PROJECT TITLE

Evolutionary dynamics of the algae: Understanding adaptive potential under environmental change

BACKGROUND

Role and origin of eukaryotic algae

Eukaryotic algae are critical for the structure and functioning of marine ecosystems. They are responsible for nearly half of the net primary production that takes place on the planet [1,2]. Their primary significance in relation to this project is that they play central roles in ecosystem functioning and climate forcing.

First, algae form a crucial component of oceanic as well as coastal ecosystems. They produce organic compounds through photosynthesis and are the mainstay of most aquatic food chains [2]. Multicellular macroscopic algae play a key role in shallow marine ecosystems, where they are often the principal structural components and facilitate a rich fauna and flora (e.g. kelp forests) [3,4].

Second, algae play a major role in climate forcing models through their function in biogeochemical cycles and oceanic CO_2 absorption [2,5]. Algal photosynthesis is the primary contributor to the absorption of atmospheric CO_2 in the ocean, the process known as the "biological pump" [6]. Photosynthesis lowers the partial pressure of CO_2 in the upper layer of the ocean and promotes the diffusion of CO_2 from the atmosphere. In addition, a substantial amount of the carbon fixed by algae in the upper water layers sinks to, and gets stored in, the interior of the ocean. Considering that virtually all of the carbon stored in sediments is of marine origin [7], it is evident that the biological pump is a very important buffer of human CO_2 emissions [8,9], acting positively to mediate the global climatic response to carbon emissions.

Eukaryotic algae originated from the merger of a heterotrophic eukaryote and a photosynthetic cyanobacterium [10,11]. In this endosymbiosis event, the heterotrophic eukaryote engulfed a type of cyanobacterium that, over time, evolved to become an organelle (the chloroplast) stably integrated into the eukaryote cell. As descendents of the resulting eukaryotic alga diversified over time, the eukaryotic algae from various lineages became involved in secondary endosymbiosis events [10,11]. In secondary endosymbiosis, a heterotrophic eukaryote engulfs a photosynthetic eukaryote, resulting in a eukaryotic alga with a complex eukaryote-derived secondary plastid. At least seven of these eukaryote-eukaryote endosymbioses have occurred, involving heterotrophic host eukaryotes from many of the major eukaryotic groups and plastids derived from various other places in the eukaryotic tree of life [11]. As a consequence, the capacity to perform photosynthesis has spread to many different eukaryotic lineages, giving rise to the major groups of eukaryotic algae (e.g. dinoflagellates, euglenophytes, etc.) existing today.
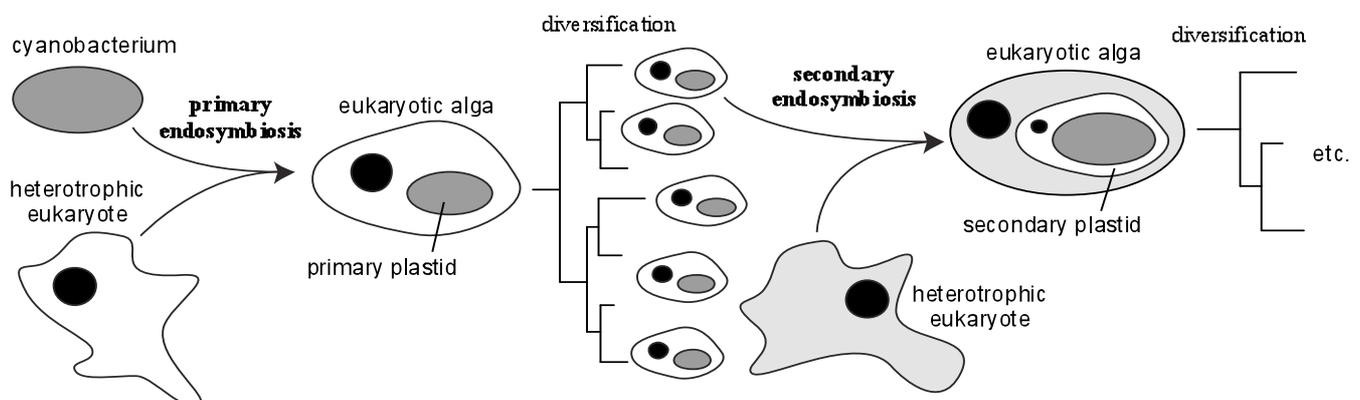
Problem statement

Over the past 200 years, human CO_2 emissions have contributed substantially to global change [8,12]. These emissions have altered the global carbon cycle [6] and have modified environmental conditions, perhaps most pronouncedly in regard to global temperatures [12]. Considering the cornerstone function of eukaryotic algae in community structure and global element cycles, it is vital to acquire knowledge of their capacity to adapt to changing environmental conditions and to efficiently act as a CO_2 pump. Many questions in this field remain unanswered. My goals are to resolve two crucial evolutionary questions relating to the flexibility of algae to deal with changing environments and develop the methods necessary for their resolution.

