systematic error and batch effects on some taxa.

5.6.3.4 Total evidence phylogenetics

The method presented here purely considers the information contained in the genetic aspect of the organisms examined. Morphological characters and fossil dates can add another dimension to the phylogenetic inference and put the evolution within a relative time frame [68].

5.6.3.5 Model comparison methods

The method presented here purely considered the information contained in the genetic aspect of the organisms examined. Morphological characters, if evolutionarily relevant ones can be identified, and fossil dates can add another dimension to the phylogenetic inference and put the evolution within a relative time frame [68].

5.7 Conclusion

This study presents a workflow for species tree inference that implemented what is currently thought to be the best practice methods. The scripts processed RNA-seq libraries through assembly, single copy gene selection to alignment for phylogenetic species inference. As a case study exemplifying organisms rife with paralogs and ancient lineages, the Gonyaulacales were selected. The resulting phylogeny showed a well resolved, well supported inference of the Gonyaulacales evolution. This was then compared to phylogenies inferred from commonly utilized methods in the literature, and potential issues arising from these methods were discussed. By presenting a statistically rigorous method and demonstrating how it overcomes common problems in phylogenetic studies, we hope that in the future such robust, reproducible, open-access approaches to process large data-sets such as the MMETSP database can become standard practice.

5.8 Acknowledgments

The GVL section of this study was conducted inside the National eResearch Collaboration Tools and Resources (NeCTAR) research cloud, an initiative by the National Research Infrastructure for Australia (NCRIS). Gratitude to the Stanley Watson foundation, the Linnaean Society of New South Wales, and the ABRS National Taxonomy Research Student Travel Bursary for funding A. L. Kretzschmar's attendance at the Molecular Evolution workshop at the Marine biological laboratory, Woods Hole, MA, USA. Shout out to the Taming the BEAST organizers & fellow attendees for a most illuminating workshop in February 2017 on BEAST methodology, and to Geneious for subsidizing A. L. Kretzschmar's attendance fee. The transcriptomic sequencing was funded by A. L. Kretzschmar's student funding and in part by an ARC Future Fellowship to S. Murray. A. L. Kretzschmar's PhD stipend was funded through a UTS Doctoral scholarship.

5.9 Supplementary material

TABLE 5.2 Supplementary culturing conditions for species processed for this study.

Species	Strain	Temp	Source location
<i>Gambierdiscus carpenteri</i>	UTSMER9A	17	Merimbula, AU
<i>Gambierdiscus lapillus</i>	HG4	27	Heron Island, AU
<i>Gambierdiscus polynesiensis</i>	CG15	27	Rarotonga, COK
<i>Gambierdiscus holmesii</i>	HG5	27	Heron Island, AU
<i>Thecadinium</i> cf. <i>kofoidii</i>	THECA	18	Gordons bay, Sydney, AU

TABLE 5.3 Supplementary transcriptomes used for study along including strain ID, source and BUSCOv2 information. MMETSP abbreviation for marine Microbial eukaryotic transcriptome sequencing project, by Moore Foundation.

Species	Strain	complete BUSCOs	single complete BUSCOs	fragmented BUSCOs	Source
Gonyaulacales transcriptomes					
<i>Alexandrium catenella</i>	OF101	110	74	3	MMETSP0790 [114]
<i>Alexandrium monilatum</i>	JR08	107	74	3	MMETSP0093 [114]
<i>Ceratium fusus</i>	PA161109	121	81	4	MMETSP1074 [114]
<i>Coolia malayensis</i>	MAB	138	100	1	[269]
<i>Cryptothecodinium cohnii</i>	Seligo	126	98	0	MMETSP0326_2 [114]
<i>Gambierdiscus carpenteri</i>	UTSMER9A	101	83	2	This study
<i>Gambierdiscus excentricus</i>	VGO790	88	83	4	[117]
<i>Gambierdiscus lapillus</i>	HG4	141	98	2	This study
<i>Gambierdiscus poly-nesiensis</i>	CG15	104	81	3	This study
<i>Gambierdiscus holmesii</i>	HG5	134	87	2	This study
<i>Gonyaulax spinifera</i>	CCMP409	83	53	2	MMETSP1439 [114]
<i>Ostreopsis ovata</i>	HER27	132	99	2	[269]
<i>Ostreopsis rhodesae</i>	HER26	131	98	1	[269])
<i>Ostreopsis siamensis</i>	BH1	132	98	1	[269]

<i>Protoceratium reticulatum</i>	CCCM535=CCMP1889	108	72	5	MMETSP0228 [114]
<i>Pyrodinium bahamense</i>	pbaha01	119	897	2	MMETSP0796 [114]
<i>Thecadinium</i> cf. <i>kofoidii</i>	THECA	93	70	5	This study
Outgroup transcriptomes					
<i>Azadinium spinosum</i>	3D9	1.8	81	4	MMETSP1036_2 [114]
<i>Dinophysis acuminata</i>	DAEP01	117	74	2	MMETSP0797 [114]
<i>Karenia brevis</i>	CCMP2229	115	85	2	MMETSP0030 [114]

TABLE 5.4 Supplementary accession numbers for ribosomal DNA sequences used for Figure 5.1. Sequences sourced from NCBI, except accession numbers with '*' sourced from the Silva database. Genes not publically available are denoted by '-'.

Species	SSU seq.	D1-D3 LSU seq.
Gonyaulacales taxa		
<i>Alexandrium catenella</i>	AB088286	AB088238
<i>Alexandrium monilatum</i>	AY883005	-
<i>Ceratium fusus</i>	AF022153	AF260390
<i>Coolia malayensis</i>	HQ897279*	KX589143
<i>Crypthecodinium cohnii</i>	M64245	-
<i>Gambierdiscus carpenteri</i>	EF202908	EF202938
<i>Gambierdiscus excentricus</i>	GETL01000157*	HQ877874

<i>Gambierdiscus lapillus</i>	KU558930	-
<i>Gambierdiscus polynesiensis</i>	EF202907	This study
<i>Gambierdiscus holmesii</i>	This study	this study
<i>Gonyaulax spinifera</i>	AF022155	DQ151558
<i>Ostreopsis ovata</i>	AF244939	KJ781420
<i>Ostreopsis rhodesae</i>	KX055855	KX055845
<i>Ostreopsis siamensis</i>	KX055868	HQ414223
<i>Protoceratium reticulatum</i>	AF274273	EF613362
<i>Pyrodinium bahamense</i>	AY456115	AB936757
<i>Thecadinium</i> cf. <i>kofoidii</i>	AY238478	KT371445
Outgroup taxa		
<i>Azadinium spinosum</i>	JN680857	JN165101
<i>Dinophysis acuminata</i>	AJ506972	EF613351
<i>Karenia brevis</i>	EF492504	AY355458

CHAPTER 6

Discussion

6.1 General discussion

Ciguatera is a global problem, which is predicted to increase in prevalence and severity as the effects of climate change unfold. Despite the public health risks posed by CFP, very little is known about how to monitor for outbreaks. To address this, a global ciguatera strategy was called into action by UNESCO's Intergovernmental Oceanographic Commission. This strategy is a concerted effort by a team of international experts to address the issues posed by CFP holistically through three elements: 1) Improve organism detection, monitoring and risk forecasting; 2) Improve toxin detection in cells and fish; and 3) Improve epidemiological data collection, reporting and assessments [105].

This thesis contributes to element 1 of the strategy by identifying a new species of *Gambierdiscus* (Chapter 2), developing molecular environmental monitoring assays for the new toxic species (Chapter 3) and then exploring the placement of *Gambierdiscus* within a greater evolutionary context including other toxin producers (Chapter 5) and by identifying gene expression patterns common to *Gambierdiscus* as well as specific to toxic specimens (Chapter 4). As outlined in the strategy, investigations into the factors at play for ciguatera risk need to be multi-faceted. This body of work approaches the issue of ciguatera risk factors from two angles:

6.1.1 Targeting the species at bottom of the food chain

For this purpose, the species diversity in a ciguateric region of the GBR was explored, which uncovered a new species, *G. lapillus*, in Chapter 2. A molecular assay for the new species was developed and tested in Chapter 3. The assay(s) allow(s) for directly monitoring environmental samples and deducing the number of cells of each species present.

6.1.2 Targeting toxin production potential directly

The first step for this approach is to characterize the toxicity and toxin profile of the species in Chapter 2. Due to the current lack of knowledge about the toxin synthesis pathway(s), a baseline

needed to be established before the enzyme evolution can be extrapolated. To do so, [Chapter 5](#) sought to infer the evolution of the genus in the greater context of toxic and non-toxic sister genera. Then to establish a baseline of gene expression in the genus, a core transcriptome for *Gambierdiscus* was derived in [Chapter 5](#). [Chapter 5](#) further described transcript clusters unique to both toxic and non-toxic species for future exploration.

6.2 Significance and future directions

Climate change is set to transform every ecological niche on the planet. One of the predicted impacts is the change in abundance, habitat range and dominance in local species composition of the genus *Gambierdiscus*. Ciguatera cases are directly linked to ciguatoxin production by *Gambierdiscus* and subsequent bioaccumulation through the food web, hence the climate induced change for *Gambierdiscus* is likely to extend to ciguatera. This thesis has refined the understanding of the diversity within the genus *Gambierdiscus* by describing a new toxic member and provided a molecular environmental monitoring tool for the new species. Further, the evolutionary analysis of *Gambierdiscus* and related taxa using current best practice methods, and the identification of a core transcriptome, provide a basis for future investigations of *Gambierdiscus*. In particular, new avenues of investigation the toxin production pathways will be supported by this work.

6.3 Thesis conclusion

In conclusion this thesis lays fundamental groundwork for understanding the genus *Gambierdiscus* from an evolutionary perspective, its associated toxicology, and provides a novel environmental monitoring tool.

