CENTER FOR CHEMICAL CURRENCIES OF A MICROBIAL PLANET (C-CoMP)

Elizabeth Kujawinski (WHOI) and Mary Ann Moran (UGA)

Half of the photosynthesis on Earth is carried out by ocean phytoplankton. Their activity is the first step in a complex system that annually withdraws 50 billion metric tons of carbon from the atmosphere, approximately half of total global carbon fixation. Of this, 25 billion metric tons participate in a rapid cycle in which biologically labile molecules are released into seawater and converted back into inorganic form by marine bacteria within hours to days. The central mechanism of this fast cycle is a chemical-microbe network, connecting the production, release, and consumption of dissolved organic carbon by surface ocean microbes; however, the controls on this network, and its links to carbon sequestration in the deep ocean, are not known. Consequently, its sensitivities to changing ocean conditions are also unknown, and responses to future climate scenarios are not predictable. ***Rapid and transformative advances are needed to understand the mechanisms that mediate a quarter of Earth's organic carbon flux, a first-order, time-sensitive question that has never been addressed within a coherent oceanographic framework.***

CENTER MISSION. The Center for Chemical Currencies of a Microbial Planet (C-CoMP) is dedicated to integrating research, education and knowledge transfer activities to reach a mechanistic understanding of surface ocean carbon flux within the context of a changing ocean and through increased participation in ocean sciences. C-CoMP will support science teams that merge biology, chemistry, modeling, and informatics to close long-standing knowledge gaps in the identities and dynamics of organic molecules that serve as the currencies of elemental transfer between the ocean and atmosphere. C-CoMP will foster education, outreach, and knowledge transfer activities that engage students of all ages, broaden participation by the next generation of ocean scientists, and extend novel research integration approaches into complementary academic and industrial communities. Center funding is critical to this mission, uniquely facilitating an open exchange of experiment and model design, methodological and conceptual challenges, results from science, education, and outreach, and collaborations that underpin an integrated science and education community. C-CoMP will embrace best practices regarding ethics and diversity in hiring and in executing meetings and field experiences to achieve a culture of inclusion.

INTELLECTUAL MERIT. Oceanographers were among the first scientists to recognize the central roles of microbial communities and their chemical products in biogeochemical cycles. Admittedly, marine microbes were hard to miss because their photosynthesis forms the basis of nearly all ocean productivity, and their products accumulate in a dissolved organic pool that is equivalent in mass to the atmospheric carbon dioxide reservoir. Yet climate-carbon feedbacks on the marine carbon reservoir are a major, unresolved uncertainty for future climate projections, and evolution of the changing ocean depends directly on microbial responses to climate change, ocean acidification, and other perturbations. C-CoMP will support interdisciplinary science teams working with novel approaches and following open and reproducible science practices to address major outstanding knowledge gaps at the heart of the global carbon cycle: (1) the chemical currencies of surface ocean carbon flux; (2) the chemical-microbe network in the surface ocean; and (3) network sensitivity and feedbacks on climate. C-CoMP will leverage emerging tools and technologies to tackle critical challenges in these themes, in synergy with existing ocean programs (e.g., EXPORTS) and consistent with NSF's Big Ideas surrounding "Rules of Life" and "Convergent Research". C-CoMP science and education themes can be efficiently addressed only through a sustained network of scientists achieving research goals while expanding ocean science literacy and broadening the workforce able to tackle multi-disciplinary problems.

Advanced chemical tools will be employed to isolate and identify molecules produced by marine microbes, together with the genes, transcripts and proteins important for their production and consumption. Data will be interpreted within the context of numerical models that test these molecules' role in chemical-microbe networks. Emerging molecular biology tools will link physiology to function across microbial clades. New informatics tools will leverage existing datasets of marine microbial and

environmental parameters, developing novel ways to merge and visualize heterogeneous data types. Laboratory model systems will validate conceptual and computational models to understand the emergent properties of the chemical-microbe network. Field studies, enhanced by autonomous sampling platforms, will provide important context for scaling up lab- and model-based insights on the chemical-microbe network to the regional and global carbon cycles. By pursuing research on a cellular level with relevance for ecosystem function, C-CoMP discoveries will be applicable to microbiomes in other habitats and host organisms.

BROADER IMPACTS. The ocean affects every aspect of life on the planet, regardless of geographical proximity to the coast. C-CoMP education and outreach activities seek to overcome barriers to ocean literacy and diversify participation in ocean research. The Center will develop (1) initiatives to expand ocean literacy in K-12 and the broader population, (2) ocean sciences undergraduate curricula and research experiences that provide multiple entry points into research, (3) post-baccalaureate programs to transition undergraduates into graduate education and careers in ocean science, and (4) interdisciplinary graduate student and postdoctoral programs that prepare the next generation of ocean researchers. The C-CoMP team includes education faculty who will conduct empirical research on the impacts of education and outreach activities and export successful STEM initiatives to the education community.

Knowledge transfer of C-CoMP innovations, technologies, and research approaches is a central priority. C-CoMP will produce immediately relevant data for other microbiome systems, including an expanded lexicon of molecules important for chemical-microbe networks. The Center's open science research environment will enhance exchange of interdisciplinary research findings in science and education. C-CoMP will partner with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) in an informatics research collaboration to develop data integration platforms valuable to the wider science community. On the broadest societal level, C-CoMP insights into factors that control the chemical-microbe network on environmental scales, and its sensitivity to a changing climate, will be transferred to policy-makers.

ORGANIZATIONS. C-CoMP engages an integrated team of researchers, including those whose field does not naturally intersect with carbon cycle inquiry. Each partner institution brings important expertise and technical capabilities to C-CoMP. The Woods Hole Oceanographic Institution (WHOI) is a leading oceanographic institution with internal project management expertise, data sharing resources, and communications to enhance C-CoMP research, education, and knowledge transfer activities. Facilities and expertise in advanced chemistry and biology will be based at The University of Georgia (UGA), WHOI, Columbia, MIT and Ohio State. Computational expertise for modeling and data mining will be based at The University of Virginia (UVA), the University of Chicago, Boston University, WHOI, and Stanford. WHOI and the Bermuda Institute of Ocean Sciences (BIOS) will facilitate field research at the Martha's Vinyard Coastal Observatory (MVCO) and Bermuda Atlantic Time-series Study (BATS) oceanographic sites, and have demonstrated experience leveraging time-series data for educational and outreach activities. UGA and Boston College will lead the educational responsibilities and coordinate programs among partner institutions and their export to STEM education communities.

LEGACY. C-CoMP will revolutionize the knowledge and technologies to study chemical transformations in microbial systems and understand their outsized impact on elemental cycles. Research will close an important and urgent knowledge gap in the mechanisms driving carbon flow between ocean and atmosphere, with global implications for predictive climate models. C-CoMP will champion open science research and showcase approaches to science where cross-disciplinary collaborations, community engagement, and inclusive practices foster strategic advances in critical science problems and STEM initiatives. With expanded participation in ocean science research and ocean literacy across the US society, the next generation of ocean scientists will more faithfully reflect the diverse US population.

I. PROBLEM DESCRIPTION and CENTER RATIONALE: (Intellectual Merit)

The balance of Earth's carbon stores is being altered drastically. Increased atmospheric carbon dioxide (CO_2) levels are causing changes in weather, growing seasons, and ice stores[9]. Increased CO_2 is making the ocean warmer and more acidic. The ocean carbon cycle is on the front lines of climate change.

Half of the photosynthesis on Earth is carried out in the ocean, with marine phytoplankton fixing 50 Pg (10^{15} g) of carbon each year. By various biological, chemical, and physical processes, some carbon is exported below the photic zone as particulate organic carbon[10]. A small fraction of this downward "biological pump" flux (about 0.5 Pg C year^{-1}) is remineralized in the deep ocean and locked away from the atmosphere for hundreds to thousands of years[11]. Sequestration of carbon derived from particulate organic carbon in the deep ocean therefore creates a critical reservoir that modulates long-term atmospheric CO_2 concentrations, and has been a central focus of modern oceanography research.