Question 1: How have algae adapted to nutrient limitations on evolutionary timescales?

Algal photosynthesis has been cited as an important factor in mitigating human CO_2 emissions due to its central function in the biological pump [8,13]. However, if the biological pump is to absorb a high proportion of anthropogenic CO_2 in the coming century and help mitigate global warming, its efficiency must increase [14]. Nutrient limitations currently hinder algal growth and, consequently, the efficiency of the biological pump, in vast areas of the world's oceans [15,16]. Seventeen elements are known to be essential to organismal life [17]. Macronutrients like C, N and P are required in relatively large quantities, whereas trace elements like Cu, Zn and Fe are needed in much smaller amounts yet are required as a cofactor in metalloproteins. About one quarter of all proteins require a trace metal ion to carry out their functions, and they are responsible for many crucial redox processes in the cell, including photosynthesis [2,18,19]. Phototrophic biomass can increase in aquatic ecosystems until one of the critical elements becomes limiting. In present-day oceans, nitrogen and trace elements (especially Fe) are known to limit phytoplankton growth [15], and the ongoing acidification of the ocean due to CO_2 emissions is likely to further decrease the bioavailability of iron [20].

It is not well understood how algae have adapted to low iron availability in the past. While some attention has been paid to transcriptional responses to iron starvation on ecological timescales [21] and adaptations of some algae to more effectively capture and store iron [22], little is known about changes in trace metal requirements over evolutionary timescales. Comparative studies of the trace elemental composition of 29 phytoplankton species grown under nutrient-replete conditions have shown differences between plastid types, as algae with green-type plastids require more Cu, Zn and Fe than those with red-type plastids [23-25]. This finding has led to the hypothesis that endosymbioses drive trace metal utilization. In other words, during the course of endosymbiosis events the host lineage adopts the trace metal utilization characteristics of the alga that is engulfed (and subsequently transformed into a plastid). However, this hypothesis has never been verified by studies focusing on the evolutionary processes behind the establishment of trace metal requirements.

In addition to trace metals, and in particular iron, nitrogen is often cited as a limiting nutrient for algal growth [15]. Nitrogen is a macro-element that is used in many macromolecules. Similar to the situation for trace elements, knowledge of the evolutionary processes acting on nitrogen utilization is key to understanding growth limitations. Several studies have documented variation of the experimental C:N:P stoichiometry between algal species [24,25], but the evolution of this trait has not been studied in detail.

Question 2: How do algal niches evolve?

Despite the critical role of algae in marine ecosystem functioning and the notion that the distribution and abundance of algal populations can cause dramatic changes in productivity, with cascading effects on ecosystems and human economies [26], relatively little is known about the environmental determinants of algal distribution ranges and how the environmental tolerances of algae evolve. The Hutchinsonian niche (hereafter referred to simply as "the niche") is defined as the broad-scale environmental conditions within which a species can persist. It is a key determinant of the geographic distribution of any given species. Three environmental factors cited frequently as having major impacts on algal species distributions are sea surface temperature, salinity and nutrient availability [27-30], all three of which are increasingly affected by human activities [12]. If we understood how algal niches evolve, we could predict their responses to significant and rapid environmental change. Unfortunately, our knowledge of the evolution of algal niches is in a very early stage. Even though a few case studies have looked at how species'

temperature affinities have changed along phylogenetic trees [31,32], much work remains to be done to fully understand the evolutionary dynamics and adaptability of algal niches.

A second unknown is how rates of diversification (speciation and extinction) relate to climate. The paleontological record of marine invertebrates suggests that rates of speciation are higher in tropical seas than in temperate oceans, and that rates of extinction show the opposite pattern [33]. Extinctions are thought to be largely diversity-independent and caused by external factors such as climate change [33,34]. Regrettably, we cannot turn to the fossil record to learn about the diversity dynamics of most algae because their soft bodies do not fossilize well and the paleontological record is very incomplete. If we are to understand the speciation and extinction dynamics underlying the present diversity of algae and how environmental factors affect diversification processes, another approach will be needed.

Question 3: How can informative evolutionary models be applied?