CHAPTER 7**Bibliography**

Bibliography

- [1] Abraham, A., Jester, E. L., Granade, H. R., Plakas, S. M., and Dickey, R. W. Caribbean ciguatoxin profile in raw and cooked fish implicated in ciguatera. *Food chemistry* 131, 1 (2012), 192–198.
- [2] Adachi, R., and Fukuyo, Y. The thecal structure of a marine toxic dinoflagellate *Gambierdiscus toxicus* gen. et sp. nov. collected in a ciguatera-endemic area. *Bulletin of the Japanese Society of Scientific Fisheries (Japan)* 45 (1979), 67–71.
- [3] Adams, M. J. An outbreak of ciguatera poisoning in a group of scuba divers. *Journal of Wilderness Medicine* 4, 3 (1993), 304–311.
- [4] Afgan, E., Sloggett, C., Goonasekera, N., Makunin, I., Benson, D., Crowe, M., Gladman, S., Kowsar, Y., Pheasant, M., Horst, R., et al. Genomics virtual laboratory: a practical bioinformatics workbench for the cloud. *PloS one* 10, 10 (2015), e0140829.
- [5] Andrews, S. FastQC: A quality control tool for high throughput sequence data, 2010. bioinformatics.babraham.ac.uk/projects/fastqc/.
- [6] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. Gene Ontology: tool for the unification of biology. *Nature genetics* 25, 1 (2000), 25.
- [7] Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., Huelsenbeck, J. P., Ronquist, F., Swofford, D. L., Cummings, M. P., et al. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Systematic biology* 61, 1 (2011), 170–173.
- [8] Bachvaroff, T. R., Concepcion, G. T., Rogers, C. R., Herman, E. M., and Delwiche, C. F. Dinoflagellate expressed sequence tag data indicate massive transfer of chloroplast genes to the nuclear genome. *Protist* 155, 1 (2004), 65–78.

- [9] Bachvaroff, T. R., and Place, A. R. From stop to start: tandem gene arrangement, copy number and trans-splicing sites in the dinoflagellate *Amphidinium carterae*. *PLoS One* 3, 8 (2008), e2929.
- [10] Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A., and Lemey, P. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular biology and evolution* 30, 2 (2012), 239–243.
- [11] Banner, A. H., and Helfrich, P. *The distribution of ciguatera in the tropical Pacific*. Hawaii Marine Laboratory, University of Hawaii, 1964.
- [12] Beauchemin, M., Roy, S., Daoust, P., Dagenais-Bellefeuille, S., Bertomeu, T., Letourneau, L., Lang, B. F., and Morse, D. Dinoflagellate tandem array gene transcripts are highly conserved and not polycistronic. *Proceedings of the National Academy of Sciences* 109, 39 (2012), 15793–15798.
- [13] Beedessee, G., Hisata, K., Roy, M., Van Dolah, F., Satoh, N., and Shoguchi, E. Comparative genomics-first approach to understand diversification of secondary metabolite biosynthetic pathways in symbiotic dinoflagellates. *bioRxiv* (2018).
- [14] Beedessee, G., Hisata, K., Roy, M. C., Satoh, N., and Shoguchi, E. Multifunctional polyketide synthase genes identified by genomic survey of the symbiotic dinoflagellate, *Symbiodinium minutum*. *BMC genomics* 16, 1 (2015), 941.
- [15] Berdalet, E., Bravo, I., Evans, J., Fraga, S., Kibler, S., Kudela, M., Larsen, J., Litaker, W., Penna, A., Tester, P., et al. Global ecology and oceanography of harmful algal blooms, GEOHAB core research project: HABs in benthic systems. *GEOHAB report* (2012).
- [16] Bolger, A. M., Lohse, M., and Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 15 (2014), 2114–2120.
- [17] Bomber, J. W. *Ecology, genetic variability and physiology of the ciguatera-causing dinoflagellate Gambierdiscus toxicus* Adachi & Fukuyo. 1987.
- [18] Bomber, J. W., Guillard, R. R., and Nelson, W. G. Roles of temperature, salinity, and light in seasonality, growth, and toxicity of ciguatera-causing *Gambierdiscus toxicus* Adachi et Fukuyo (Dinophyceae). *Journal of experimental marine biology and ecology* 115, 1 (1988), 53–65.
- [19] Bomber, J. W., Rubio, M. G., and Norris, D. R. Epiphytism of dinoflagellates associated with the disease ciguatera: substrate specificity and nutrition. *Phycologia* 28, 3 (1989), 360–368.
- [20] Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., et al. BEAST 2.5: An advanced

- software platform for Bayesian evolutionary analysis. *PLoS computational biology* 15, 4 (2019), e1006650.
- [21] Box, G. E. All models are wrong, but some are useful. *Robustness in Statistics* 202 (1979).
- [22] Bravo, I., Figueroa, R. I., and Fraga, S. Cellular and nuclear morphological variability within a single species of the toxigenic dinoflagellate genus *Gambierdiscus*: Relationship to life-cycle processes. *Harmful Algae* 40 (2014), 1–8.
- [23] Brown, C., Scott, C., Crusoe, M., Sheneman, L., Rosenthal, J., and Howe, A. khmer-protocols documentation, 2013. <http://dx.doi.org/10.6084/m9.figshare.878460>.
- [24] Caillaud, A., De la Iglesia, P., Darius, H. T., Pauillac, S., Aligizaki, K., Fraga, S., Chinain, M., and Diogène, J. Update on methodologies available for ciguatera toxin determination: perspectives to confront the onset of ciguatera fish poisoning in Europe. *Marine drugs* 8, 6 (2010), 1838–1907.
- [25] Casabianca, S., Cornetti, L., Capellacci, S., Vernesi, C., and Penna, A. Genome complexity of harmful microalgae. *Harmful algae* 63 (2017), 7–12.
- [26] Cavender-Bares, J., Kozak, K. H., Fine, P. V., and Kembel, S. W. The merging of community ecology and phylogenetic biology. *Ecology letters* 12, 7 (2009), 693–715.
- [27] Cerveau, N., and Jackson, D. J. Combining independent de novo assemblies optimizes the coding transcriptome for nonconventional model eukaryotic organisms. *BMC bioinformatics* 17, 1 (2016), 525.
- [28] Chan, W. H., Mak, Y. L., Wu, J. J., Jin, L., Sit, W. H., Lam, J. C. W., Sadovy de Mitcheson, Y., Chan, L. L., Lam, P. K. S., and Murphy, M. B. Spatial distribution of ciguateric fish in the Republic of Kiribati. *Chemosphere* 84, 1 (2011), 117–123.
- [29] Chinain, M., Darius, H. T., Ung, A., Cruchet, P., Wang, Z., Ponton, D., Laurent, D., and Pauillac, S. Growth and toxin production in the ciguatera-causing dinoflagellate *Gambierdiscus polynesiensis* (Dinophyceae) in culture. *Toxicon* 56, 5 (2010), 739–750.
- [30] Chinain, M., Darius, H. T., Ung, A., Fouc, M. T., Revel, T., Cruchet, P., Pauillac, S., and Laurent, D. Ciguatera risk management in French Polynesia: the case study of Raivavae Island (Australes Archipelago). *Toxicon* 56, 5 (2010), 674–690.
- [31] Chinain, M., Faust, M. A., and Pauillac, S. Morphology and molecular analyses of three toxic species of *Gambierdiscus* (Dinophyceae): *G. pacificus*, sp. nov., *G. australes*, sp. nov., and *G. polynesiensis*, sp. nov. *Journal of Phycology* 35, 6 (1999), 1282–1296.

- [32] Chinain, M., Germain, M., Deparis, X., Pauillac, S., and Legrand, A.-M. Seasonal abundance and toxicity of the dinoflagellate *Gambierdiscus* spp.(Dinophyceae), the causative agent of ciguatera in Tahiti, French Polynesia. *Marine Biology* 135, 2 (1999), 259–267.
- [33] Chinain, M., Germain, M., Sako, Y., Pauillac, S., and Legrand, A.-M. Intraspecific variation in the dinoflagellate *Gambierdiscus toxicus* (Dinophyceae). I. isozyme analysis. *Journal of Phycology* 33, 1 (1997), 36–43.
- [34] Clausing, R. J., Losen, B., Oberhaensli, F. R., Darius, H. T., Sibat, M., Hess, P., Swarzenski, P. W., Chinain, M., and Bottein, M.-Y. D. Experimental evidence of dietary ciguatoxin accumulation in an herbivorous coral reef fish. *Aquatic Toxicology* 200 (2018), 257–265.
- [35] Consortium, G. O. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic acids research* 45, D1 (2016), D331–D338.
- [36] Conte, J., Potoczniak, M. J., and Tobe, S. S. Using synthetic oligonucleotides as standards in probe-based qPCR. *Biotechniques* 64, 4 (2018), 177–179.
- [37] Coordinators, N. R. Database resources of the national center for biotechnology information. *Nucleic acids research* 45, Database issue (2017), D12.
- [38] Cruz-Rivera, E., and Villareal, T. A. Macroalgal palatability and the flux of ciguatera toxins through marine food webs. *Harmful Algae* 5, 5 (2006), 497–525.
- [39] Dai, X., Mak, Y. L., Lu, C.-K., Mei, H.-H., Wu, J. J., Lee, W. H., Chan, L. L., Lim, P. T., Mustapa, N. I., Lim, H. C., et al. Taxonomic assignment of the benthic toxigenic dinoflagellate *Gambierdiscus* sp. type 6 as *Gambierdiscus balechii* (Dinophyceae), including its distribution and ciguatoxicity. *Harmful Algae* 67 (2017), 107–118.
- [40] Darby, C. A., Stolzer, M., Ropp, P. J., Barker, D., and Durand, D. Xenolog classification. *Bioinformatics* 33, 5 (2016), 640–649.
- [41] Darius, H., Ponton, D., Revel, T., Cruchet, P., Ung, A., Tchou Fouc, M., and Chinain, M. Ciguatera risk assessment in two toxic sites of French Polynesia using the receptor-binding assay. *Toxicon* 50, 5 (2007), 612–626.
- [42] Darius, H. T., Roué, M., Sibat, M., Viallon, J., Vandersea, M. W., Tester, P. A., Litaker, R. W., Amzil, Z., Hess, P., Chinain, M., et al. *Tectus niloticus* (Tegulidae, Gastropod) as a novel vector of ciguatera poisoning: detection of Pacific ciguatoxins in toxic samples from Nuku Hiva Island (French Polynesia). *Toxins* 10, 1 (2017), 2.
- [43] Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 8 (2011), 1164–1165.