Yet an equally significant fraction of annual marine photosynthate is processed through the labile dissolved organic carbon (DOC) pool (Fig. 1), a poorly understood step in the ocean carbon cycle by which upper-ocean microbes return approximately 25 Pg C year^{-1} to inorganic form within hours to days of its fixation[12]. The central mechanism of this process is the production, release, and consumption of bioreactive molecules by ocean microbes. This mechanism's sensitivity to changing ocean conditions is unknown, and its responses to future climate scenarios are therefore not predictable based on current knowledge[13]. ***The goal of our proposed Center for Chemical Currencies of a Microbial Planet (C-CoMP) is to bring rapid and transformative advances to understanding of the bioreactive molecules of the ocean carbon cycle that are responsible for a quarter of Earth's annual organic carbon flux.***

The molecules of the labile DOC reservoir are produced and consumed by diverse microbes in the surface ocean (Fig. 1). Each compound is present in only trace amounts due to the efficiency with which it is recognized and assimilated by microbes, and each is embedded in a much larger pool of nonlabile marine DOC estimated to contain tens of thousands of distinct compounds[14]. Nevertheless, it is through this pool of unknown molecules, turning over once every three days[12], that a quarter of Earth's net primary production is transformed. ***We propose to learn the rules for creating, accessing, and recycling this immense bioreactive chemical space, a subject never tackled within a coherent oceanographic framework.***

The enormous flux of labile DOC through the network of chemicals that link ocean microbes makes even minor changes quantitatively significant on a global scale, with decadal if not annual consequences. First-order, time-sensitive questions include: As the oceans change, will the proportion of carbon moving through labile DOC change? Will an altered chemical-microbe network impact carbon transformation efficiency? Will the proportion of other major elements that link Earth's biogeochemical cycles through dissolved organic carbon be affected? Will shifts in labile DOC formation and flux influence ocean eutrophication, hypoxia, and/or harmful algal bloom events? Will modifications to labile DOC cycling alter the formation and loss of longer-lived, deep-ocean non-labile organic carbon? As the Earth's climate changes, the urgency to close these knowledge gaps is acute. The role of microbially-produced bioreactive molecules as a major route for carbon and energy flux in the global carbon cycle was recognized 45 years ago[15,16], and an understanding of its ecological vulnerabilities is long overdue.

Why is an STC required to address this scientific challenge, and why now? Surface ocean carbon cycling is primed for discovery. First, C-CoMP is synergistic with federal science investments, such as NSF's and NASA's investments in EXport Processes in the Ocean from Remote Sensing (EXPORTS). Understanding the downward flux of particulate organic carbon from net primary production and resulting carbon storage in the ocean interior is the EXPORTS side of the coin. C-CoMP focuses on the flip side – the controls on the release and remineralization of bioreactive dissolved organic carbon at the ocean surface. Both processes are globally significant and identifying the connectivity between them is critical. Second,

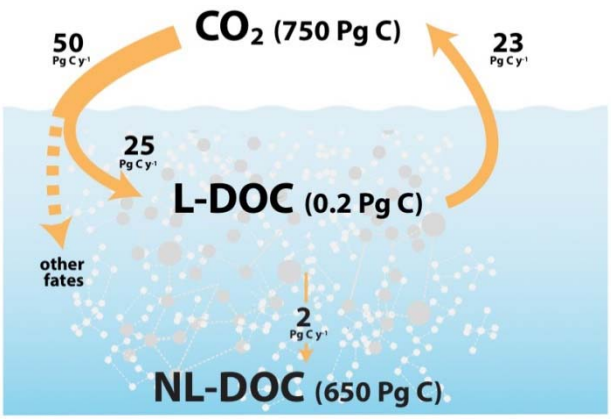

Figure 1. Metabolites in the labile dissolved organic carbon pool (L-DOC) are synthesized by phytoplankton and other microbes and released into seawater through exudation, leakage, and viral lysis. They are consumed primarily by heterotrophic bacteria within hours to days, as bacterial respiration rapidly converts organic carbon back to inorganic form. Equilibration of the inorganic carbon reservoir with atmospheric CO_2 effects a net transfer of about half the oceanic primary production through the L-DOC pool and back to the atmosphere. A small fraction of L-DOC is converted to non-labile DOC (NL-DOC) and participates in long-term carbon storage in the ocean. Primary production not converted to L-DOC is processed as particulate organic carbon.

C-CoMP explicitly leverages recent advances in analytical chemistry, molecular biology, numerical modeling, and big data integration. Supported by investments from federal programs and private foundations, capabilities now exist to assess chemical reactions on a global scale, constrain cellular network and ecosystem models based on substrates and catalysts, leverage modern sequencing datasets to ocean-level understanding, and employ microbes to identify keystone chemicals[17-19]. Third, C-CoMP engages NSF's Big Ideas. The Center addresses "the rules of life" through a cellular level focus on processes with oceanic-dimension relevance, and through discoveries broadly applicable to microbiomes in other habitats and host organisms. It addresses NSF's goal of stimulating "convergent research" through a focus on a major scientific question of significance to Earth's future habitability, and integration of multiple disciplines into a common research framework.

Fourth, generating transformative advances in our understanding of carbon flux in the global ocean requires integration at scales not possible through standard funding opportunities. Our Center will provide a critical platform to: (1) structure an interdisciplinary and highly collaborative team for idea exchange, scalable resources, and integrative data analysis and synthesis; (2) leverage the distinctive perspectives of laboratory model systems, ocean observatories, and numerical approaches to catalyze novel discoveries; (3) enable investigations that span orders of magnitude of scale, from molecules, to microbiomes, to global biogeochemical cycles; (4) train diverse early career scientists to integrate across scientific and informatic disciplines; and (5) build resources for students of all ages to connect to the "ocean microbiome" whose functions maintain the planet. C-CoMP will establish a new paradigm for the chemical currencies within microbiome networks and their impact on global elemental cycles, while providing a diverse framework for societal engagement in ocean science.

II. CENTER RESEARCH OBJECTIVES:

C-CoMP will support 14 senior personnel working in interdisciplinary teams with novel approaches and open science practices to resolve major outstanding knowledge gaps ("Science Themes") at the heart of the global carbon cycle: (1) the chemical currencies of surface ocean carbon flux; (2) the chemical-microbe network in the surface ocean; and (3) network sensitivity and feedbacks on climate (Fig. 2). In Science Theme 1, first-order approaches will query the chemical environment created by cycles of microbial life and death, employing novel strategies and computational frameworks to identify and quantify the molecules driving carbon flux. In Science Theme 2, the dynamics and evolution of microbial ecological networks will be explored to determine forces shaping production and exchange of chemical currencies. In Science Theme 3, vulnerabilities of the chemical-microbe network mediating surface ocean carbon flux in a changing climate will be addressed. C-CoMP scientific themes are broadly acknowledged as stubborn knowledge gaps within oceanography, resistant to study since being recognized as a central component of Earth's carbon cycle[15,16,20,21] and not yet integrated into global earth systems models.

Given their complexity and urgency, the C-CoMP themes can be efficiently addressed only through a sustained network of scientists achieving research goals while expanding ocean science literacy and broadening the workforce able to tackle multi-disciplinary problems. Ocean ecosystems require understanding of both the principles shaping microbial communities and their sensitivities to external perturbations. These forces are fundamental to the biology and chemistry of varied microbial systems, both natural and human-built. Thus, while the surface ocean is C-CoMP's focus, the relevance of our efforts to understand the metabolic currencies of microbial interactions extend well beyond oceanography.

C-CoMP is structured on three scaffolds that underpin research and link to education and outreach efforts (Fig. 2). At the center are the *Science Themes*, representing critical knowledge gaps in metabolite-driven carbon cycling that governs the fate of a quarter of Earth's primary production (in other words, the knowledge we seek). The *Critical Challenges* scaffold represents current technological and conceptual roadblocks to knowledge building (why the science is lacking). The *Enabling Tools and Technologies* scaffold represents the emerging technologies to be created or repurposed to resolve scientific roadblocks (why major knowledge advances are now possible).