Macroevolutionary biology is concerned with change that occurs above the level of species, in other words changes occurring over geological rather than ecological timescales. It studies the evolution of characters and the dynamics of diversification (speciation and extinction) in groups of organisms. Macroevolution has long been exclusively the domain of paleontologist, who can observe trait evolution, speciation and extinction by studying geological time series. However, in the past few decades molecular phylogenies, along with data on the species' characteristics, have increasingly been used to answer the same sort of questions with the assistance of mathematical modeling techniques [35-37]. This approach is not just useful for algae, but can be applied to any group of organisms with a poor fossil record.

Until very recently, only very simple models could be applied to study macroevolutionary patterns from phylogenetic data. For example, the models assumed that rates of trait evolution or speciation were constant throughout the phylogenetic tree – an assumption known to be incorrect in many cases. Although several minor improvements have been made to the models in the past few years, many macroevolutionary hypotheses cannot be tested using existing methods. Answering the questions that workers in the field regularly pose involves fitting complex models for which likelihood functions are difficult or impossible to formulate. While the concepts and the data required to answer these questions are available, a lack of methods stands in the way of answering them.

AIMS AND APPROACH

This project aims to address these important and pressing questions in algal biology and develop the analytical methods needed to solve them.

Goal 1: Characterization of the evolutionary dynamics of elemental stoichiometry in algae. I will study changes in trace metal utilization that occur in association with endosymbiosis events and alterations of C:N:P stoichiometry that happened during the evolutionary history of the algae.

Goal 2: Characterization of niche evolution in macroalgae and its impact on diversification dynamics. I will assess how rapidly the algal niche evolves, which groups have more conserved niches, and which can adapt more rapidly. I will also investigate how the environment affects rates of speciation and extinction.

Goal 3: Development of approximate Bayesian computation (ABC) to fit complex evolutionary models. I will develop ABC methods to fit the complex models of continuous trait evolution and speciation-extinction dynamics that are needed to address goals 1 and 2.

Conceptual framework: Studying the past to inform the present and future

I wish to study the evolutionary processes that underlie and perhaps hinder the algae's capacity to act as a carbon sink and adapt to environmental change. Characterizing the relevant evolutionary processes that have acted in the past will increase our understanding of present patterns and is a first step to predicting (and more effectively managing) the future.

Integrating different sources of data and techniques from various disciplines is crucial to reach these goals. I will use knowledge and techniques from various disciplines (environmental science, algal physiology, molecular biology, paleoceanography, statistical modeling) to characterize the evolutionary

processes related to algal function in the global biological pump and their adaptability to environmental change.

The general approach toward goals 1 and 2 involves three steps:
1. Gather data for traits (e.g. Fe utilization, temperature affinities) for an appropriate set of species.
2. Generate molecular sequence data to infer a high-quality phylogenetic tree of the same species.
3. Use statistical modeling to characterize the evolution of the traits along the phylogeny.

I will now summarize the specific approach that will be taken for each of the goals.

Approach to goal 1: Elemental stoichiometry

For this goal, I aim to characterize the evolution of nitrogen and trace element utilization. Because it can be anticipated that change in these features occurs over long timescales [24], the taxon sampling will include representative species of all major groups of algae. Two types of information are needed and are being gathered to answer the questions:

1. Data about trace element utilization of algae grown under nutrient-replete conditions. Data on the utilization of all the important cations (Mg, Ca, Sr, Fe, Mn, Zn, Cu, Co, Cd, Mo) have already been published for 29 species [24,25]. My student Jens Frickel (Ghent University) is currently doing his Masters thesis with the goal of generating similar data for another 34 species, bringing the total to 63 species spread across the different algal groups.
2. Experimental data on the C:N:P stoichiometry of the same 63 species will be available from the literature, and from our own experiments.

To generate phylogenetic trees, multi-gene alignments will be generated for the above-mentioned 63 species. Due to the various plastid endosymbiosis events mentioned above, the evolutionary history of the eukaryotic host taxa differs from that of the endosymbiotic plastids [38]. To accommodate these differences, two phylogenies will be generated: one of the host cells using nuclear genes that belong to the host, and a second phylogeny of the plastid lineage based on genes from the plastid genome.