- [44] de Silva, E. Contribution a l'étude du microplancton de dakar et des régions maritimes voisines. *Bull. Inst. Fr. Afr. noire, Sér. A* 18, 2 (1956), 335–371.
- [45] Dechraoui, M.-Y., Naar, J., Pauillac, S., and Legrand, A.-M. Ciguatoxins and brevetoxins, neurotoxic polyether compounds active on sodium channels. *Toxicon* 37, 1 (1999), 125–143.
- [46] Degnan, J. H., and Rosenberg, N. A. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in ecology & evolution* 24, 6 (2009), 332–340.
- [47] Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., and Notredame, C. Nextflow enables reproducible computational workflows. *Nature biotechnology* 35, 4 (2017), 316.
- [48] Dickey, R. W., and Plakas, S. M. Ciguatera: a public health perspective. *Toxicon* 56, 2 (2010), 123–136.
- [49] Diogèned, J. The chemistry of ciguatoxins: From the first records to current challenges of monitoring programs. *Toxins and Biologically Active Compounds from Microalgae* 1 (2014), 176.
- [50] Drummond, A. J., and Suchard, M. A. Bayesian random local clocks, or one rate to rule them all. *BMC biology* 8, 1 (2010), 114.
- [51] Du, P., Hahn, M. W., and Nakhleh, L. Species tree inference under the Multispecies Coalescent on data with paralogs is accurate. *bioRxiv* (2019), 498378.
- [52] Eddy, S., and Wheeler, T. HMMER: biosequence analysis using profile hidden Markov models, 2015. hmmer.org/.
- [53] Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32, 5 (2004), 1792–1797.
- [54] Eichholz, K., Beszteri, B., and John, U. Putative monofunctional type I polyketide synthase units: a dinoflagellate-specific feature? *PLoS One* 7, 11 (2012), e48624.
- [55] Farrell, H., Murray, S. A., Zammit, A., and Edwards, A. W. Management of ciguatoxin risk in Eastern Australia. *Toxins* 9, 11 (2017), 367.
- [56] Farrell, H., Zammit, A., Harwood, D. T., McNabb, P., Shadbolt, C., Manning, J., Turahui, J. A., van den Berg, D. J., and Szabo, L. Clinical diagnosis and chemical confirmation of ciguatera fish poisoning in New South Wales, Australia. *Communicable Diseases Intelligence* 40, 1 (2016).
- [57] Faust, M. A. Observation of sand-dwelling toxic dinoflagellates (Dinophyceae) from widely differing sites, including two new species. *Journal of Phycology* 31, 6 (1995), 996–1003.

- [58] Felsenstein, J. PHYLIP (phylogeny inference package) distributed by the author. *Department of Genome Sciences, University of Washington, Seattle*, Version 3 (2005).
- [59] Fitch, W. M. Distinguishing homologous from analogous proteins. *Systematic zoology* 19, 2 (1970), 99–113.
- [60] Fitch, W. M. Homology: a personal view on some of the problems. *Trends in genetics* 16, 5 (2000), 227–231.
- [61] Fraga, S., and Rodríguez, F. Genus *Gambierdiscus* in the Canary Islands (NE Atlantic Ocean) with description of *Gambierdiscus silvae* sp. nov., a new potentially toxic epiphytic benthic dinoflagellate. *Protist* 165, 6 (2014), 839–853.
- [62] Fraga, S., Rodríguez, F., Caillaud, A., Diogène, J., Raho, N., and Zapata, M. *Gambierdiscus excentricus* sp. nov. (Dinophyceae), a benthic toxic dinoflagellate from the Canary Islands (NE Atlantic Ocean). *Harmful Algae* 11 (2011), 10–22.
- [63] Fraga, S., Rodríguez, F., Riobó, P., and Bravo, I. *Gambierdiscus balechii* sp. nov. (Dinophyceae), a new benthic toxic dinoflagellate from the Celebes Sea (SW Pacific Ocean). *Harmful Algae* (2016).
- [64] Friedman, M. A., Fernandez, M., Backer, L. C., Dickey, R. W., Bernstein, J., Schrank, K., Kibler, S., Stephan, W., Gribble, M. O., Bienfang, P., et al. An updated review of ciguatera fish poisoning: Clinical, epidemiological, environmental, and public health management. *Marine drugs* 15, 3 (2017), 72.
- [65] Friedman, M. A., Fleming, L. E., Fernandez, M., Bienfang, P., Schrank, K., Dickey, R., Bottein, M.-Y., Backer, L., Ayyar, R., Weisman, R., et al. Ciguatera fish poisoning: treatment, prevention and management. *Marine drugs* 6, 3 (2008), 456–479.
- [66] Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 23 (2012), 3150–3152.
- [67] Gabaldón, T. Large-scale assignment of orthology: back to phylogenetics? *Genome biology* 9, 10 (2008), 235.
- [68] Gavryushkina, A., Heath, T. A., Ksepka, D. T., Stadler, T., Welch, D., and Drummond, A. J. Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic biology* 66, 1 (2017), 57–73.
- [69] Gimp, G. Image manipulation program. *User Manual, Edge-Detect Filters, Sobel, The GIMP Documentation Team* 8, 2 (2008), 8–7.

- [70] Gingold, D. B., Strickland, M. J., and Hess, J. J. Ciguatera fish poisoning and climate change: analysis of national poison center data in the United States, 2001–2011. *Environ Health Perspect* 122, 6 (2014), 580–586.
- [71] Gómez, F. A list of free-living dinoflagellate species in the world's oceans. *Acta Botanica Croatica* 64, 1 (2005), 129–212.
- [72] Gómez, F. A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). *Systematics and Biodiversity* 10, 3 (2012), 267–275.
- [73] Gómez, F., Qiu, D., Lopes, R. M., and Lin, S. *Fukuyoa paulensis* gen. et sp. nov., a new genus for the globular species of the dinoflagellate *Gambierdiscus* (Dinophyceae). *PLoS One* 10, 4 (2015), e0119676.
- [74] Gornik, S. G., Ford, K. L., Mulhern, T. D., Bacic, A., McFadden, G. I., and Waller, R. F. Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. *Current Biology* 22, 24 (2012), 2303–2312.
- [75] Grzebyk, D., Berland, B., Thomassin, B. A., Bosi, C., and Arnoux, A. Ecology of ciguateric dinoflagellates in the coral reef complex of Mayotte Island (SW Indian Ocean). *Journal of Experimental Marine biology and ecology* 178, 1 (1994), 51–66.
- [76] Guindon, S., and Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* 52, 5 (2003), 696–704.
- [77] Guo, R., and Ki, J.-S. Spliced leader sequences detected in EST data of the dinoflagellates *Cochlodinium polykrikoides* and *Prorocentrum minimum*. *Algae* 26, 3 (2011), 229–235.
- [78] Haas, B., and Papanicolaou, A. TransDecoder (find coding regions within transcripts), 2016.
- [79] Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* 8, 8 (2013), 1494–1512.
- [80] Hallegraeff, G. M., Bolch, C., Hill, D., Jameson, I., LeRoi, J., McMinn, A., Murray, S., De Salas, M., Saunders, K., de Salas, M., et al. *Algae of Australia: phytoplankton of temperate coastal waters*. 2010.
- [81] Hamilton, B., Hurbungs, M., Jones, A., and Lewis, R. J. Multiple ciguatoxins present in Indian Ocean reef fish. *Toxicon* 40, 9 (2002), 1347–1353.
- [82] Hamilton, B., Hurbungs, M., Vernoux, J.-P., Jones, A., and Lewis, R. J. Isolation and characterisation of Indian Ocean ciguatoxin. *Toxicon* 40, 6 (2002), 685–693.

- [83] Hamilton, B., Whittle, N., Shaw, G., Eaglesham, G., Moore, M. R., and Lewis, R. J. Human fatality associated with Pacific ciguatoxin contaminated fish. *Toxicon* 56, 5 (2010), 668–673.
- [84] Hariganeya, N., Tanimoto, Y., Yamaguchi, H., Nishimura, T., Tawong, W., Sakanari, H., Yoshimatsu, T., Sato, S., Preston, C. M., and Adachi, M. Quantitative pcr method for enumeration of cells of cryptic species of the toxic marine dinoflagellate *Ostreopsis* spp. in coastal waters of Japan. *PloS one* 8, 3 (2013), e57627.
- [85] Harke, M. J., Juhl, A. R., Haley, S. T., Alexander, H., and Dyhrman, S. T. Conserved transcriptional responses to nutrient stress in bloom-forming algae. *Frontiers in microbiology* 8 (2017), 1279.
- [86] He, F., and Maslov, S. Pan-and core-network analysis of co-expression genes in a model plant. *Scientific reports* 6 (2016), 38956.
- [87] Heath, T. A., Hedtke, S. M., Hillis, D. M., et al. Taxon sampling and the accuracy of phylogenetic analyses. *J Syst Evol* 46, 3 (2008), 239–257.
- [88] Heberle, H., Meirelles, G., da Silva, F., Telles, G., and Minghim, R. Interactivenn: a web-based tool for the analysis of sets through venn diagrams. *BMC bioinformatics* 16 (2015), 169.
- [89] Hokama, Y., Ebesu, J., Nishimura, K., Oishi, S., Mizuo, B., Stiles, M., Sakamoto, B., Takenaka, W., and Nagai, H. Human intoxications from hawaiian reef fishes associated with diverse marine toxins. *Journal of Natural Toxins* 5, 2 (1996), 235–247.
- [90] Holland, W. C., Litaker, R. W., Tomas, C. R., Kibler, S. R., Place, A. R., Davenport, E. D., and Tester, P. A. Differences in the toxicity of six *Gambierdiscus* (Dinophyceae) species measured using an *in vitro* human erythrocyte lysis assay. *Toxicon* 65 (2013), 15–33.
- [91] Holmes, M. J. *Gambierdiscus yasumotoi* sp. nov.(Dinophyceae), a toxic benthic dinoflagellate from southeastern Asia. *Journal of Phycology* 34, 4 (1998), 661–668.
- [92] Holmes, M. J., and Lewis, R. J. Purification and characterisation of large and small maitotoxins from cultured *Gambierdiscus toxicus*. *Natural toxins* 2, 2 (1994), 64–72.
- [93] Holmes, M. J., Lewis, R. J., and Gillespie, N. C. Toxicity of Australian and French Polynesian strains of *Gambierdiscus toxicus* (Dinophyceae) grown in culture: Characterization of a new type of maitotoxin. *Toxicon* 28, 10 (1990), 1159–1172.
- [94] Holmes, M. J., Lewis, R. J., Poli, M. A., and Gillespie, N. C. Strain dependent production of ciguatoxin precursors (gambiertoxins) by *Gambierdiscus toxicus* (Dinophyceae) in culture. *Toxicon* 29, 6 (1991), 761–775.