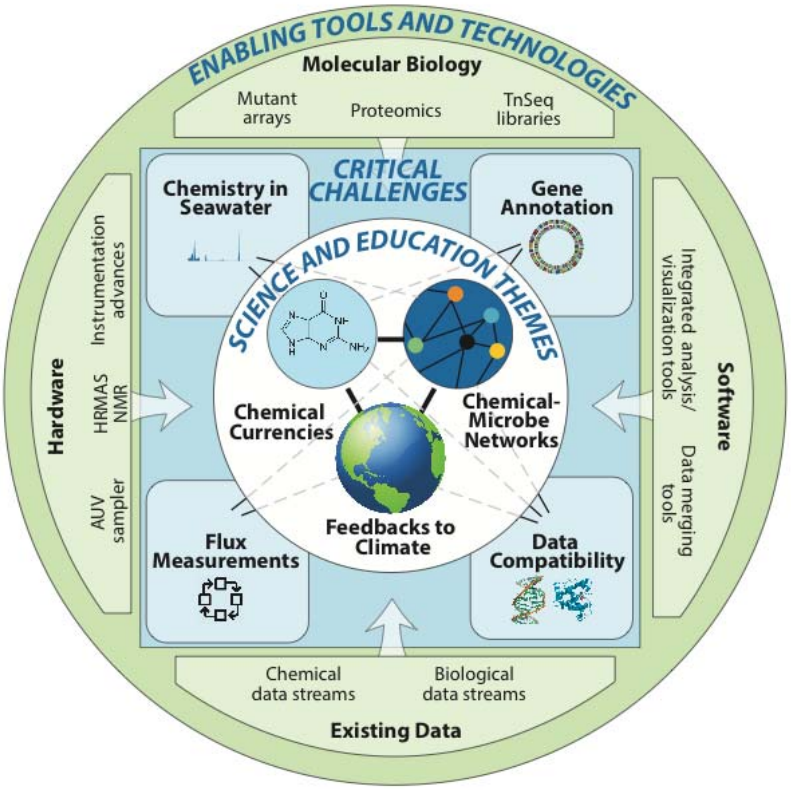

Fig. 2. C-CoMP's three scaffolds. ***Science themes*** (center circle) represent research and education efforts at the heart of the ocean carbon cycle. Addressing these themes has been hindered by long-recognized science and technology ***critical challenges*** (blue square) that are now solvable with ***enabling tools and technologies*** (outer green circle) in an interdisciplinary setting.

We first introduce the *Critical Challenges,* followed by the *Enabling Tools and Technologies* that will make progress possible. Last, we describe the *Science Themes* and strategies to tackle them.

A. The Critical Challenges *("why the science is lacking")*

Major gaps in knowledge of controls over the rapid remineralization of a large fraction of Earth's fixed carbon, despite decades of research, can be traced to four primary scientific/technological roadblocks. Each roadblock impacts progress in two or more Science Themes (Fig. 2), and each has emerged unfailingly in microbial oceanography research for over four decades. Our collective experience tells us that solutions to these critical challenges will be highly transformational, not only in ocean sciences but in microbial-based research across life and physical sciences. Our collective experience also tells us that progress in these areas is now within reach given a coordinated, integrated, and interdisciplinary team assembled through the STC framework.

Critical Challenge 1: Chemistry in Seawater – The seawater organic matter pool is conservatively composed of tens to hundreds of thousands of distinct organic compounds[14], most present in only trace

amounts (femto- to pico-molar) and with very short half-lives (minutes to hours). *Surface ocean metabolites are phenomenally complex and stubbornly resistant to study – and yet are central to evaluating the current state of the ocean and tracking its response to changing climate*[13]. Methods to measure small, polar metabolites in seawater[22,23], increase sensitivity of chemical detection[24], and populate ecologically-relevant metabolite databases are essential for progress.

Critical Challenge 2: Gene/Protein Annotation – Discovery of microbial gene function through traditional strategies is laborious and low-throughput. High-throughput pipelines that replace one-gene-at-a-time annotation strategies are among the most important routes to characterizing the chemical-microbe network that mediates global-scale carbon flux. *There is no debate that a scalable strategy to solve the gene/protein annotation bottleneck will drive revolutionary advances in microbial oceanography, and touch virtually all fields in biology and chemistry,* but a concerted and interdisciplinary effort is needed to do so. Expertise in chemistry, biology, computation, modeling, and technology are essential to solving the puzzle of microbial gene and protein function in the ocean.

Critical Challenge 3: Flux Quantification – While labile metabolites make up only 0.03% by mass of the DOC and 0.0005% of total ocean carbon, the flux through this pool accounts for close to 50% of the annual CO_2 exchange between the ocean and atmosphere. With few exceptions, *methods to measure fluxes of metabolites through the chemical-microbe network are not available*, not at the bulk level and definitely not at the temporal and micron-sized spatial scales experienced by microbes in the ocean[25]. Without quantitative measures of rates and controls on carbon flux and transformation in the surface ocean, some of the largest uncertainties in the global carbon cycle cannot be addressed.

Critical Challenge 4: Data Compatibility – Integration of new and existing data sources across biology, chemistry, and biogeochemistry is central for our scientific mission and for computational empowerment in education and community outreach. *Yet there is no computational framework to co-mine these data sets and address the biochemical and physiological dynamics that stretch from microbial cells to biogeochemical and evolutionary dynamics of ecosystems.* A common scientific framework for data types that are heterogeneous across time, space, and levels of biological and biogeochemical organization is a major impediment to progress in understanding the rules of life.

B. Enabling Tools and Technologies ("why major knowledge advances are now possible")

The collective expertise represented by C-CoMP scientists (Table 2, below) covers critical areas of informatic and technological advances required to address the structure and vulnerabilities of the surface ocean carbon cycle. The Center will explicitly invest personnel and resources in these areas, concentrated in Years 1 and 2, to spark rapid progress toward our science goals. The ET&Ts serve as key linkage points between science and education/outreach activities (Fig. 7, below).

(ET&T1) ***Enabling Tools and Technologies 1: Ocean data*** – Data that will advance our understanding of labile marine DOC cycling are already being generated at tremendous rates, and even larger and more informative data streams are expected. Existing data sources cover ocean biology through isolate, single-cell, and metagenome-assembled genomes, metagenomes, transcriptomes, metatranscriptomes, and metaproteomes (e.g., NCBI, JGI-IMG, KBase GTDB, GORG-BATS, BATS, MVCO, HOT, Tara Oceans, Ocean Sampling Day, iMicrobe, Ocean Protein Portal), ocean chemistry (e.g., Ocean Observatories Initiative data, GEOTRACES), and environmental parameters (e.g., BATS, HOT, Argo, and Ocean Observatories Initiative data). These public data facilitate the immediate development of technically-feasible data compatibility and analysis advances. C-CoMP is supporting a full-time Digital Coordinator co-located with WHOI-based Biological and Chemical Oceanography Data Management Office (BCO-DMO) staff to ingest and integrate existing and forthcoming data in support of C-CoMP science.

(ET&T2) ***Enabling Tools and Technologies 2: Software*** – New oceanographic, bioinformatic, and modeling software tools responding to the interdisciplinary needs of microbiome science are emerging, and these are critical to comprehensively extract insights from ocean data. C-CoMP scientists have expertise in creating modular and scalable open-source software tools that support heterogeneous data analyses; these include: anvi'o[26], an open-source bioinformatic analysis and visualization platform that enables applications of genome-resolved metagenomics, pangenomics, phylogenomics, and microbial population genetics to marine systems over temporal and spatial scales (*Eren*; Fig. 3); the Ocean Protein Portal, a data sharing platform that interrogates ocean metaproteomic data for abundance and taxonomic origin (*Saito*); the Flux Balance Analysis (FBA)-based metabolic modeling platform COMETS that performs computer simulations of microbial metabolism in spatially structured communities[27] (*Segrè*; Fig. 3). We will spur data integration and software development in the first two years of C-CoMP by supporting a programmer and a postdoctoral researcher (*Eren, Saito*).

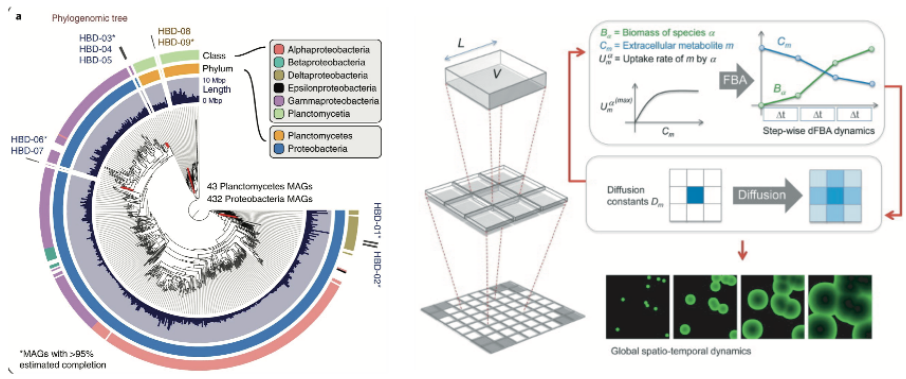

Fig. 3. Enabling software. Left: anvi'o phylogenomic analysis of 475 metagenome-assembled-genomes from Tara Oceans using phylogenetic marker genes; outer rings indicate genome size and taxonomy. Right: Key steps of COMETS simulations.