The great majority of the DNA sequence data needed to construct these phylogenies is already available on GenBank (http://www.ncbi.nlm.nih.gov/genbank/). Nonetheless, two critical data gaps need to be filled before comprehensive and statistically supported trees can be inferred. As part of this project, I aim to generate additional sequence data for a number of crucial taxa. Currently, the placement of the plastids of the dinoflagellates and chlorarachniophytes is not well characterized. Although a few studies have addressed this question [39,40], these only place the groups in question in a sparsely sampled tree. I aim to generate phylogenetic datasets with a richer taxon sampling that will pinpoint the origins of these groups more accurately. For chlorarachniophytes, complete plastid genome data are available [11], but very little sequence information is available for the lineage of green algae that gave rise to the progenitor of the chlorarachniophyte plastid. To fill this gap, I will sequence the plastid genomes of twelve Ulvophyceae and Trebouxiophyceae. A second knowledge gap is the phylogenetic affinities of the dinoflagellate plastids. It is known that dinoflagellates have independently initiated symbioses at least five times from plastids of various origins [41], but the exact origins of the plastids have remained difficult to pinpoint. In this case, the problem is not a lack of information about the lineages that sprouted the plastid, but is due to the paucity of sequence data of the dinoflagellates themselves. In dinoflagellates, plastid genes are often transferred to the nucleus while the endosymbiont genome deteriorates to a few minicircles [42,43]. I aim to sequence the minicircles and the plastid genes transferred to the nucleus of six dinoflagellates belonging to three plastid types. DNA and RNA isolation and preparation will be carried out with established methods [44-46]. All sequencing will use Illumina high-throughput sequencing technology to obtain large amounts of sequence data at moderate cost. I will use the latest generation sequencing kits that permit running paired-end runs of 150 nucleotides each, alleviating the problems associated with short runs in earlier kits.

After the DNA sequences have been generated, they will be combined with sequences available on GenBank and assembled into multi-gene datasets, one containing nuclear genes and another containing plastid genes. Host and plastid phylogenies will then be inferred from these datasets using model-based phylogenetic inference methods [47-51] and calibrated in geological time using relaxed molecular clock techniques [47,52-54] and previously published fossil information. Even though elucidating the history of

endosymbiosis events with molecular phylogenetics has proven difficult, the application of high-throughput sequencing now allows great increases in the number of taxa sampled and, together with more advanced phylogenetic techniques, is rapidly improving the situation [38]. Furthermore, any uncertainty that may remain in the final trees can easily be integrated in the next level of analysis using Bayesian analysis techniques.

With the phylogenetic trees and the data about trace element and C:N:P stoichiometry, models can be designed to answer some key questions. First, the hypothesis relating trace metal utilization to plastid endosymbiosis will be tested in an explicit evolutionary framework using novel models of trait evolution (see below, goal 3). These models will help infer the degree to which the host and endosymbiont contribute to the elemental stoichiometry of the resulting cell. Second, I will examine which lineages have developed lower dependencies on important limiting nutrients using ancestral character state estimation. Third, I will develop a model that relates the ocean redox state to rates of diversification in groups with different trace metal requirements to test the hypothesis that changes in ocean chemistry have led to a reduction of phytoplankton with green-type plastids relative to species with red-type plastids [23,55-58].

This Future Fellowship will form the basis of future research projects. For example, it lays the foundation for detailed analyses of the genomic changes underlying the evolution of elemental stoichiometry. The genomes of strains that were selected for determination of trace element composition in my lab have been or are being sequenced. The genome can provide very rich information about elemental stoichiometry because it holds the key to how much C, N and P are used in macromolecules [59,60] and the metalloproteome can explain changes in trace metal utilization [61-63]. Using bioinformatics techniques from the new field of stoichiogenomics [64], I aim to characterize environmental selection pressures on components of the photosynthetic apparatus where metal substitutions are possible. This will identify the molecular background of nutrient-driven growth limitation in nature and offer potential solutions.