- [95] Honaas, L. A., Wafula, E. K., Wickett, N. J., Der, J. P., Zhang, Y., Edger, P. P., Altman, N. S., Pires, J. C., Leebens-Mack, J. H., et al. Selecting superior de novo transcriptome assemblies: lessons learned by leveraging the best plant genome. *PLoS One* 11, 1 (2016), e0146062.
- [96] Hoppenrath, M. Dinoflagellate taxonomy—a review and proposal of a revised classification. *Marine Biodiversity* 47, 2 (2017), 381–403.
- [97] Hoppenrath, M., Chomérat, N., Horiguchi, T., Schweikert, M., Nagahama, Y., and Murray, S. Taxonomy and phylogeny of the benthic *Prorocentrum* species (Dinophyceae) — a proposal and review. *Harmful Algae* 27 (2013), 1–28.
- [98] Hoppenrath, M., and Leander, B. S. Dinoflagellate phylogeny as inferred from heat shock protein 90 and ribosomal gene sequences. *PloS one* 5, 10 (2010), e13220.
- [99] Hoppenrath, M., Murray, S. A., Chomérat, N., and Horiguchi, T. Marine benthic dinoflagellates-unveiling their worldwide biodiversity.
- [100] Horiguchi, E. M. H. S. A. M. N. C. T. *Marine benthic dinoflagellates - unveiling their worldwide biodiversity*. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart, 07 2014.
- [101] Hou, Y., and Lin, S. Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: gene content estimation for dinoflagellate genomes. *PLoS One* 4, 9 (2009), e6978.
- [102] Huson, D. H., Richter, D. C., Rausch, C., DeZulian, T., Franz, M., and Rupp, R. Dendroscope: An interactive viewer for large phylogenetic trees. *BMC bioinformatics* 8, 1 (2007), 460.
- [103] Igarashi, T., Aritake, S., and Yasumoto, T. Mechanisms underlying the hemolytic and ichthyotoxic activities of maitotoxin. *Natural toxins* 7, 2 (1999), 71–79.
- [104] Ikehara, T., Kuniyoshi, K., Oshiro, N., and Yasumoto, T. Biooxidation of ciguatoxins leads to species-specific toxin profiles. *Toxins* 9, 7 (2017), 205.
- [105] Intergovernmental Oceanographic Commission of UNESCO. *Global Ciguatera Strategy*, 2015 (accessed July 29, 2018). http://hab.ioc-unesco.org/index.php?option=com_content&view=article&id=47:ciguatera&catid=29:activities.
- [106] Irwin, N. A., Martin, B. J., Young, B. P., Browne, M. J., Flaus, A., Loewen, C. J., Keeling, P. J., and Howe, L. J. Viral proteins as a potential driver of histone depletion in dinoflagellates. *Nature communications* 9, 1 (2018), 1535.
- [107] Ito, E., Suzuki-Toyota, F., Toshimori, K., Fuwa, H., Tachibana, K., Satake, M., and Sasaki, M. Pathological effects on mice by gambierol, possibly one of the ciguatera toxins. *Toxicon* 42, 7 (2003), 733–740.

- [108] Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. Phylogenomics: the beginning of incongruence? *TRENDS in Genetics* 22, 4 (2006), 225–231.
- [109] Jin, M., Liu, H., He, C., Fu, J., Xiao, Y., Wang, Y., Xie, W., Wang, G., and Yan, J. Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. *Scientific reports* 6 (2016), 18936.
- [110] Johnson, L. K., Alexander, H., and Brown, C. T. Re-assembly, quality evaluation, and annotation of 678 microbial eukaryotic reference transcriptomes.
- [111] Kahlke, T. GOSUM: Gene Ontology Summarizer version 0.1, 2018. <http://doi.org/10.5281/zenodo.1344306>.
- [112] Kato, K. H., Moriyama, A., Huitorel, P., Cosson, J., Cachon, M., and Sato, H. Isolation of the major basic nuclear protein and its localization on chromosomes of the dinoflagellate, *Oxyrrhis marina*. *Biology of the Cell* 89, 1 (1997), 43–52.
- [113] Kears, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 12 (2012), 1647–1649.
- [114] Keeling, P. J., Burki, F., Wilcox, H. M., Allam, B., Allen, E. E., Amaral-Zettler, L. A., Armbrust, E. V., Archibald, J. M., Bharti, A. K., Bell, C. J., et al. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PloS one* (2014).
- [115] Kibler, S. R., Litaker, R. W., Holland, W. C., Vandersea, M. W., and Tester, P. A. Growth of eight *Gambierdiscus* (Dinophyceae) species: Effects of temperature, salinity and irradiance. *Harmful Algae* 19 (2012), 1–14.
- [116] Kimura, K., Okuda, S., Nakayama, K., Shikata, T., Takahashi, F., Yamaguchi, H., Skamoto, S., Yamaguchi, M., and Tomaru, Y. Rna sequencing revealed numerous polyketide synthase genes in the harmful dinoflagellate *Karenia mikimotoi*. *PloS one* 10, 11 (2015), e0142731.
- [117] Kohli, G. S., Campbell, K., John, U., Smith, K. F., Fraga, S., Rhodes, L. L., and Murray, S. A. Role of modular polyketide synthases in the production of polyether ladder compounds in ciguatoxin-producing *Gambierdiscus polynesiensis* and *G. excentricus* (Dinophyceae). *Journal of Eukaryotic Microbiology* 64, 5 (2017), 691–706.
- [118] Kohli, G. S., Haslauer, K., Sarowar, C., Kretzschmar, A. L., Boulter, M., Harwood, D. T., Laczka, O., and Murray, S. A. Qualitative and quantitative assessment of the presence of ciguatoxin, P-CTX-1B, in Spanish Mackerel (*Scomberomorus commerson*) from waters in New South Wales (Australia). *Toxicology Reports* (2017).

- [119] Kohli, G. S., John, U., Figueroa, R. I., Rhodes, L. L., Harwood, D. T., Groth, M., Bolch, C. J., and Murray, S. A. Polyketide synthesis genes associated with toxin production in two species of *Gambierdiscus* (Dinophyceae). *BMC genomics* 16, 1 (2015), 410.
- [120] Kohli, G. S., John, U., Van Dolah, F. M., and Murray, S. A. Evolutionary distinctiveness of fatty acid and polyketide synthesis in eukaryotes. *The ISME journal* 10, 8 (2016), 1877.
- [121] Kohli, G. S., Murray, S. A., Neilan, B. A., Rhodes, L. L., Harwood, D. T., Smith, K. F., Meyer, L., Capper, A., Brett, S., and Hallegraeff, G. M. High abundance of the potentially maitotoxic dinoflagellate *Gambierdiscus carpenteri* in temperate waters of New South Wales, Australia. *Harmful Algae* 39 (2014), 134–145.
- [122] Kohli, G. S., Neilan, B. A., Brown, M. V., Hoppenrath, M., and Murray, S. A. Cob gene pyrosequencing enables characterization of benthic dinoflagellate diversity and biogeography. *Environmental microbiology* 16, 2 (2014), 467–485.
- [123] Kohli, G. S., Papiol, G. G., Rhodes, L. L., Harwood, D. T., Selwood, A., Jerrett, A., Murray, S. A., and Neilan, B. A. A feeding study to probe the uptake of maitotoxin by snapper (*Pagrus auratus*). *Harmful Algae* 37 (2014), 125–132.
- [124] Koid, A. E., Liu, Z., Terrado, R., Jones, A. C., Caron, D. A., and Heidelberg, K. B. Comparative transcriptome analysis of four prymnesiophyte algae. *PLoS One* 9, 6 (2014), e97801.
- [125] Kon, N. F., Teng, S. T., Hii, K. S., Yek, L. H., Mujahid, A., Lim, H. C., Lim, P. T., and Leaw, C. P. Spatial distribution of toxic *Alexandrium tamiyavanichii* (Dinophyceae) in the south-eastern South China Sea-Sulu Sea: A molecular-based assessment using real-time quantitative PCR (qPCR) assay. *Harmful Algae* 50 (2015), 8–20.
- [126] Koonin, E. V. Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* 39 (2005), 309–338.
- [127] Kraeer, K., and van Essen-Fishman, L. *Pseudocaranx dentex* (white trevally). ian.umces.edu/imagelibrary/, 2010. Integration and Application Network, University of Maryland Center for Environmental Science.
- [128] Kretzschmar, A. L., Verma, A., Harwood, T., Hoppenrath, M., and Murray, S. Characterization of *Gambierdiscus lapillus* sp. nov. (Gonyaulacales, Dinophyceae): A new toxic dinoflagellate from the Great Barrier Reef (Australia). *Journal of phycology* 53, 2 (2017), 283–297.
- [129] Kubatko, L. S., and Degnan, J. H. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* 56, 1 (2007), 17–24.

- [130] Kuno, S., Kamikawa, R., Yoshimatsu, S., Sagara, T., Nishio, S., and Sako, Y. Genetic diversity of *Gambierdiscus* spp.(Gonyaulacales, Dinophyceae) in Japanese coastal areas. *Phycological research* 58, 1 (2010), 44–52.
- [131] LaJeunesse, T. C., Lambert, G., Andersen, R. A., Coffroth, M. A., and Galbraith, D. W. *Symbiodinium* (Pyrrophyta) genome sizes (DNA content) are smallest among dinoflagellates. *Journal of Phycology* 41, 4 (2005), 880–886.
- [132] LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., and Santos, S. R. Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Current Biology* 28, 16 (2018), 2570–2580.
- [133] Lapierre, P., and Gogarten, J. P. Estimating the size of the bacterial pan-genome. *Trends in genetics* 25, 3 (2009), 107–110.
- [134] Larsson, M. E., Laczka, O. F., Harwood, D. T., Lewis, R. J., Himaya, S., Murray, S. A., and Doblin, M. A. Toxicology of *Gambierdiscus* spp.(dinophyceae) from Tropical and Temperate Australian Waters. *Marine drugs* 16, 1 (2018), 7.
- [135] Lauritano, C., De Luca, D., Ferrarini, A., Avanzato, C., Minio, A., Esposito, F., and Ianora, A. De novo transcriptome of the cosmopolitan dinoflagellate *Amphidinium carterae* to identify enzymes with biotechnological potential. *Scientific reports* 7, 1 (2017), 11701.
- [136] Laza-Martínez, A., David, H., Riobó, P., Miguel, I., and Orive, E. Characterization of a strain of *Fukuyoa paulensis* (Dinophyceae) from the Western Mediterranean Sea. *Journal of Eukaryotic Microbiology* (2016).
- [137] Legrand, A., Teai, T., Crochet, P., Satake, M., Murata, K., and Yasumoto, T. Two structural types of ciguatoxins involved on ciguatera fish poisoning in French Polynesia. In *VIII International conference on Harmful algae- Abstracts and Posters Classification*. vp. 1997. (1997).
- [138] Lehane, L., and Lewis, R. J. Ciguatera: recent advances but the risk remains. *International journal of food microbiology* 61, 2 (2000), 91–125.
- [139] Lemmon, A. R., and Moriarty, E. C. The importance of proper model assumption in Bayesian phylogenetics. *Systematic Biology* 53, 2 (2004), 265–277.
- [140] Lewis, F., Hughes, G. J., Rambaut, A., Pozniak, A., and Brown, A. J. L. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS medicine* 5, 3 (2008), e50.
- [141] Lewis, R. Socioeconomic impacts and management ciguatera in the Pacific. *Bulletin de la Societe de pathologie exotique* (1990) 85, 5 Pt 2 (1992), 427–434.