(ET&T3) ***Enabling Tools and Technologies 3: Molecular Biology Tools*** – With the emergence of model organisms in ocean science, molecular biology tools open up access to experimental manipulations of microbes and microbial communities. For example, loss-of-function mutant libraries[6,28], CRISPR-cas9[29], TnSeq[30], and high-throughput functional phenotyping[31], among others, offer enormous unrealized potential for probing the foundation of the microbial carbon cycle (e.g., gene and protein annotation, functional links between microbes, metabolite sources and sinks), and for explicitly testing theories emerging from laboratory, field, and modeling efforts in microbial strains and model communities (e.g., mechanisms of carbon excretion, carbon use efficiency, feedbacks across levels of biological complexity).

Initial efforts will focus on a molecular tool with immediate technical feasibility: arrayed mutant libraries of model bacterial strains. In a novel iteration of this technology, saturated random transposon mutant libraries with barcoded transposons[32] will be arrayed in 384-well plates, and low-cost PCR amplification used to link the transposon insertion point of each mutant (i.e., the gene that is disrupted) with its physical location in the plate. The resulting arrayed library includes mutants for virtually all non-essential genes, and these can be individually accessed for experimentation. In Year 1, C-CoMP will invest in 4-6 arrayed mutant libraries for bacteria selected for experimental and modeling focus (see ST1b, below; Fig. 5) through a collaboration with Dr. Chris Reisch (University of Florida). Transposon library-based methods in general (e.g., TnSeq[32], DubSeq[33]) are ideal for marine isolates because they enable high-throughput screening for function yet do not require full genetic systems; these methods will be fully exploited in experimental frameworks to gain insights into microbial physiological and ecology.

(ET&T4) ***Enabling Tools and Technologies 4: Hardware*** – Advances in instrumentation for ocean sampling and chemical analysis are galvanizing surface ocean carbon cycle research. C-CoMP will invest in the novel autonomous underwater vehicle (AUV) Clio that overcomes the perennial problem of low

microbial biomass in ocean sampling through a large volume filtration system. Clio can conduct high-resolution vertical sampling of 100s of liters of seawater, permitting simultaneous chemical, biochemical, and microbial measurements (Fig. 4). Clio will be modified to add field collection capabilities for metabolite analysis (Year 2), and will undertake coordinated multi-'omics sampling in C-CoMP research cruises (Years 3, 4; *Saito*, *Clio team*).

Improved chemistry technologies are opening up access to seawater DOC pools. New nuclear magnetic resonance spectroscopy (NMR) approaches leverage advances in probe technology[34] for extremely high sensitivity in ^{13}C measurements with less salt interference than conventional ^{1}H NMR[35]. High Resolution Magic Angle Spinning (HRMAS) NMR developments offer real-time, untargeted measures of natural abundance substrate flux into microbial cells[36], with long observation windows (potentially up to several days) and minimal sample preparation (*Edison*). Ultrahigh resolution mass spectrometry (e.g., Orbitrap MS) detects hundreds of molecules per analysis, providing rapid fingerprints of chemical complexity in environmental samples. When combined with liquid chromatography (LC), assessments of compound structure, molecular mass and abundance can be achieved in short (<15 min) analyses. Low-molecular weight or highly polar molecules can be isolated after chemical derivatization, further opening the analytical window to encompass a broad range of chemical currencies in seawater. Detection limits for these analyses are in the pM-nM range for quite small (10 ml) sample sizes (Fig. 4), making seawater metabolite measurements tractable even in oligotrophic ocean settings (*Kujawinski*).

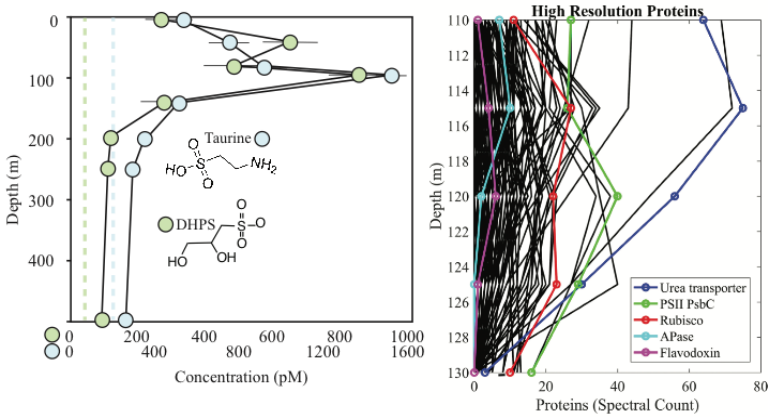

Fig. 4. Enabling hardware. Left: Concentrations of metabolites (organic sulfur compounds dihydroxypropanesulfonate (DHPS) and taurine) detected at pM levels in oligotrophic seawater at BATS using new LC-MS methods. Right: Clio-generated metaproteomic profile at BATS highlighting *Prochlorococcus* proteins enhanced near the chlorophyll maximum (PsbC, Rubisco, flavodoxin) or at the surface (urea transporter and alkaline phosphatase), consistent with patterns of nutrient limitation.

C. Science Themes *("the knowledge we seek")*

The Center structure supports synergistic and simultaneous efforts over three science themes, enabled by emerging solutions (ET&Ts, above) to roadblocks (Critical Challenges, above) that have hindered understanding of the surface ocean carbon cycle. The ability to update and realign strategic research priorities across the themes as research matures is critical for success, and will be enabled through C-CoMP communication avenues and open data policies (see Center Management, below).

(ST1) ***Science Theme 1: Chemical Currencies of Surface Ocean Microbes***

Molecules synthesized and consumed by ocean microbes are the currencies by which elements are transferred between ocean reservoirs. These bioreactive molecules are released into seawater through mechanisms such as exudation from microbial cells[37] and cell lysis within the microbial food web[38,39]. They are consumed by hundreds of heterotrophic bacterial species, subspecies, and ecotypes that efficiently recognize and assimilate organic matter[15], often within minutes to hours of release. At a bare minimum, released metabolites take the form of monocarboxylic and dicarboxylic acids[40], glycerols and fatty acids[40,41], nitrogenous compounds such as taurine, choline, sarcosine, polyamines, methylamines, ectoine, and amino acids[42], one-carbon compounds[41,42], sulfonates and sulfonium compounds[17,43], and

polysaccharides[44]. The quantity of metabolites released directly by marine phytoplankton is estimated to account for ~30% of primary production, although this value is poorly constrained and variable across species and growth conditions[44,45]. The exchanged metabolites embody an extraordinary diversity of functions that can be summarized loosely into *substrates* that are produced and consumed in quantities that sustain growth and reproduction; *co-factors* that are released in small quantities to mediate reactions, such as vitamins[46] and siderophores[47]; and *signals* that act as coded chemical information passed between microbes to alter function. Some will fall into multiple categories. The entirety of microbial production and consumption of labile DOC sums to the largest annual carbon flux through the ocean and atmospheric reservoirs, yet we have only the most rudimentary knowledge of the identity and fate of the chemical players. *C-CoMP will identify the central molecules or molecule classes that move carbon and other elements through the surface ocean or have key regulatory roles in these fluxes.*

ST1a. Strategic chemical identification – Bioreactive molecules serving as microbial currencies in the surface ocean are not easily identified or quantified. They are dilute (~1,000,000-fold lower in concentration than salts), structurally complex, and require extraction or modification protocols prior to analysis. C-CoMP will leverage recent advances in analytical chemistry that offer increased sensitivity, reproducibility, and resolution (LC-MS; *Kujawinski* and NMR; *Edison*), with a focus on new derivatization protocols and advances in extraction and separation tools that further extend the analytical power of LC-MS and NMR in a seawater matrix. Derivatization methods for LC-MS and concentration methods for NMR now enable untargeted quantification of microbial metabolites in seawater[22,23]. HRMAS is an NMR method for tracking consumption of molecules by living cells in real time[36]. For all microbial metabolite measurements, we will use data science approaches to extract an ontology of substrates, co-factors, and signals based on common patterns in structure, occurrence, and fate.