Approach to goal 2: Niche evolution

The second goal is to characterize the evolution of macroecological niches and determine the environmental correlates of speciation and extinction rates. The concept map below illustrates the approach I will take to achieve this goal.
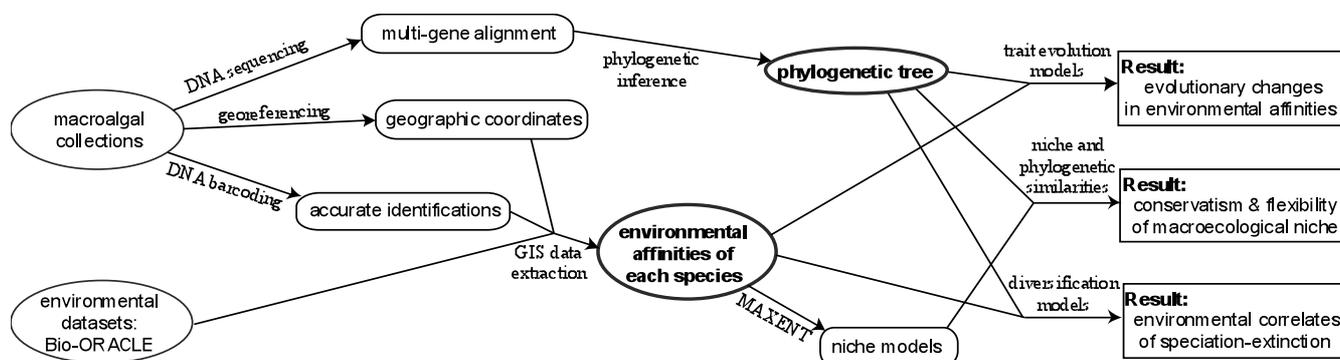
Two types of data are needed to address this goal: (1) information about the environmental affinities of species; and (2) multigene alignments to build phylogenetic trees. I will gather information about the environmental affinities of species by combining species occurrence data with global datasets of environmental variables. Obtaining species occurrence data is slightly more complicated in algae than in many other organisms, the reason being that in many cases algal morphology is not representative of species boundaries, and specimens identified morphologically and referred to specific geographical localities in herbaria, published records and various databases such as AlgaeBase or AussiAlgae are often untrustworthy. Therefore, species identifications will be achieved by applying species delimitation algorithms to collections of DNA barcodes, which yields much more reliable results [32,65]. These methods detect the interface between within- and above-species processes in sequence evolution (micro- vs. macroevolution) and define species boundaries accordingly [66,67]. Once specimens have been assigned to species, it is a straightforward task to assemble a database of species distribution data by georeferencing the sequenced specimens [32]. Due to the restriction of having to work with DNA

barcodes for species identification, I will focus on three groups of macroalgae for which extensive sets of DNA barcodes are already available: the Ulvophyceae (> 8,000 sequences), Dictyotales (> 2,000) and Rhodophyta (> 25,000). At present, there are some geographical sampling biases in the datasets of DNA barcodes, with tropical biodiversity in particular being underrepresented. Therefore, I aim to carry out fieldwork in two understudied Indo-Pacific regions in which algal biodiversity is known to be especially high: Papua New Guinea and the Marshall Islands. Macroalgal specimens will be collected and processed appropriately, the GPS coordinates stored, and a DNA barcode will be generated for each specimen.

This work will be a comprehensive dataset of accurately identified algal specimens and geographic coordinates specifying where they were collected. From this, the environmental (macroecological) affinities of species will be determined by extracting environmental information from the Bio-ORACLE global GIS datasets that my colleagues and I recently produced for these sorts of applications [68].

Phylogenetic trees will be generated from multi-gene alignments. Where possible, these data will be mined from GenBank [69], although but some gaps in the datasets will need to be filled by generating novel sequence data from a selection of specimens. Phylogenetic trees will be inferred as described above.

With the environmental affinities of species and phylogenetic trees in hand, three paths will be taken to accomplish the desired goals. First, models of trait evolution will be applied to characterize the evolutionary history of species' environmental affinities. This approach, which is illustrated in my recent publication on the green seaweed genus *Halimeda* [32], will yield information about which lineages have adapted to particular environmental conditions and when this happened in geological time. Joint interpretation of these results with paleoceanographic information can potentially identify the environmental changes driving the observed evolutionary patterns [70]. The second path involves an extra step in which niche models are generated for each species using maximum entropy techniques [71,72]. By comparing niche similarities [73,74] and phylogenetic affinities between species pairs, this approach will provide detailed information about how conserved (or dynamic) the niche is and about the role of macroecological niche differentiation in marine speciation [75]. The third path aims to identify the environmental correlates of rates of speciation and extinction. Identifying correlates of rates of speciation and extinction is generally done by paleontologists, but molecular phylogenies also contain the signature of past speciation and extinction events, information that can be extracted with diversification models [76,77]. I will develop new methods that permit identifying the relationship between climate (sea surface temperature) and rates of speciation and extinction (see below). Taken together, these three approaches will yield detailed information about how the Hutchinsonian niche of algae determines their distribution ranges, how algal niches change in relation to global climate variations, and how this influences rates of speciation and extinction. These results have applications in predicting how algal distribution ranges will shift under future climate scenarios, estimating how fast algae can adapt to climate change, and how climate change may alter speciation-extinction dynamics.