- [142] Lewis, R., Molgo, J., and Adams, D. J. Ciguatera toxins: Pharmacology of toxins involved in ciguatera and related fish poisonings. *Seafood and freshwater toxins: pharmacology, physiology, and detection* (2000), 419.
- [143] Lewis, R. J. The changing face of ciguatera. *Toxicon* 39, 1 (2001), 97–106.
- [144] Lewis, R. J. Ciguatera: Australian perspectives on a global problem. *Toxicon* 48, 7 (2006), 799–809.
- [145] Lewis, R. J., and Holmes, M. J. Origin and transfer of toxins involved in ciguatera. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology* 106, 3 (1993), 615–628.
- [146] Lewis, R. J., Inserra, M., Vetter, I., Holland, W. C., Hardison, D. R., Tester, P. A., and Litaker, R. W. Rapid extraction and identification of maitotoxin and ciguatoxin-like toxins from Caribbean and Pacific *Gambierdiscus* using a new functional bioassay. *PLoS One* 11, 7 (2016), e0160006.
- [147] Lewis, R. J., Sellin, M., Poli, M. A., Norton, R. S., MacLeod, J. K., and Sheil, M. M. Purification and characterization of ciguatoxins from moray eel (*Lycodontis javanicus*, Muraenidae). *Toxicon* 29, 9 (1991), 1115–1127.
- [148] Lewis, R. J., Vernoux, J.-P., and Brereton, I. M. Structure of Caribbean ciguatoxin isolated from *Caranx latus*. *Journal of the American Chemical Society* 120, 24 (1998), 5914–5920.
- [149] Lewis, R. J., Wong Hoy, A. W., and McGiffin, D. C. Action of ciguatoxin on human atrial trabeculae. *Toxicon* 30, 8 (1992), 907–914.
- [150] Li, Y.-h., Zhou, G., Ma, J., Jiang, W., Jin, L.-g., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L., et al. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology* 32, 10 (2014), 1045.
- [151] Lidie, K. B., and Van Dolah, F. M. Spliced leader RNA-mediated trans-splicing in a dinoflagellate, *Karenia brevis*. *Journal of Eukaryotic Microbiology* 54, 5 (2007), 427–435.
- [152] Lin, S. Genomic understanding of dinoflagellates. *Research in microbiology* 162, 6 (2011), 551–569.
- [153] Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., Li, L., Zhang, Y., Zhang, H., Ji, Z., et al. The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* 350, 6261 (2015), 691–694.
- [154] Lin, S., and Zhang, H. Dinoflagellate meta-transcriptomics enabled by spliced leader. In *13th International Conference on Harmful Algae. Hong Kong* (2010), pp. 166–170.

- [155] Lin, S., Zhang, H., Zhuang, Y., Tran, B., and Gill, J. Spliced leader - based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. *Proceedings of the National Academy of Sciences* 107, 46 (2010), 20033–20038.
- [156] Lindsay, J. A. Chronic sequelae of foodborne disease. *Emerging infectious diseases* 3, 4 (1997), 443.
- [157] Litaker, R. W., Vandersea, M. W., Faust, M. A., Kibler, S. R., Chinain, M., Holmes, M. J., Holland, W. C., and Tester, P. A. Taxonomy of *Gambierdiscus* including four new species, *Gambierdiscus caribaeus*, *Gambierdiscus carolinianus*, *Gambierdiscus carpenteri* and *Gambierdiscus ruetzleri* (Gonyaulacales, Dinophyceae). *Phycologia* 48, 5 (2009).
- [158] Litaker, R. W., Vandersea, M. W., Faust, M. A., Kibler, S. R., Nau, A. W., Holland, W. C., Chinain, M., Holmes, M. J., and Tester, P. A. Global distribution of ciguatera causing dinoflagellates in the genus *Gambierdiscus*. *Toxicon* 56, 5 (2010), 711–730.
- [159] Litaker, W., Steidinger, K. A., Mason, P. L., Landsberg, J. H., Shields, J. D., Reece, K. S., Haas, L. W., Vogelbein, W. K., Vandersea, M. W., Kibler, S. R., et al. The reclassification of *Pfiesteria shumwayae* (Dinophyceae): *Pseudopfiesteria*, gen. nov. *Journal of Phycology* 41, 3 (2005), 643–651.
- [160] Liu, L., Xi, Z., and Davis, C. C. Coalescent methods are robust to the simultaneous effects of long branches and incomplete lineage sorting. *Molecular biology and evolution* 32, 3 (2014), 791–805.
- [161] Lobel, P. S., Anderson, D. M., and Durand-Clement, M. Assessment of ciguatera dinoflagellate populations: sample variability and algal substrate selection. *The Biological Bulletin* 175, 1 (1988), 94–101.
- [162] Loeffler, C. R., Richlen, M. L., Brandt, M. E., and Smith, T. B. Effects of grazing, nutrients, and depth on the ciguatera-causing dinoflagellate *Gambierdiscus* in the US Virgin Islands. *Marine Ecology Progress Series* 531 (2015), 91–104.
- [163] MacManes, M. On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in Genetics* 5 (2014).
- [164] Maddison, W. P., and Knowles, L. L. Inferring phylogeny despite incomplete lineage sorting. *Systematic biology* 55, 1 (2006), 21–30.
- [165] Mak, Y. L., Wai, T.-C., Murphy, M. B., Chan, W. H., Wu, J. J., Lam, J. C., Chan, L. L., and Lam, P. K. Pacific ciguatoxins in food web components of coral reef systems in the Republic of Kiribati. *Environmental science & technology* 47, 24 (2013), 14070–14079.

- [166] Mattei, C., Dechraoui, M.-Y., Molgo, J., Meunier, F. A., Legrand, A.-M., and Benoit, E. Neurotoxins targetting receptor site 5 of voltage-dependent sodium channels increase the nodal volume of myelinated axons. *Journal of neuroscience research* 55, 6 (1999), 666–673.
- [167] McKinney, W. Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference* (2010), S. van der Walt and J. Millman, Eds., pp. 51–56.
- [168] McTavish, E. J., Drew, B. T., Redelings, B., and Cranston, K. A. How and why to build a unified Tree of Life. *BioEssays* 39, 11 (2017).
- [169] Mendes, F. K., and Hahn, M. W. Why concatenation fails near the anomaly zone. *Systematic biology* 67, 1 (2017), 158–169.
- [170] Meyer, J. M., Rödelberger, C., Eichholz, K., Tillmann, U., Cembella, A., McGaughran, A., and John, U. Transcriptomic characterisation and genomic glimps into the toxigenic dinoflagellate *Azadinium spinosum*, with emphasis on polyketide synthase genes. *BMC genomics* 16, 1 (2015), 27.
- [171] Meyer, J. M., Rödelberger, C., Eichholz, K., Tillmann, U., Cembella, A., McGaughran, A., and John, U. Transcriptomic characterisation and genomic glimps into the toxigenic dinoflagellate *Azadinium spinosum*, with emphasis on polyketide synthase genes. *BMC genomics* 16, 1 (2015), 27.
- [172] Minor, T., Lasher, A., Klontz, K., Brown, B., Nardinelli, C., and Zorn, D. The Per Case and Total Annual Costs of Foodborne Illness in the United States. *Risk Analysis* (2014).
- [173] Moldowan, J. M., and Talyzina, N. M. Biogeochemical evidence for dinoflagellate ancestors in the Early Cambrian. *Science* 281, 5380 (1998), 1168–1170.
- [174] Molgo, J., and Adams, D. J. Ciguatera toxins: pharmacology of toxins involved in ciguatera and related fish poisonings. *Seafood and freshwater toxins: pharmacology, physiology, and detection* (2000), 419.
- [175] Monroe, E. A., Johnson, J. G., Wang, Z., Pierce, R. K., and Van Dolah, F. M. Characterization and expression of nuclear-encoded polyketide synthases in the brevetoxin-producing dinoflagellate *Karenia brevis*. *Journal of Phycology* 46, 3 (2010), 541–552.
- [176] Monroe, E. A., and Van Dolah, F. M. The toxic dinoflagellate *Karenia brevis* encodes novel type I-like polyketide synthases containing discrete catalytic domains. *Protist* 159, 3 (2008), 471–482.
- [177] Morey, J. S., and Van Dolah, F. M. Global analysis of mRNA half-lives and de novo transcription in a dinoflagellate, *Karenia brevis*. *PLoS One* 8, 6 (2013), e66347.

- [178] Munday, R., Murray, S., Rhodes, L. L., Larsson, M. E., and Harwood, D. T. Ciguatoxins and maitotoxins in extracts of sixteen gambierdiscus isolates and one fukuyoa isolate from the south pacific and their toxicity to mice by intraperitoneal and oral administration. *Marine drugs* 15, 7 (2017), 208.
- [179] Murata, M., Legrand, A. M., Ishibashi, Y., Fukui, M., and Yasumoto, T. Structures and configurations of ciguatoxin from the moray eel *Gymnothorax javanicus* and its likely precursor from the dinoflagellate *Gambierdiscus toxicus*. *Journal of the American Chemical Society* 112, 11 (1990), 4380–4386.
- [180] Murata, M., Legrand, A. M., Ishibashi, Y., and Yasumoto, T. Structures of ciguatoxin and its congener. *Journal of the American chemical Society* 111, 24 (1989), 8929–8931.
- [181] Murata, M., Naoki, H., Iwashita, T., Matsunaga, S., Sasaki, M., Yokoyama, A., and Yasumoto, T. Structure of maitotoxin. *Journal of the American Chemical Society* 115, 5 (1993), 2060–2062.
- [182] Murata, M., Naoki, H., Matsunaga, S., Satake, M., and Yasumoto, T. Structure and partial stereochemical assignments for maitotoxin, the most toxic and largest natural non-biopolymer. *Journal of the American Chemical Society* 116, 16 (1994), 7098–7107.
- [183] Murray, S., Jørgensen, M. F., Ho, S. Y., Patterson, D. J., and Jermini, L. S. Improving the analysis of dinoflagellate phylogeny based on rDNA. *Protist* 156, 3 (2005), 269–286.
- [184] Murray, S., Momigliano, P., Heimann, K., and Blair, D. Molecular phylogenetics and morphology of *Gambierdiscus yasumotoi* from tropical eastern Australia. *Harmful Algae* 39 (2014), 242–252.
- [185] Murray, S., Momigliano, P., Heimann, K., and Blair, D. Molecular phylogenetics and morphology of *Gambierdiscus yasumotoi* from tropical eastern Australia. *Harmful Algae* 39 (2014), 242–252.
- [186] Murray, S. A., Diwan, R., Orr, R. J., Kohli, G. S., and John, U. Gene duplication, loss and selection in the evolution of saxitoxin biosynthesis in alveolates. *Molecular phylogenetics and evolution* 92 (2015), 165–180.
- [187] Murray, S. A., Garby, T., Hoppenrath, M., and Neilan, B. A. Genetic diversity, morphological uniformity and polyketide production in dinoflagellates (*Amphidinium*, Dinoflagellata). *PloS one* 7, 6 (2012), e38253.
- [188] Murray, S. A., Suggett, D. J., Doblin, M. A., Kohli, G. S., Seymour, J. R., Fabris, M., and Ralph, P. J. Unravelling the functional genetics of dinoflagellates: a review of approaches and opportunities. *Perspect. Phycol* 3, 1 (2016), 37–52.