ST1b. Focused metabolic annotation – A critical companion to chemical identification is an understanding of the biology that generates and consumes bioreactive molecules. Gene and protein expression assays will be used to pinpoint the microbial molecules being actively cycled (*Dyhrman, Moran, Saito*). For example, expression patterns of microbial transporters suggest which molecules are bioreactive within complex DOC pools[6] and transposon mutant libraries act as sensors of metabolite production and consumption through loss of those functions (*Moran*). Computational models identify new chemical targets and integrate gene, protein, and metabolite data to detect the corresponding chemical reactions (*Covert, Segrè*). For example, linearly optimized flux-balance analysis, by exploring the optimally efficient states of these networks, can determine by-product secretion rates[48,49] and predict costless metabolites supporting cooperative microbial interactions[50]. Phylogenomic analyses informed with metagenomic and metabolomic data uncover pathway diversification across evolutionary timescales with metabolic consequences[1]. For example, metabolic analysis of the dominant marine phytoplankter *Prochlorococcus* hypothesized an evolutionary increase in excretion rates of the low molecular weight organic acids glycolate, pyruvate, and citrate[1]. Iterative and complementary laboratory and modeling studies focusing on a selected suite of autotrophic and heterotrophic taxa (Fig. 5) will feed into model systems research (ST2a, below) and cell- and community-scale modeling (ST2c, below) to identify key bioreactive molecules of the surface ocean.

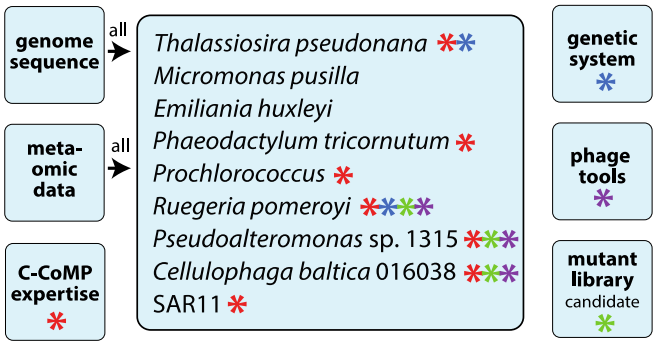

Fig. 5. Candidate marine microbes for focused metabolite and gene identification will be chosen based on phylogenetic affiliation, ecological role, and available data/tools[1-8].

ST1c. Chemical and biological data integration – The C-CoMP metabolite ontology will be integrated with existing data platforms (e.g., GNPS[51], MAGI[52]). Improved predictability of functions for poorly known or newly discovered molecules[53] will support extrapolation to biogeochemical roles. Data-driven strategies will identify overlaps between "pathways without observed metabolites" and "observed metabolites without known pathways"[52]. C-CoMP will develop novel computational strategies (*Eren*) to support data integration, such as machine learning to uncover relationships between chemical structure and biological reactivity, and computational prediction of enzymes or transporters that co-occur with metabolites.

ANTICIPATED ACCOMPLISHMENTS: Science Theme 1 will build the first ontology of the bioreactive molecules directing carbon through the surface ocean, supporting ST2 and ST3 (below). Addressing this major knowledge gap is possible by the entrainment of investigators with new domain specialties into carbon cycle science, and leveraging nascent tools and data integration approaches. Early career scientists will be trained in interdisciplinary research (see ET2-4, below). "Best practice" data formats with broad applicability in microbial ecosystems will be developed. Software platforms for co-analyzing diverse data streams at scale will be made accessible to the broader science community (see Knowledge Transfer).

Science Theme 1 research will be jumpstarted by the enabling tools and technologies (ET&T) for ocean data ingestion (e.g., extracting patterns in structure vs. distribution of marine metabolites), software (e.g., computational prediction of microbial genes that co-occur with bioactive molecules), molecular biology (e.g., identifying metabolites by new functional gene annotations), and hardware (e.g., identifying metabolites through advances in chemical instrumentation).

(ST2) ***Science Theme 2: Rules of the Chemical-Microbial Network***

The dynamic chemical-microbe network controls labile marine DOC synthesis, release, and transformation and determines the fate of 25% of annual global carbon fixation. Factors that influence the residence time of this newly-fixed carbon include diversity in chemicals and organisms; diel, seasonal, and annual cycles of the chemical and biological players; and variation in physical forces acting over time and space at scales from micron to ocean basin. We will develop and advance a theoretical framework of why and how cells and microbial communities produce and consume labile DOC. *C-CoMP will determine the rules of the chemical-microbe network that regulate the massive transfer of carbon between global reservoirs.*

ST2a. Model systems – Laboratory model systems will be a cornerstone for building and validating conceptual and computational models to capture the emergent properties of the chemical-microbe network. Experimental systems that span ranges in throughput, scale, homogeneity, and microbial and viral components will be leveraged to build quantitative insights into molecule flux and the factors that control them (*Dyhrman, Edison, Kujawinski, Moran, Saito, Sullivan*). For example, we will employ microbial co-cultures and synthetic communities[6,54], and high-throughput arenas that characterize microbial interactions under shifting chemical regimes[31,54]. Model systems will be analyzed using tools in 'omics (genomics, transcriptomics, proteomics, metabolomics), isotope labeling, and genetic manipulations of microbial phenotypes. We will address relationships among chemical features (functional groups, heteroatoms, element stoichiometry) and biological reactivity, microbial uptake and release rates, microbial growth versus efficiency trade-offs, and viral rewiring of metabolism[38,55] to identify predictable rules of the network. Experiments focusing on representative autotrophic and heterotrophic taxa (Fig. 5) will be the focus here, and will link directly to cell- and community-scale modeling (ST2c, below). They will be included in undergraduate course curricula and research experiences (see ET2, below).

ST2b. Field observations – Studies of simplified components of the chemical-microbe network will complement those at the environmental scale accessed through field-based research. How individual strategies and tradeoffs from model system components are manifested in the ocean will be a focus of this effort. Partnerships with two time-series stations in the Atlantic Ocean, the open-ocean Bermuda Atlantic

Time Series Study (BATS; established in 1988) and the coastal Martha's Vineyard Coastal Observatory (MVCO; established in 2000) will support a conceptual and experimental scale-up. BATS and MVCO have strong annual cycles in ocean physics, nutrient regimes, and microbial communities within which to situate field experiments. Each program has the benefit of ongoing multi-disciplinary data collection through core programs and collaborative research by affiliated investigators, including C-CoMP scientists (*Bates, Dyhrman, Kujawinski, Saito*). Accessibility from Bermuda (BATS) and Woods Hole (MVCO) make these ideal partners for field research, and they offer platforms for comparative analyses to other well-studied ocean systems (e.g., San Pedro Ocean Time series (SPOT), Hawaii Ocean Time-series (HOT)).

Methodology suites similar to those for the model system studies will be used in the field to develop and test hypotheses about chemical-microbe interactions and identify molecules and microbes that best predict features of carbon flux[18]. Autonomous oceanographic instruments will be used to measure concentrations and fluxes of target bioreactive molecules at the necessary spatial and temporal coverage[56] and for collecting targeted surface ocean samples for 'omics research[57]. The AUV Clio (ET&T4, above) will be outfitted with a new metabolomics sampling module to complement existing tools for high-throughput water column sampling of proteins, DNA, and RNA[19] (*Saito, Clio team*) to generate coordinated measures of seawater metabolomics and gene and protein inventories (*Bates, Dyhrman, Kujawinski, Moran, Saito, Sullivan*). Clio samples will build an integrated dataset of the natural chemical currencies produced and consumed in a complex ecosystem, querying many co-occurring biomolecules in a seasonally dynamic carbon cycle. Undergraduates in capstone experiences and Bridge to PhD (B2P) Fellows (see ET2 and 3, below) will be involved in method testing for MVCO field surveys and two C-CoMP cruises (Years 3 and 4).

ST2c. Chemical-microbe network models – C-CoMP will develop **data-driven models** that scale and integrate heterogeneous data, both new and existing from synthetic and natural microbial communities in the North Atlantic and other ocean basins (e.g., BATS, HOT, MVCO, Tara Oceans, GORG-BATS). Data will be searched for patterns, dependencies, and inconsistencies that offer insights into the roles of microbial metabolites. The models will consist of networks of correlations in abundance, occurrence, and other metrics of relatedness[58] connecting microbes, microbial genes, and metabolites. This framework will be enhanced with machine learning algorithms to discern microbe-microbe and microbe-metabolite interaction patterns[59-61] to generate and test hypotheses about underlying principles (*Segrè*).