Approach to goal 3: ABC for complex models

The third goal of the project is to develop models of trait evolution and of evolutionary diversification that will be necessary for satisfactory inference under goals 1 and 2. A large body of literature lays the foundation for these models, and many of the simple models have already been implemented, mostly in the R environment for statistical computation [78-83]. However, these simple models will not suffice to accurately describe the evolution of traits like elemental stoichiometry and climate-dependent speciation-extinction dynamics.

Evolutionary models of elemental stoichiometry need to take the proposed impact of endosymbiosis events into account and allow measurement of the contribution that host and plastid have on trace metal utilization. In other words, the models need to distinguish between the background rate of evolution of the trait due to vertical inheritance and the changes happening as a consequence of an endosymbiosis event. The BMC model serves as a first step towards achieving this goal [84]. It allows for two different rates of continuous trait evolution, one rate for branches where no changes in a discrete trait occur and a second rate for branches where the discrete trait changes. It is obvious how this model applies to the evolution of trace metal utilization, with one rate for branches along which an endosymbiosis event occurs and another rate for branches without endosymbiosis. An alternative model that would better approximate the

anticipated pattern of evolution of trace metal utilization has a single background rate throughout the tree, with pulsed changes on branches along which endosymbiosis events take place. I will implement and evaluate this model. Both these models, when applied to the host phylogeny, will indicate the impact of the plastid on the evolution of trace element utilization. Similarly, their application to the plastid phylogeny will quantify the contribution of the host cell.

To test the hypothesis that rates of diversification in groups with different trace metal requirements depend on the redox state of the ocean, I will implement a model that relates the rate of diversification to both trace metal requirements and trace metal availability. Whereas the former is treated as a species trait, the latter will change gradually through time, affecting all lineages in the tree similarly. Trace metal availability will be derived from paleoceanographic data, i.e. the redox state of the ocean that largely determines the bioavailability of trace metals [85].

Models of climate-dependent diversification need to relate climatic variables (e.g. sea surface temperature) to rates of speciation and extinction. The foundation for such models has already been laid with the YWC model that relates the logit of the net rate of diversification to a linear combination of predictors [86]. Two properties of this model limit its utility for my purpose. First, the model requires known ancestral trait values, which are difficult to obtain for organisms' environmental affinities (but see [75,87]). Second, the model is a birth-only model in which extinction is not taken into account. I will extend this model to relate environmental variables to both speciation and extinction, and remove the dependency on ancestral state values using an approximate Bayesian computation (ABC) framework.

Due to the difficulty or impossibility of defining and computing the likelihood functions of the complex models that are needed to evaluate some macroevolutionary hypotheses, a new approach is needed to undertake complex, yet robust statistical inference. ABC is a powerful and flexible simulation technique that can be used to infer parameters and choose between models in the complicated scenarios that have to be considered here [88-90]. It replaces the likelihood computation by a step that involves simulating artificial data for different parameter values and comparing summary statistics of the simulated data to summary statistics of the observed data in a rejection sampling framework. The use of ABC in macroevolutionary inference is in its infancy, and to my knowledge only one study using this approach has been published [77]. There exists a great opportunity to pioneer the development of this approach to permit inference on some of the most significant questions in macroevolution. One opportunity for doing this may be to use the R packages MECCA or TreEvo to apply ABC to macroevolutionary data. This software is being developed by Graham Slater at UCLA (MECCA) and Brian O'Meara at University of Tennessee (TreEvo) but has not been released to the public yet. After defining the models described above mathematically, MECCA, TreEvo (or a similar tool) will be used to infer their parameters and compare between different models.

Research schedule

The tasks I aim to complete in four years are ambitious, and thorough planning and good coordination with students and other collaborators will be necessary. The table below illustrates the anticipated work schedule for the duration of the project.

	Year 1				Year 2				Year 3				Year 4			
Logistical planning with collaborators	■															
Goal 1: Gather phylogenetic data					■	■	■	■	■	■	■					
Goal 1: Infer phylogenies and model evolution											■	■	■	■		
Goal 2: Field work		■			■				■							
Goal 2: Gather DNA barcodes & phylogenetic data		■	■	■	■	■	■	■	■	■						
Goal 2: Gather species distribution data	■	■	■	■	■	■	■	■	■							
Goal 2: Infer phylogenies and model evolution									■	■	■	■				
Goal 3: Define models & evaluate ABC approach	■	■														
Publication and communication		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

Significance and Innovation

Addressing important problems

Poor understanding of climate change impacts on biodiversity and ecosystem functioning is an issue of global and national significance. Within Australia, the potentially adverse impacts of future climate change on biodiversity conservation are widely acknowledged, notably through the listing of anthropogenic climate change as a Key Threatening Process under Commonwealth and State legislation.