- [189] Mutreja, A., Kim, D. W., Thomson, N. R., Connor, T. R., Lee, J. H., Kariuki, S., Croucher, N. J., Choi, S. Y., Harris, S. R., Lebens, M., et al. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* 477, 7365 (2011), 462.
- [190] Nagai, H., Murata, M., Torigoe, K., Satake, M., and Yasumoto, T. Gambieric acids, new potent antifungal substances with unprecedented polyether structures from a marine dinoflagellate *Gambierdiscus toxicus*. *The Journal of Organic Chemistry* 57, 20 (1992), 5448–5453.
- [191] Nakajima, I., Oshima, Y., and Yasumoto, T. Toxicity of benthic dinoflagellates in Okinawa (Japan). *Bulletin of the Japanese Society of Scientific Fisheries (Japan)* (1981).
- [192] Nishimura, T., Hariganeya, N., Tawong, W., Sakanari, H., Yamaguchi, H., and Adachi, M. Quantitative pcr assay for detection and enumeration of ciguatera-causing dinoflagellate *Gambierdiscus* spp.(gonyaulacales) in coastal areas of japan. *Harmful Algae* 52 (2016), 11–22.
- [193] Nishimura, T., Sato, S., Tawong, W., Sakanari, H., Uehara, K., Shah, M. M. R., Suda, S., Yasumoto, T., Taira, Y., Yamaguchi, H., et al. Genetic diversity and distribution of the ciguatera-causing dinoflagellate *Gambierdiscus* spp.(Dinophyceae) in coastal areas of Japan. *PloS one* 8, 4 (2013), e60882.
- [194] Nishimura, T., Sato, S., Tawong, W., Sakanari, H., Yamaguchi, H., and Adachi, M. Morphology of *Gambierdiscus scabrosus* sp. nov.(Gonyaulacales): a new epiphytic toxic dinoflagellate from coastal areas of Japan. *Journal of Phycology* 50, 3 (2014), 506–514.
- [195] Nonomura, T., Sasaki, M., Matsumori, N., Murata, M., Tachibana, K., and Yasumoto, T. The complete structure of maitotoxin, part II: Configuration of the C135 C142 side chain and absolute configuration of the entire molecule. *Angewandte Chemie International Edition in English* 35, 15 (1996), 1675–1678.
- [196] Parsons, M. L., Aligizaki, K., Bottein, M.-Y. D., Fraga, S., Morton, S. L., Penna, A., and Rhodes, L. *Gambierdiscus* and *Ostreopsis*: Reassessment of the state of knowledge of their taxonomy, geography, ecophysiology, and toxicology. *Harmful algae* 14 (2012), 107–129.
- [197] Parsons, M. L., Settlemier, C. J., and Ballauer, J. M. An examination of the epiphytic nature of *Gambierdiscus toxicus*, a dinoflagellate involved in ciguatera fish poisoning. *Harmful algae* 10, 6 (2011), 598–605.
- [198] Pawlowicz, R., Morey, J., Darius, H., Chinain, M., and Van Dolah, F. Transcriptome sequencing reveals single domain Type I-like polyketide synthases in the toxic dinoflagellate *Gambierdiscus polynesiensis*. *Harmful Algae* 36 (2014), 29–37.

- [199] Philippe, H., Snell, E. A., Baptiste, E., Lopez, P., Holland, P. W., and Casane, D. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Molecular biology and evolution* 21, 9 (2004), 1740–1752.
- [200] Pisapia, F., Sibat, M., Herrenknecht, C., Lhaute, K., Gaiani, G., Ferron, P.-J., Fessard, V., Fraga, S., Nascimento, S. M., Litaker, R. W., et al. Maitotoxin-4, a novel MTX analog produced by *Gambierdiscus excentricus*. *Marine drugs* 15, 7 (2017), 220.
- [201] Plissonneau, C., Hartmann, F. E., and Croll, D. Pangenome analyses of the wheat pathogen *Zymoseptoria tritici* reveal the structural basis of a highly plastic eukaryotic genome. *BMC biology* 16, 1 (2018), 5.
- [202] Posnien, N., Zeng, V., Schwager, E. E., Pechmann, M., Hilbrant, M., Keefe, J. D., Damen, W. G., Prpic, N.-M., McGregor, A. P., and Extavour, C. G. A comprehensive reference transcriptome resource for the common house spider *Parasteatoda tepidariorum*. *PLoS One* 9, 8 (2014), e104885.
- [203] Pottier, I., Hamilton, B., Jones, A., Lewis, R. J., and Vernoux, J. P. Identification of slow and fast-acting toxins in a highly ciguatoxic barracuda (*Sphyrna barracuda*) by HPLC/MS and radiolabelled ligand binding. *Toxicon* 42, 6 (2003), 663–672.
- [204] Pottier, I., Vernoux, J.-P., Jones, A., and Lewis, R. J. Characterisation of multiple Caribbean ciguatoxins and congeners in individual specimens of horse-eye jack (*Caranx latus*) by high-performance liquid chromatography/mass spectrometry. *Toxicon* 40, 7 (2002), 929–939.
- [205] Price, D. C., and Bhattacharya, D. Robust dinoflagellata phylogeny inferred from public transcriptome databases. *Journal of Phycology* (2017).
- [206] Price, D. C., Farinholt, N., Gates, C., Shumaker, A., Wagner, N. E., Bienfang, P., and Bhattacharya, D. Analysis of *Gambierdiscus* transcriptome data supports ancient origins of mixotrophic pathways in dinoflagellates. *Environmental Microbiology* (2016).
- [207] Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F. O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41, D1 (2013), D590–D596.
- [208] Queensland Government, Queensland Health. *Notifiable conditions annual reporting*, 2016 (accessed December 30, 2016). <https://www.health.qld.gov.au/clinical-practice/guidelines-procedures/diseases-infection/surveillance/reports/notifiable/annual>.
- [209] Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. InterProScan: protein domains identifier. *Nucleic acids research* 33, suppl_2 (2005), W116–W120.

- [210] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [211] Rae, P. Hydroxymethyluracil in eukaryote DNA: a natural feature of the pyrophyta (Dinoflagellates). *Science* 194, 4269 (1976), 1062–1064.
- [212] Rae, P. M., and Steele, R. E. Modified bases in the DNAs of unicellular eukaryotes: an examination of distributions and possible roles, with emphasis on hydroxymethyluracil in dinoflagellates. *Biosystems* 10, 1-2 (1978), 37–53.
- [213] Rains, L. K., and Parsons, M. L. *Gambierdiscus* species exhibit different epiphytic behaviors toward a variety of macroalgal hosts. *Harmful algae* 49 (2015), 29–39.
- [214] Read, B. A., Kegel, J., Klute, M. J., Kuo, A., Lefebvre, S. C., Maumus, F., Mayer, C., Miller, J., Monier, A., Salamov, A., et al. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* 499, 7457 (2013), 209.
- [215] Rein, K. S., and Borrone, J. Polyketides from dinoflagellates: origins, pharmacology and biosynthesis. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 124, 2 (1999), 117–131.
- [216] Rhodes, L., Giménez Papiol, G., Smith, K., and Harwood, T. *Gambierdiscus* cf. *yasumotoi* (Dinophyceae) isolated from New Zealand’s sub-tropical northern coastal waters. *New Zealand Journal of Marine and Freshwater Research*, ahead-of-print (2014), 1–8.
- [217] Rhodes, L., Harwood, T., Smith, K., Argyle, P., and Munday, R. Production of ciguatoxin and maitotoxin by strains of *Gambierdiscus australes*, *G. pacificus* and *G. polynesiensis* (Dinophyceae) isolated from Rarotonga, Cook Islands. *Harmful Algae* 39 (2014), 185–190.
- [218] Rhodes, L., Smith, K. F., Verma, A., Curley, B. G., Harwood, D. T., Murray, S., Kohli, G. S., Solomona, D., Rongo, T., Munday, R., and Murray, S. A. A new species of *Gambierdiscus* (Dinophyceae) from the south-west Pacific: *Gambierdiscus honu* sp. nov. *Harmful Algae* 65 (2017), 61–70.
- [219] Rhodes, L. L., Smith, K. F., Munday, R., Selwood, A. I., McNabb, P. S., Holland, P. T., and Bottein, M.-Y. Toxic dinoflagellates (Dinophyceae) from Rarotonga, Cook Islands. *Toxicon* 56, 5 (2010), 751–758.
- [220] Rhodes, L. L., Smith, K. F., Murray, S., Harwood, D. T., Trnski, T., and Munday, R. The epiphytic genus *Gambierdiscus* (Dinophyceae) in the Kermadec Islands and Zealandia regions of the southwestern Pacific and the associated risk of ciguatera fish poisoning. *Marine drugs* 15, 7 (2017), 219.