C-CoMP will develop **mechanistic models** that translate 'omics information and environmental boundary conditions into quantitative predictions of expected phenotypes. These models will be compared with observations, enabling iterative assessment of how metabolite composition and exchange shape communities. Given that chemical-microbe networks can span many levels of biological organization and multiple orders of magnitude in space and time, tradeoffs are necessary between comprehensive, high-detail approaches with many parameters, and simplified, few-parameter frameworks that capture broad patterns but have less nuanced predictions (Fig. 6). Our team has deep expertise across these multiple scales and model types. To facilitate comparisons and integration at the highest level, we will focus initially on representative autotrophic and heterotrophic taxa for which genomic data and genetic and phage tools are available, and for which physiological constraints will be investigated in model organism and field studies (Fig. 5; ST2a,b, above). Genome-scale reconstructions of the metabolic networks of these microbes[48,62] (*Covert, Segrè*) will link intracellular circuits with influx/outflux of metabolites across cell boundaries, as the basis for implementing flux balance analysis (FBA) models. FBA models assume a steady-state and do not require detailed intracellular kinetic parameters, but nonetheless provide realistic, testable estimates of uptake/secretion fluxes based on environmental limitations. C-CoMP expertise in the biology and ecology of representative microbes (*Braakman, Dyhrman, Moran, Sullivan*) will support model development and testing.

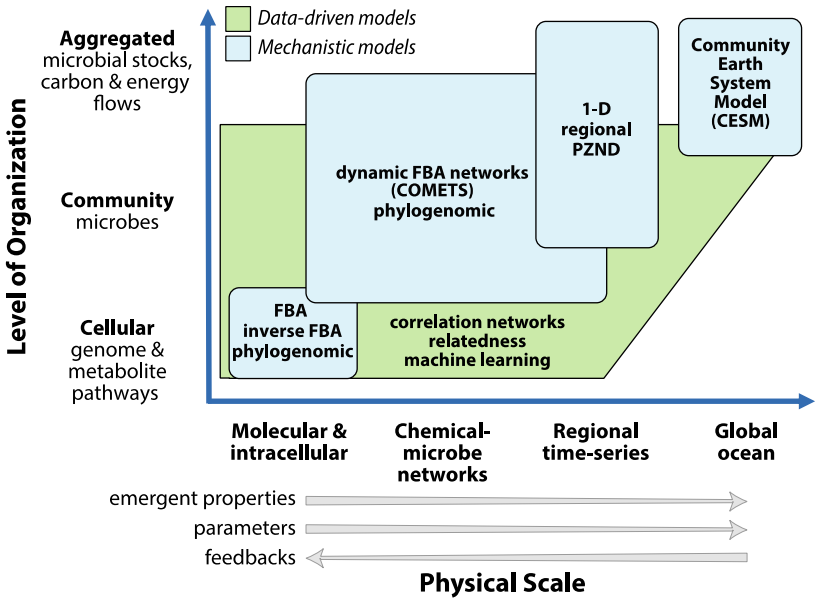

Fig. 6. Modeling the chemical-microbe network over physical scales from cells to global, and organizational levels from cellular to aggregated systems. Mechanistic models (blue) and data-driven models (green) cover the spatial and organizational space. Gray arrows show key links between model scales.

FBA models will be used in different modalities at the cellular and community scales (Fig. 6). First, based on the hypothesis that microbes have evolved to produce their own biomass with high efficiency, standard biomass optimizations to predict growth rates and metabolite uptake/ secretion (*Covert, Segrè*) will be validated using measurements from model system experiments, field studies, and data-driven models (ST2a,b, above). In the second modality, recently developed algorithms for inverse FBA (invFBA)[63] will test implementation of both standard biomass optimization and more complex metabolic "goals" as drivers of network function, e.g., excretion of metabolites that mediate interactions. Pathways leading to the largest gains in quality of the fit to experimental data will suggest rules governing cell excretion of dissolved organic carbon.

The third parallel modality will use "dynamic FBA" (dFBA) to include exchange between microbial cells, explicitly modeling the dynamics of multi-species communities and predicting, as emergent properties, interactions arising from competition or metabolite exchange. Our COMETS (Computation of Microbial Ecosystems in Time and Space) dFBA platform[27] enables community simulations with spatial resolution (e.g., taking into account light and nutrient gradients) in which cells communicate via diffusion and local experience rather than bulk conditions[64]. As with implementations of FBA models for individual microbes, these community models will provide predictions of changes in environmental metabolite abundances that can be tested or validated through the model system experiments (ST2a), field studies (ST2b), and data-driven models (above). Further, FBA-based models can estimate carbon use efficiency [65], a key parameter of surface ocean models, and can serve as a mechanistic bridge between field data from natural microbial communities and laboratory data from simplified model microbial communities. Motivating questions include: What governs labile DOC release from different cell types? Why does bacterial DOC uptake rate, respiration, and carbon use efficiency vary by cell type and metabolite? How do nutrient limitation, temperature, and pH alter the chemical-microbe network?

FBA models will be complemented with phylogenomic models aimed at understanding the forces and tradeoffs shaping metabolic pathways by evolutionary reconstruction. For the same representative taxa (Fig. 5), phylogenomic models will map metabolic pathways inferred through genomics and metabolomics onto phylogenies to generate hypotheses about metabolic innovations leading to metabolite production and consumption (*Braakman*). These hypotheses can be directly tested with microbial and microbial/ viral model systems (*Dyhrman, Moran, Sullivan*). For example, phylogenomic reconstruction led to the hypothesis that *Prochlorococcus* pathways for excreting organic acids evolved as a metabolic strategy for assimilating limited nutrients[1]; subsequent laboratory experiments confirmed this link[66]. Phylogenomic model results will provide constraints on objective functions of FBA models, and a framework by which metagenomic and metatranscriptomic data can be leveraged to address metabolite

exchange in the ocean. For example, ongoing work by *Braakman* and *Kujawinski* identified pathways mediating cross-feeding of nitrogen-rich nucleotides between *Prochlorococcus* and the heterotrophic bacterium SAR11, two of the most abundant microbes in the surface ocean; metagenomic data showed global patterns in these pathways consistent with nutrient availability. Connections between environmental constraints and metabolite release are directly relevant for regional- and global-scale models (ST3).

ANTICIPATED ACCOMPLISHMENTS: Identifying the mechanisms that underpin the vast and dynamic chemical-microbe network in the surface ocean is critical to understanding the controls on labile DOC synthesis, release, and transformation. C-CoMP will define the rules that govern these carbon cycle transformations and quantify carbon flux through the chemical-microbe network for parameterization of regional- and global-scale biogeochemistry models (ST3, below). New approaches to analyze and integrate biochemical and genomic data sets will be fostered, led by team members who have pioneered informatic and modeling approaches in oceanography and biological science. Theory will emerge that addresses metabolite production and release, explicitly representing costs and benefits in the context of environmental stressors.

ET&Ts will jumpstart ST2 research through ocean data ingestion (e.g., data-driven exploration of chemical-microbe co-occurrence patterns), software (e.g., visualization of multi-'omics model system and field data), molecular biology (e.g., experimentally testing predicted chemical-microbe linkages), and hardware (e.g., targeted field samples for hypothesis testing).

(ST3) ***Science Theme 3: Network Sensitivity and Feedbacks on Climate***

Oceanographers were among the first scientists to recognize the central biogeochemical roles of microbial communities and their chemical products, with initial descriptions of carbon transfer through ocean microbes appearing over 40 years ago[16]. Admittedly, microbes were hard to miss in the ocean because their CO_2 fixation forms the basis of nearly all ocean productivity, and because the accumulation of their products has created a pool of non-labile DOC that stores as much carbon as the atmospheric CO_2 reservoir[12]. Yet climate-carbon feedbacks on the marine carbon reservoir are a major, unresolved uncertainty for future climate projections, and evolution of the changing ocean depends directly on microbial responses to temperature increases, ocean acidification, and other environmental perturbations. *C-CoMP will uncover the principles that control the fate of labile DOC on regional to global scales and assess sensitivity to climate feedbacks.*

ST3a. Linking microbe-DOC models across scales of organization – Data interrogation and hierarchical chemical-microbe network models (ST1 and ST2) are the foundations for exploring microbial carbon cycle processes that are most sensitive to environmental perturbations. Presently, global upper-ocean biogeochemical dynamics are captured through phytoplankton-zooplankton-nutrient-detritus (PZND) modules that approximate the details of intracellular metabolic networks and extracellular diffusive transport with macro-level, bulk functional forms (e.g., Michaelis-Menten equations for substrate uptake and linear terms for DOC excretion). Current allocation approaches in PZND models use simplified rules to track substrates and energy through core metabolic pathways (e.g., carbon acquisition, photosynthesis, Calvin Cycle fixation, and nutrient uptake) and to allocate carbon and nutrients to functional pools (e.g., photosynthetic apparatus, ribosomes and biosynthetic machinery, nutrient acquisition proteins; e.g., ref[67]). As environmental conditions evolve, schemes are optimized for maximum cell growth.