If we are to understand the effects of changes in the environment on keystone species of marine ecosystems and the functioning of the biological pump, there is an urgent need to characterize how the organisms involved have responded to changes in the environment in the past. The present project uses state-of-the-art evolutionary approaches, combined with relevant ecological and physiological data, to move this objective forward.

Research on evolutionary dynamics of environmental niches and rates of diversification tends to be high impact research. Much of the recent work has been published in ERA ranked A* journals like *Science*, *PNAS* and *Global Ecology & Biogeography*. Similarly, recent papers about elemental stoichiometry have been published in widely read journals like *Nature*, *PNAS*, *Proceedings B* and *New Phytologist*. Novel techniques in evolutionary modeling are commonly published in A* journals such as *Systematic Biology* and *American Naturalist*.

Advancing the knowledge base

The current project advances the knowledge base by integrating the disciplines of algal physiology, macroecology and paleoceanography with molecular phylogenetics and mathematical modeling. In what follows, I will give a brief overview of how the several disciplines are combined to reach each of the goals, what sorts of new insights this is expected to yield, and the new perspectives that this brings.

Goal 1: Here, physiological and paleoceanographical data are integrated with molecular phylogenies to model the evolution of elemental stoichiometry and its effects on diversification. This approach will permit documenting the impact of endosymbiosis events on this important physiological process and evaluating whether the regime shift from green algal dominance to chromalveolate dominance across the Paleozoic could have been due to altered speciation-extinction dynamics following changes in the bioavailability of trace elements. Finally, this research has the potential to identify lineages that have evolved lower requirements for limiting nutrients, which can in turn instigate research into the genomic features responsible for these lower nutrient requirements.

Goal 2: To reach this goal, macroecological data are to be integrated with molecular phylogenetics. This advances the knowledge base in two ways; first, the rate of macroecological niche evolution of multiple algae lineages will be estimated; second, information about the environmental correlates of rates of speciation and extinction will become better understood. This knowledge has various applications in climate change biology. First, studies of marine community dynamics under climate change scenarios can use the information gained about the macroecological niche of the algae that form the base of the food chain and are important habitat creators. Second, it becomes possible to assess whether species may be able to adapt to climate variations quickly enough to keep up with global change. Third, one can derive information about how global change is expected to affect rates of speciation and extinction in the long run.

Goal 3: The novelties in the modeling aspects of the study include the definition of fairly detailed models of trait evolution and species diversification. Fitting these models using ABC approaches is a major step forward in computational macroevolutionary studies. The application of these techniques to our large datasets will not only produce insights into the questions at hand, but will also illustrate the strengths and potential weaknesses of this approach to address macroevolutionary questions. Furthermore, the developed models can be used to answer similar questions in other taxa and in other situations where gradual evolution with occasional pulses of change may be hypothesized to have taken place.

Innovative research methods

The interdisciplinary approach used here is innovative in many ways. In addition to using state-of-the-art data acquisition and analysis techniques, the development of novel models linking data and knowledge from different fields in an evolutionary framework is an important innovation with applications in many other evolutionary questions.

With the application of ABC in macroevolutionary questions, this project is on the cutting edge of the field. Bayesian methods in general and ABC in particular are proving to be a very powerful complement to more classical frequentist approaches. Simulation is a very useful technique for learning about complex systems, and the ABC framework allows simulating the parameter values of the complex models needed to answer today's questions. The power of ABC in the context of this project is that it will allow me to integrate ideas and data from various disciplines (algal physiology, environmental sciences, paleoceanography) into evolutionary models. This will lead to models that are more biologically reasonable and that yield much more informative results than the presently available models.

In addition to developing these novel inference tools, the project will use several state-of-the-art techniques in data acquisition and analysis:

High-throughput DNA sequencing: The chloroplast genome, minicircle DNA and transcriptome sequencing to be done for phylogenomic inference of plastid and host phylogenies in Goal 1 will be entirely carried out using the latest generation of Illumina high-throughput sequencing. At present, this allows runs of paired-end 150 x 150 bp reads, yielding a total of 6 Gbp per lane or 48 Gbp per flow cell.

Automated DNA barcoding: While DNA barcoding is still carried out using traditional capillary sequencing, significant advances have recently been made in automating this process. I will use the ALGA (Algal Life Global Audit) facilities at the University of New Brunswick to generate greatly accelerated numbers of DNA barcodes. At this facility, DNA extraction and PCR have been automated using robots. Quality control of the resulting sequences and their addition to reference alignments have been automated with scripts that I have developed.