- [221] Richlen, M. L., Morton, S. L., Barber, P. H., and Lobel, P. S. Phylogeography, morphological variation and taxonomy of the toxic dinoflagellate *Gambierdiscus toxicus* (Dinophyceae). *Harmful Algae* 7, 5 (2008), 614–629.
- [222] Ridgway, K. Long-term trend and decadal variability of the southward penetration of the East Australian Current. *Geophysical Research Letters* 34, 13 (2007).
- [223] Rio, D. C., Ares, M., Hannon, G. J., and Nilsen, T. W. Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harbor Protocols* 2010, 6 (2010), pdb-prot5439.
- [224] Rizzo, P. J. The enigma of the dinoflagellate chromosome. *The Journal of protozoology* 38, 3 (1991), 246–252.
- [225] Roch, S., and Steel, M. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theoretical population biology* 100 (2015), 56–62.
- [226] Rodríguez, I., Genta-Jouve, G., Alfonso, C., Calabro, K., Alonso, E., Sánchez, J. A., Alfonso, A., Thomas, O. P., and Botana, L. M. Gambierone, a ladder-shaped polyether from the dinoflagellate *Gambierdiscus belizeanus*. *Organic letters* 17, 10 (2015), 2392–2395.
- [227] Rongo, T., and van Woesik, R. Ciguatera poisoning in Rarotonga, southern Cook Islands. *Harmful Algae* 10, 4 (2011), 345–355.
- [228] Rongo, T., and van Woesik, R. The effects of natural disturbances, reef state, and herbivorous fish densities on ciguatera poisoning in Rarotonga, southern Cook Islands. *Toxicon* 64 (2013), 87–95.
- [229] Ronquist, F., and Huelsenbeck, J. P. MrBayes 3- Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 12 (2003), 1572–1574.
- [230] Roué, M., Darius, H. T., Picot, S., Ung, A., Viallon, J., Gaertner-Mazouni, N., Sibat, M., Amzil, Z., and Chinain, M. Evidence of the bioaccumulation of ciguatoxins in giant clams (*Tridacna maxima*) exposed to *Gambierdiscus* spp. cells. *Harmful Algae* 57 (2016), 78–87.
- [231] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2015.
- [232] Ryan, D. E., Pepper, A. E., and Campbell, L. De novo assembly and characterization of the transcriptome of the toxic dinoflagellate *Karenia brevis*. *BMC genomics* 15, 1 (2014), 888.
- [233] Salcedo, T., Upadhyay, R. J., Nagasaki, K., and Bhattacharya, D. Dozens of toxin-related genes are expressed in a nontoxic strain of the dinoflagellate *Heterocapsa circularisquama*. *Molecular biology and evolution* 29, 6 (2012), 1503–1506.

- [234] Saldarriaga, J. F., Cavalier-Smith, T., Menden-Deuer, S., Keeling, P. J., et al. Molecular data and the evolutionary history of dinoflagellates. *European journal of protistology* 40, 1 (2004), 85–111.
- [235] Satake, M., Ishida, S., Yasumoto, T., Murata, M., Utsumi, H., and Hinomoto, T. Structural confirmation of maitotoxin based on complete ¹³C NMR assignments and the three-dimensional PFG NOESY-HMQC spectrum. *Journal of the American Chemical Society* 117, 26 (1995), 7019–7020.
- [236] Satake, M., Murata, M., and Yasumoto, T. Gambierol: a new toxic polyether compound isolated from the marine dinoflagellate *Gambierdiscus toxicus*. *Journal of the American Chemical Society* 115, 1 (1993), 361–362.
- [237] Satake, M., Murata, M., and Yasumoto, T. The structure of CTX3C, a ciguatoxin congener isolated from cultured *Gambierdiscus toxicus*. *Tetrahedron letters* 34, 12 (1993), 1975–1978.
- [238] Saxby, T. *Scomberomorus munroi* (Australian spotted mackerel). ian.umces.edu/imagegallery/, 2004. Integration and Application Network, University of Maryland Center for Environmental Science.
- [239] SHALCHIAN-TABRIZI, K., MINGE, M. A., CAVALIER-SMITH, T., NEDREKLEPP, J. M., Klaveness, D., and JAKOBSEN, K. S. Combined heat shock protein 90 and ribosomal RNA sequence phylogeny supports multiple replacements of dinoflagellate plastids. *Journal of Eukaryotic Microbiology* 53, 3 (2006), 217–224.
- [240] Shoguchi, E., Beedessee, G., Tada, I., Hisata, K., Kawashima, T., Takeuchi, T., Arakaki, N., Fujie, M., Koyanagi, R., Roy, M. C., et al. Two divergent *Symbiodinium* genomes reveal conservation of a gene cluster for sunscreen biosynthesis and recently lost genes. *BMC genomics* 19, 1 (2018), 458.
- [241] Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., et al. Draft assembly of the *Symbiodinium minutum* nuclear genome reveals dinoflagellate gene structure. *Current biology* 23, 15 (2013), 1399–1408.
- [242] Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 19 (2015), 3210–3212.
- [243] Sims, J. K. A theoretical discourse on the pharmacology of toxic marine ingestions. *Annals of emergency medicine* 16, 9 (1987), 1006–1015.

- [244] Sites Jr, J. W., Reeder, T. W., and Wiens, J. J. Phylogenetic insights on evolutionary novelties in lizards and snakes: sex, birth, bodies, niches, and venom. *Annual Review of Ecology, Evolution, and Systematics* 42 (2011), 227–244.
- [245] Skinner, M. P., Brewer, T. D., Johnstone, R., Fleming, L. E., and Lewis, R. J. Ciguatera fish poisoning in the Pacific Islands (1998 to 2008). *PLoS Negl Trop Dis* 5, 12 (2011), e1416.
- [246] Slamovits, C. H., and Keeling, P. J. Widespread recycling of processed cDNAs in dinoflagellates. *Current Biology* 18, 13 (2008), R550–R552.
- [247] Smith, K. F., Biessy, L., Argyle, P. A., Trnski, T., Halafihi, T., and Rhodes, L. L. Molecular identification of *Gambierdiscus* and *Fukuyoa* (Dinophyceae) from environmental samples. *Marine Drugs* 15, 8 (2017), 243.
- [248] Smith, K. F., Rhodes, L., Verma, A., Curley, B. G., Harwood, D. T., Kohli, G. S., Solomona, D., Rongo, T., Munday, R., and Murray, S. A. A new *Gambierdiscus* species (Dinophyceae) from Rarotonga, Cook Islands: *Gambierdiscus cheloniae* sp. nov. *Harmful Algae* 60 (2016), 45–56.
- [249] Song, G., Dickins, B. J., Demeter, J., Engel, S., Dunn, B., and Cherry, J. M. AGAPE (Automated Genome Analysis PipelinE) for pan-genome analysis of *Saccharomyces cerevisiae*. *PLoS One* 10, 3 (2015), e0120671.
- [250] Sparrow, L., Momigliano, P., Russ, G. R., and Heimann, K. Effects of temperature, salinity and composition of the dinoflagellate assemblage on the growth of *Gambierdiscus carpenteri* isolated from the Great Barrier Reef. *Harmful Algae* 65 (2017), 52–60.
- [251] Sperr, A., and Doucette, G. Variation in growth rate and ciguatera toxin production among geographically distinct isolates of *Gambierdiscus toxicus*. *Harmful and Toxic Algal Blooms*. (1996), 309–312.
- [252] Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 9 (2014), 1312–1313.
- [253] Staunton, J., and Weissman, K. J. Polyketide biosynthesis: a millennium review. *Natural product reports* 18, 4 (2001), 380–416.
- [254] Stephens, T. G., Ragan, M. A., Bhattacharya, D., and Chan, C. X. Core genes in diverse dinoflagellate lineages include a wealth of conserved dark genes with unknown functions. *Scientific reports* 8, 1 (2018), 17175.
- [255] Stevens, T., and Boucher, W. *Python Programming for Biology: Bioinformatics and Beyond*. Cambridge University Press, 2018.

- [256] Suzuki, Y., Glazko, G. V., and Nei, M. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences* 99, 25 (2002), 16138–16143.
- [257] Taylor, F., and Gustavson, M. An underwater survey of the organism chiefly responsible for “ciguatera” fish poisoning in the eastern Caribbean region: the benthic dinoflagellate *Gambierdiscus toxicus*. In *Proceedings of the 7th. International Diving Science Symposium* (1986), CMAS, University of Padua, pp. 95–111.
- [258] Tester, P. A., Kibler, S. R., Holland, W. C., Usup, G., Vandersea, M. W., Leaw, C. P., Teen, L. P., Larsen, J., Mohammad-Noor, N., Faust, M. A., et al. Sampling harmful benthic dinoflagellates: Comparison of artificial and natural substrate methods. *Harmful Algae* 39 (2014), 8–25.
- [259] Tester, P. A., Vandersea, M. W., Buckel, C. A., Kibler, S. R., Holland, W. C., Davenport, E. D., Clark, R. D., Edwards, K. F., Taylor, J. C., Pluym, J. L. V., et al. *Gambierdiscus* (Dinophyceae) species diversity in the Flower Garden Banks National Marine Sanctuary, Northern Gulf of Mexico, USA. *Harmful Algae* 29 (2013), 1–9.
- [260] Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proceedings of the National Academy of Sciences* 102, 39 (2005), 13950–13955.
- [261] Ting, J., and Brown, A. Ciguatera poisoning: a global issue with common management problems. *European Journal of Emergency Medicine* 8, 4 (2001), 295–300.
- [262] Tonge, J., Battey, Y., Forbes, J., Grant, E., et al. Ciguatera poisoning: a report of two outbreaks and a probable fatal case in Queensland. *Medical Journal of Australia* 2, 24 (1967), 1088–90.
- [263] Tubaro, A., Dell’Ovo, V., Sosa, S., and Florio, C. Yessotoxins: A toxicological overview. *Toxicon* 56, 2 (2010), 163–172.
- [264] Van Dolah, F. M. Diversity of marine and freshwater algal toxins. *FOOD SCIENCE AND TECHNOLOGY-NEW YORK-MARCEL DEKKER-* (2000), 19–44.
- [265] Van Dolah, F. M., Kohli, G. S., Morey, J. S., and Murray, S. A. Both modular and single-domain Type I polyketide synthases are expressed in the brevetoxin-producing dinoflagellate, *Karenia brevis* (Dinophyceae). *Journal of phycology* 53, 6 (2017), 1325–1339.