There is conceptual convergence between the allocation approaches in PZND models and the aggregated output from steady-state and dynamic FBA models from ST2c; this nexus will be exploited to formulate new resource allocation models for marine microbes. We will employ a systematic approach to model development and evaluation, one that is grounded in theory and encapsulates rules of the chemical-

microbe network (*Braakman, Covert, Doney, Segrè*). We will search for predicted emergent feedbacks across levels of biological organization and characterize functional responses of phenotypic metrics – carbon excretion, carbon use efficiency, and carbon transformation rates – to improve parameterizations[68]. Starting with highly simplified models, the development path will follow an iterative process of coarse-graining to assess trade-offs between added explanatory power and computational costs as more detailed processes, metabolites, and emergent properties are included (*Doney*). Identifying greatest gains in explanatory power will feed back to inform theories/ models/ experiments relevant to ST1 and ST2 (Fig. 6). Representations of chemical-microbe interactions will focus on responses to warming, stratification, nutrient limitation, and acidification. Some of these effects (e.g. stratification) can be gleaned from spatially resolved dFBA (COMETS). Other effects (warming, acidification) are at the frontier of metabolic modeling research, but will be guided by FBA that considers reaction free-energies as a function of temperature and pH[69]. We will determine the sufficient, but computationally tractable, number of aggregated microbe groups and labile metabolites to include in carbon models; identify slow-evolving community metabolic pathways that must be treated explicitly (prognostically) compared to faster processes and quasi-steady state stocks that can be approximated implicitly from overall system state; and determine functional forms, saturation, and limitation parameters.

ST3b. Labile DOC in a regional biogeochemical perspective – Translation of C-CoMP science to regional scales requires that hard-fought details of the chemical-microbe network from ST2 be simplified, although not trivialized, in the PZND-biogeochemical modules. C-CoMP efforts will build on the multi-species models from FBA COMETS (ST2c) explicitly addressing the composition, characteristics, and rates of metabolic currency exchange between microbes that predict outcomes of the fast and outsized carbon loop linking the surface ocean and atmospheric reservoirs (*Braakman, Covert, Doney, Segrè*). These model approaches will be evaluated in regional 1-D biogeochemical simulations against field data collected at the ocean time-series sites (*Bates, Edison, Kujawinski, Saito*). Results will in turn be used to predict the distribution of metabolic pathways and functions in the ocean, which can be examined using metagenomics.

Mechanistic simulations from both microbial network (COMETS) (ST2c) and PZND-biogeochemistry approaches will be tailored for the BATS and MVCO microbial systems and leveraged with long-term oceanographic and microbial observations generated at these sites. Physical and biogeochemical drivers for BATS and MVCO will be incorporated into the regional models by extending a regional test-bed, data assimilation system developed for other marine environments and time-series locations (e.g. refs[70,71]).

ST3c. Labile DOC in a global biogeochemical perspective – The well-recognized downward export of particulate organic carbon (the "biological pump") to deep ocean waters and sediments is globally significant because it isolates carbon from the atmospheric pool for hundreds to thousands of years. The poorly-constrained microbial remineralization of labile DOC is *also* globally significant because it mediates the reverse process – rapid conversion of recently-fixed carbon back to inorganic form and its subsequent return to the atmospheric pool – while simultaneously providing energy for the processes that lead to export. According to most current-generation global carbon models, changes in the marine inorganic carbon reservoir driven by biological climate-carbon feedbacks (i.e., microbially-driven) will be balanced largely by changes in physical transport[72]. Reluctance to explicitly add principles governing labile molecule release from microbial cells can be traced to the high computational cost, limited conceptual foundation, and lack of evaluation data for formulating a more complex model. The extent to which carbon reservoirs are sensitive to alterations in the rates and efficiencies of processing recent photosynthate has therefore not been addressed on a global scale. We will implement rules of the chemical-microbe network (ST2) as constraints on global models.

The new, more advanced model treatments of chemical-microbe dynamics developed in ST3a and tested against regional field data in ST3b will be incorporated into PZND global ocean simulations to explore

carbon cycle ramifications (*Doney*). We will use the ocean biogeochemical component[73] of the Community Earth System Model (CESM) for these experiments because CESM is an NSF-supported activity with an open source code, publicly available reference simulations, sophisticated model analysis toolkits, and substantial model development and user communities[74]. Although more biologically sophisticated modeling approaches are used in some large-scale ocean models to incorporate explicit competition and selection among plankton species[75] and for trait-based approaches[76,77], the PZND-based CESM is used widely for international model comparisons of historical and future climate change and human carbon cycle perturbations[67,68]. As with regional models, results from global models will be used to make predictions on the distribution of pathways that can be examined using metagenomics.

ANTICIPATED ACCOMPLISHMENTS: Quantitative consideration of labile DOC flux using regional and global models will close one of the largest uncertainties in describing the global carbon cycle as it is today, and will substantially improve understanding of vulnerabilities to regional and decadal changes. C-CoMP will build a scientific framework for merging individual efforts of the experimental, field, and modeling communities whose members typically self-identify with different scales and levels of biological organization. Components emerging from community-scale models that improve predictions of regional- and global-scale models will be identified. Robust parameterizations of chemical-microbe network dynamics will improve predictions of global climate change Earth system model simulations, with results communicated to science and policy working groups and incorporated into data visualizations for expanding ocean literacy (see ET1, Knowledge Transfer, below).

ET&Ts will jumpstart ST3 through ocean data ingestion (e.g., supporting model development and testing) and molecular biology advances (e.g., investigating model predictions of links between microbial traits and carbon fate).

D. Research Timeline

C-CoMP's 5-year timeline emphasizes Science Themes (ST) in metabolite identification and annotation, model systems, and chemical-microbe network models in Years 1-3, stimulated by implementing enabling tools and technologies (ET&T) for data ingestion and analysis, and developing molecular tools for model microorganisms. In later years, global scale modeling and field are emphasized, enabled by tools that improve chemical detection and ocean sampling.

Research Objectives Timeline	Y1	Y2	Y3	Y4	Y5	Leads
Strategic chemical identification **(ST1a)**						Edison
Focused metabolite annotation **(ST1b)**						Moran
Chemical & biological data integration **(ST1c)**						Eren, Braakman
Model systems **(ST2a)**						Dyhrman
Field observations (BATS, MVCO) **(ST2b)**						Bates
Chemical-microbe network models **(ST2c)**						Covert
Modeling across scales of complexity **(ST3a)**						Segrè
Regional biogeochemical models **(ST3b)**						Doney
Global biogeochemical models **(ST3c)**						Doney
Data ingestion **(ET&T1)**						Saito
Software development **(ET&T2)**						Eren
Molecular biology tools **(ET&T3)**						Moran
Clio, chemical methods development **(ET&T4)**						Kujawinski, Saito

Table 1. Research timeline for C-CoMP.

III. EDUCATION AND HUMAN RESOURCE DEVELOPMENT OBJECTIVES

The ocean affects every aspect of life on the planet, regardless of geographical proximity to the coast. As we investigate one of Earth's largest carbon fluxes, we will engage students of all ages in understanding the functioning of the ocean ecosystem. C-CoMP will develop initiatives that expand ocean literacy in K-12 classrooms and the broader population (Education Theme 1; ET1), undergraduate curricula and research experiences that showcase the ocean sciences and provide multiple entry points to the ocean science research community (ET2), post-baccalaureate programs to help transition undergraduates into graduate education and careers in ocean science research (ET3), and interdisciplinary graduate student and postdoctoral programs that prepare the next generation of oceanographers (ET4; Fig. 7). Each of our initiatives leverages strengths within C-CoMP science and education senior personnel and furthers goals for recruitment, retention, and inclusion of diverse perspectives and backgrounds in the ocean sciences. Together, efforts will bring knowledge of ocean functions to people often not captured in ocean education programs, such as racial, gender, and socioeconomic underrepresented groups. All senior personnel will participate in education activities. They will choose the activity best suited to their scientific interest and accessible audiences. Here we describe the educational initiatives; in the next section (Broadening Participation) we highlight their potential to diversify the ocean science community.