High-performance computing: Phylogenetic inference and ABC simulations are very computationally expensive techniques, but fortunately both are highly parallelizable. I will use the super computer facilities of the recently established Victorian Life Sciences Computation Initiative, one of the largest high-performance computing facilities in life sciences worldwide, which is hosted by the University of Melbourne.

COLLABORATION

The interdisciplinary nature of the project will require collaboration with scientists from several disciplines. I have made contact and reached agreement to collaborate with key scientists working in the various fields that come together in my project. I will also continue long-standing collaborations with my present colleagues Olivier De Clerck and Frederik Leliaert in the Phycology Research Group at Ghent University (Belgium).

I have chosen to relocate to the University of Melbourne because it provides the perfect intellectual and physical environment for this research. Michael McCarthy and Brendan Wintle are experts in ecological modeling and Bayesian inference and will assist in developing complex models and inferring parameters using ABC. Ross Waller's lab specializes in the molecular biology and genome biology of dinoflagellates, and his expertise will help in developing tools to sequence the dinoflagellate minicircles. Jane Elith's expertise in maximum entropy niche modeling will be relevant to the inference of species distribution models needed to assess rates of niche evolution. Several other scientists at the University of Melbourne have research programs that overlap with mine, including algal biology (Rick Wetherbee), climate change impact and the genetics and ecology of adaptation (Ary Hoffmann, David Karoly, Mike Kearney), molecular phylogenetics (Mike Bayly) and trace metal physiology (Ian Woodrow). In conclusion, at the University of Melbourne I will be surrounded with people who have expertise in many domains relevant

to my project, and this would provide an ideal environment in which to continue and expand my research. My expertise in evolutionary inference techniques, as well as marine botanical biodiversity and ecology, will strengthen the research, teaching and postgraduate training of the School of Botany, and my appointment will hopefully encourage other faculty collaborators to expand their research programs into the marine realm and integrate evolutionary approaches into explaining the processes they study.

List of collaborators for this project:

Modeling and Bayesian statistics: Michael McCarthy & Brendan Wintle (University of Melbourne)

Algal DNA barcoding and phylogenetics: ALGA (Algal Life Global Audit, led by Gary Saunders, University of New Brunswick, Canada), Phycology Research Group at Ghent University

Molecular biology, genomics, and Tree of Life: Ross Waller (University of Melbourne), Hwan Su Yoon (Bigelow Lab of Ocean Science, USA)

Paleoceanography and Paleontology: Andy Knoll & Dave Johnston (Harvard University, USA), Ariel Anbar (Arizona State University), Steve LoDuca (Eastern Michigan University, USA)

Niche modeling: Jane Elith (University of Melbourne)

Algal physiology and analytical chemistry: John Beardall (Monash University, Australia), Frank Vanhaecke (Ghent University, Belgium), Phycology Research Group at Ghent University

Ghent University and Harvard University are host organizations in this project. I will visit Ghent University in year 2 and Harvard University in year 3 of the Fellowship.

NATIONAL RESEARCH PRIORITIES AND TARGETED PRIORITY AREAS

This project will make significant contributions to two national research priorities (NRP): (1) An environmentally sustainable Australia; and (2) Frontier technologies for building and transforming Australian industries.

The NRP "An environmentally sustainable Australia" includes the priority goals "Reducing and capturing emissions in transport and energy generation" and "Responding to climate change and variability". The proposed work on characterizing the evolution of trace metal utilization is relevant to our understanding of the nutrient limitations that currently hinder the biological pump and its function in capturing CO_2 emissions. The proposed work on evolutionary dynamics of the environmental niche and speciation-extinction processes in relation to the environment is directly relevant to the "Responding to climate change and variability" goal.

The NRP "Frontier technologies for building and transforming Australian industries" includes the priority goal "Frontier technologies". Several frontier technologies will be used in this project, among which are high-throughput sequencing technologies and high-performance computing for phylogenetic inference and ABC. Furthermore, even though the outcomes of the project do specifically not include applications, the increased knowledge base will offer new perspectives for industries working in the field of biological carbon sequestration.

COMMUNICATION OF RESULTS

I have an excellent track record of communicating my research to the scientific community. This includes publications in high-profile refereed journals, presentations at national and international meetings, and participation in workshops. If I am a successful applicant, I will continue to use these avenues to communicate the research findings of my Future Fellowship, and I am keen to further disseminate my results in working groups and popular magazines. I have communicated my research results to a broader audience through newsletters, and my research has featured in several popular media (BBC, United Press International, science blogs).