- [266] Van Dolah, F. M., Lidie, K. B., Monroe, E. A., Bhattacharya, D., Campbell, L., Doucette, G. J., and Kamykowski, D. The Florida red tide dinoflagellate *Karenia brevis*: new insights into cellular and molecular processes underlying bloom dynamics. *Harmful Algae* 8, 4 (2009), 562–572.
- [267] Vandersea, M. W., Kibler, S. R., Holland, W. C., Tester, P. A., Schultz, T. F., Faust, M. A., Holmes, M. J., Chinain, M., and Wayne Litaker, R. Development of semi-quantitative pcr assays for the detection and enumeration of *Gambierdiscus* species (Gonyaulacales, Dinophyceae) 1. *Journal of Phycology* 48, 4 (2012), 902–915.
- [268] Verma, A., Hoppenrath, M., Dorantes-Aranda, J. J., Harwood, D. T., and Murray, S. A. Molecular and phylogenetic characterization of *Ostreopsis* (dinophyceae) and the description of a new species, *Ostreopsis rhodesae* sp. nov., from a subtropical Australian lagoon. *Harmful algae* 60 (2016), 116–130.
- [269] Verma, A., Kohli, G., , Harwood, D., Ralph, P., and Murray, S. Transcriptomic investigation into polyketide toxin synthesis in *Ostreopsis* (Dinophyceae) species.). *Environmental Microbiology, In review* (2019).
- [270] Vernikos, G., Medini, D., Riley, D. R., and Tettelin, H. Ten years of pan-genome analyses. *Current opinion in microbiology* 23 (2015), 148–154.
- [271] Vernoux, J.-P., and Lewis, R. J. Isolation and characterisation of Caribbean ciguatoxins from the horse-eye jack (*Caranx latus*). *Toxicon* 35, 6 (1997), 889–900.
- [272] Vinuesa, P., and Contreras-Moreira, B. Robust identification of orthologues and paralogues for microbial pan-genomics using GET_HOMOLOGUES: a case study of pIncA/C plasmids. In *Bacterial Pangenomics*. Springer, 2015, pp. 203–232.
- [273] Watanabe, R., Uchida, H., Suzuki, T., Matsushima, R., Nagae, M., Toyohara, Y., Satake, M., Oshima, Y., Inoue, A., and Yasumoto, T. Gambieroxide, a novel epoxy polyether compound from the dinoflagellate *Gambierdiscus toxicus* GTP2 strain. *Tetrahedron* 69, 48 (2013), 10299–10303.
- [274] Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., and Zdobnov, E. M. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution* (2017).
- [275] Whelan, S., and Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular biology and evolution* 18, 5 (2001), 691–699.
- [276] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

- [277] Wisecaver, J. H., and Hackett, J. D. Dinoflagellate genome evolution. *Annual review of microbiology* 65 (2011), 369–387.
- [278] Withers, N. W. Ciguatera research in the northwestern hawaiian islands: laboratory and field studies on ciguatoxic dinoflagellates in the Hawaiian Archipelago. *Proc. Res. Invest. NWHI* (1984), 144.
- [279] Woerner, J. *Padina* (brown algae)). ian.umces.edu/imageLibrary/, 2011. Integration and Application Network, University of Maryland Center for Environmental Science.
- [280] Xie, W., Lewis, P. O., Fan, Y., Kuo, L., and Chen, M.-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic biology* 60, 2 (2010), 150–160.
- [281] Xie, W., Lewis, P. O., Fan, Y., Kuo, L., and Chen, M.-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology* 60, 2 (2011), 150–160.
- [282] Xu, Y., Richlen, M. L., Morton, S. L., Mak, Y. L., Chan, L. L., Tekiau, A., and Anderson, D. M. Distribution, abundance and diversity of *Gambierdiscus* spp. from a ciguatera-endemic area in Marakei, Republic of Kiribati. *Harmful Algae* 34 (2014), 56–68.
- [283] Yang, Z. *Molecular evolution: a statistical approach*. Oxford University Press, 2014.
- [284] Yasumoto, T., Bagnis, R., and Vernoux, J. P. Toxicity of the surgeon-fishes, 2: Properties of the principal water-soluble toxin. *Bulletin of the Japanese Society of Scientific Fisheries* (1976).
- [285] Yasumoto, T., Igarashi, T., Legrand, A.-M., Cruchet, P., Chinain, M., Fujita, T., and Naoki, H. Structural elucidation of ciguatoxin congeners by fast-atom bombardment tandem mass spectroscopy. *Journal of the American Chemical Society* 122, 20 (2000), 4988–4989.
- [286] Yasumoto, T., Nakajima, I., Bagnis, R., and Adachi, R. Finding of a dinoflagellate as a likely culprit of ciguatera. *Bulletin of the Japanese Society of Scientific Fisheries (Japan)* (1977).
- [287] Yogi, K., Oshiro, N., Inafuku, Y., Hiramata, M., and Yasumoto, T. Detailed LC-MS/MS analysis of ciguatoxins revealing distinct regional and species characteristics in fish and causative alga from the Pacific. *Analytical chemistry* 83, 23 (2011), 8886–8891.
- [288] Yokoyama, A., Murata, M., Oshima, Y., Iwashita, T., and Yasumoto, T. Some chemical properties of maitotoxin, a putative calcium channel agonist isolated from a marine dinoflagellate. *Journal of biochemistry* 104, 2 (1988), 184–187.

- [289] Yoshimatsu, T., Tie, C., Yamaguchi, H., Funaki, H., Honma, C., Tanaka, K., and Adachi, M. The effects of light intensity on the growth of Japanese *Gambierdiscus* spp. (Dinophyceae). *Harmful algae* 60 (2016), 107–115.
- [290] Zhang, H., Bhattacharya, D., and Lin, S. Phylogeny of dinoflagellates based on mitochondrial cytochrome b and nuclear small subunit rDNA sequence comparisons. *Journal of Phy-cology* 41, 2 (2005), 411–420.
- [291] Zhang, H., Bhattacharya, D., and Lin, S. A three-gene dinoflagellate phylogeny suggests monophyly of prorocentrales and a basal position for *Amphidinium* and *Heterocapsa*. *Jour-nal of Molecular Evolution* 65, 4 (2007), 463–474.
- [292] Zhang, H., Bhattacharya, D., Maranda, L., and Lin, S. Mitochondrial cob and cox1 genes and editing of the corresponding mRNAs in *Dinophysis acuminata* from Narragansett Bay, with special reference to the phylogenetic position of the genus *Dinophysis*. *Applied and environmental microbiology* 74, 5 (2008), 1546–1554.
- [293] Zhang, H., Hou, Y., Miranda, L., Campbell, D. A., Sturm, N. R., Gaasterland, T., and Lin, S. Spliced leader RNA trans-splicing in dinoflagellates. *Proceedings of the National Academy of Sciences* 104, 11 (2007), 4618–4623.
- [294] Zhang, H., and Lin, S. Retrieval of missing spliced leader in dinoflagellates. *PLoS One* 4, 1 (2009), e4129.
- [295] Zhang, H., Zhuang, Y., Gill, J., and Lin, S. Proof that dinoflagellate spliced leader (DinoSL) is a useful hook for fishing dinoflagellate transcripts from mixed microbial samples: *Symbiodinium kawagutii* as a case study. *Protist* 164, 4 (2013), 510–527.
- [296] Zheng, W., DeMattei, J. A., Wu, J.-P., Duan, J. J.-W., Cook, L. R., Oinuma, H., and Kishi, Y. Complete relative stereochemistry of maitotoxin. *Journal of the American Chemical Society* 118, 34 (1996), 7946–7968.
- [297] Zhou, Z., Miwa, M., and Hogetsu, T. Analysis of genetic structure of a *Suillus grevillei* pop-ulation in a *Larix kaempferi* stand by polymorphism of inter-simple sequence repeat (ISSR). *New Phytologist* 144, 1 (1999), 55–63.

CHAPTER 8

Appendix.

8.1 Published manuscripts

8.1.1 Qualitative and quantitative assessment of the presence of ciguatoxin, P-CTX-1B, in Spanish Mackerel (*Scomberomorus commerson*) from waters in New South Wales (Australia)

Overview: Citizen science project investigating the presence of CTX analogues in Spanish mackerel from New South Wales, Australia, the primary vector in reported CFP intoxication in Queensland, Australia.

DOI: <https://doi.org/10.1016/j.toxrep.2017.06.006>

Authors: Gurjeet S. Kohli, Kristina Haslauer, Chowdhury Sarowar, Anna Liza Kretzschmar, Mark Boulter, D.Tim Harwood, Olivier Laczka, Shauna A. Murray

Submitted: 31 January 2017

Accepted: 13 June 2017

8.1.2 Response to “More surprises in the global greenhouse: Human health impacts form recent toxic marine aerosol formulations, due to centennial alterations or world-wide coastal food webs”

Overview: Response to manuscript grossly mis-citing the *Gambierdiscus lapillus* sp. nov. manuscript **Chapter 2** to draw unsupported conclusions about *Gambierdiscus* associated health risks.

DOI: <http://dx.doi.org/10.1016/j.marpolbul.2017.05.035>

Authors: Shauna A. Murray, Penelope Ajani, Anna Liza Kretzschmar, Arjun Verma

Available online: 5 June 2017

8.2 Submitted manuscripts

8.2.1 Morphological and molecular phylogenetic identification and record verification of *Gambierdiscus excentricus* (Dinophyceae) from Madeira Island (NE Atlantic Ocean).

Overview: First verified report of the toxic species *G. excentricus* southwest of Portugal.

Authors: Mona Hoppenrath, Anna Liza Kretzschmar, Michaela E. Larsson, Mona Hoppenrath, Manfred Kaufmann, Shauna A. Murray

Submitted: Marine Biodiversity Records in August 2018, in revision.

8.2.2 Characterisation of two toxic *Gambierdiscus* spp. (Gonyaulacales, Dinophyceae) from the Great Barrier Reef (Australia): *G. lewisii* sp. nov. and *G. holmesii* sp. nov.

Overview: Characterization of two new toxic *Gambierdiscus* species, co-isolated with *G. lapillus*, from the Great Barrier Reef.

Authors: Anna Liza Kretzschmar, Michaela E. Larsson, Mona Hoppenrath, Martina A. Doblin, Shauna A. Murray

Submitted: Protist in October 2018, in revision.

8.2.3 Development of a quantitative PCR assay for the detection and enumeration of a potentially ciguatoxin-producing dinoflagellate, *Gambierdiscus lapillus* (Gonyaulacales, Dinophyceae).

Overview: Molecular monitoring tool development for *G. lapillus*, validated for environmental screening with samples from Heron Island, GBR.

Authors: Anna Liza Kretzschmar, Arjun Verma, Gurjeet Kohli & Shauna Murray

DOI: <https://doi.org/10.1101/544247>

Submitted: PLoS ONE in February 2019, in revision.

8.2.4 Trial by phylogenetics - Evaluating the Multi-Species Coalescent for evolutionary inference on taxa with high levels of paralogy (Gonyaulacales, Dinophyceae).

Overview: Elucidation of the evolutionary placement of *Gambierdiscus* within the gonyaulacales, with a focus on common data and model choices in the field.

Authors: Anna Liza Kretzschmar, Arjun Verma, Shauna Murray, Tim Kahlke, Mathieu Fourment

& Aaron E. Darling

DOI: <https://doi.org/10.1101/683383>

Submitted: peerJ in June 2019.