(ET1) *Education Theme 1: Expanding Ocean Literacy*

ET1a. K-12 populations – Efforts for expanding ocean literacy will be grounded in a standards-based framework, following the National Research Council proposal to shift from memorization-based to inquiry-based approaches[78]. Based on this framework, delegations from US states developed the Next Generation Science Standards (NGSS) in 2013. Since then, 20 states and the District of Columbia adopted NGSS, and an additional 24 adopted science standards based on the NRC Framework; together, these states (including C-CoMP states IL, CA, MA, NY, GA) represent 71% of US children[79]. Similarly, marine scientists and educators ("The Ocean Literacy Network") developed the Ocean Literacy Principles, seven concepts of ocean science that articulate the foundation of ocean knowledge[80]. Ocean literacy initiatives will work at the nexus of these two efforts. C-CoMP science objectives align with Ocean Literacy Principles #5 (ocean life), #6 (ocean and society interconnections), and #7 (ocean exploration); we seek to incorporate ocean sciences into NGSS standards for life and physical sciences covered in C-CoMP research. NSF-supported Consortia for Ocean Science Exploration and Engagement (COSEE) identified numerous opportunities for expanding ocean themes within NGSS disciplines, but these efforts have

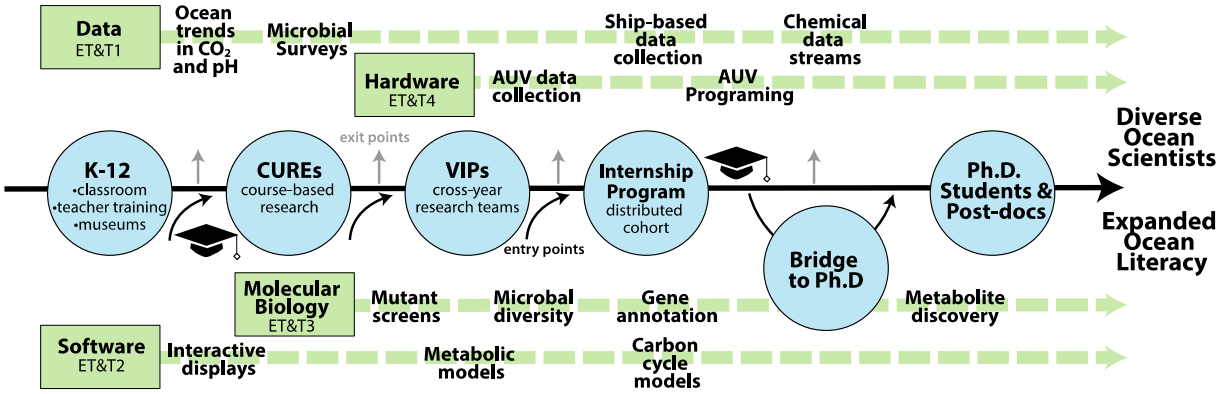

Fig. 7. Outreach and education activities (blue circles) offer multiple pathways to ocean literacy and engagement in ocean science research; these align with C-CoMP ***emerging tools and technologies*** at appropriate levels of complexity (green squares and dotted arrows with example activities). C-CoMP outreach and education initiatives focus interventions on known "exit points" by offering content and structures for retaining learners and providing entry points for new learners. The effectiveness of these initiatives will be studied through educational research (*Dolan, O'Dwyer*) and successful activities exported to the broader STEM education community.

been limited since COSEE funding ended in 2015. *C-CoMP will advance ocean education goals and develop best practices for expanding ocean literacy in US society.*

K-12 curricular materials will close the gaps between the Ocean Literacy Principles on ocean life (#5), ocean and society (#6), and ocean exploration (#7), and the "Disciplines" in NGSS. C-CoMP will leverage training and learning opportunities of PI-established, standards-based, outreach programs that engage middle and high school students and their teachers. Outreach activities at C-CoMP partner institutions include science kits for ocean acidification and plankton; camps and informal educational programs; teacher workshops; and on-line data visualizations. Our PhD-level Education Coordinator will develop additional K-12 curricular materials in collaboration with project senior personnel and assist with testing and implementation. Technology-focused programs will engage K-12 students and teachers to use ocean science data as a learning tool. The Education Coordinator will work with *O'Dwyer* to identify programs with greatest impact, and these will be expanded to national levels through engagement with developers such as OpenSciEd and submission to the NGSS Peer Review Panel (PRP) for science. Inclusion in NGSS-related on-line repositories will distribute C-CoMP ocean literacy materials to target audiences.

As an example, *Dyhrman*'s Artistic Oceanographer Program (AOP) is a standards-based program focusing on Ocean Literacy Principles #5 and #6 through integration of phytoplankton biology with art[81]. Exercises are based on NGSS in microbiology and ecology and involve autonomous inquiry supported with a resources kit provided to the classroom (microscope, slides, sample nets, art materials). Middle schools are the current focus of AOP, as this is a transition point in STEM education when many students lose interest in science and math[78]. With C-CoMP support, *Dyhrman* will expand the AOP in the NYC metropolitan area with school district partnerships and teacher training workshops, working with Earth2Class (E2C) to present hands-on workshops. E2C supports "Saturday Workshops for Educators" at Columbia, and since 1998 has provided more than 120 workshops for teachers from public and private schools, many of which serve student populations typically underrepresented in STEM.

ET1b. US society – Activities that touch US citizens in both science-focused and non-traditional venues will support the expansion of ocean literacy. C-CoMP will work with science museums, zoos, and aquaria in urban areas to develop exhibits that highlight essential ocean functions that affect all US citizens (Ocean Literacy Principle #6). As an example, *Sullivan*'s partnership with the Center of Science and Industry (COSI) in Columbus, OH[82] produced an exhibit on viruses and microbes in ocean carbon cycling. *Sullivan* will expand this exhibit through C-CoMP to advance ocean-oriented examples in chemistry and engineering to COSI's more than 250,000 annual visitors. Year 1 activities will augment existing PI partnerships to strengthen links with C-CoMP educational goals, and research will be undertaken to evaluate effectiveness in promoting interest in ocean research and understanding of the role of science in society (*O'Dwyer*).

(ET2) ***Education Theme 2: Undergraduate engagement in ocean science research***

With collaborative ocean-based research as the theme, C-CoMP undergraduate education initiatives will offer multiple pathways into research, from simply learning how research is done (ET2a), to engaging more deeply in research as a potential career path (ET2b), to preparing for science and technology careers (ET2c, ET3). These activities target points in STEM education where students may be excluded from science, especially those from traditionally marginalized backgrounds[83]. They are framed from an asset-based perspective which recognizes that higher education structures need to change in order for the ocean science community to cultivate all available talent[84].

ET2a. Undergraduate curricula – C-CoMP will develop introductory-level Course-based Undergraduate Research Experiences (CUREs), in which students collectively tackle a research question as a component of regular coursework (e.g., in introductory biology, chemistry, oceanography). CUREs are recognized for their potential to broaden access to research experiences and promote persistence and success regardless

of student backgrounds[85-87]. Under this model, C-CoMP faculty will incorporate a hands-on research element into their new or existing course curricula, with options for data-based (i.e., computational, modeling) or classroom-based (i.e., "wet lab") elements. With the course-based structure, CURE students earn credit toward their major while learning what research is, how it is done, and whether they would like to pursue it further[88]. Questions will be tractable for undergraduates, will benefit from the cohesion of many minds and hands, and will be relevant to C-CoMP scientific goals. For example, students will mine long-term records of microbial DNA from BATS for genes that indicate specific capabilities of marine microbes (ST1b, ST2b). Or students will isolate marine bacteria, measure growth rates, and parameterize simplified carbon cycle models (ST2a, ST3a). CURE courses will be offered both at partner institutions and in collaborations with faculty at two-year colleges and minority-serving institutions through CUREnet, a national network of people and programs established with NSF funds (*Dolan*). Faculty representing 78 colleges and universities have participated in CUREnet professional development and offered 221 CURE courses (Fig. 8). For teaching-intensive CUREnet faculty eager to engage their undergraduates in course-based research but lacking pedagogical expertise to do so, C-CoMP members will provide scientific guidance and infrastructure. Coastal facilities affiliated with our institutions (e.g., WHOI, UGA, UVA, BIOS) will host capstone research weekends for CURE students and faculty.

Fig. 8. Top left: Students participate in a research cruise on the R/V Atlantic Explorer. Top right: A CUREnet professional development workshop co-led by *Dolan*. Bottom: The annual NYC Marine Science Day attracts ~7000 visitors to *Dyhrman*'s Artistic Oceanographer Program booth.

ET2b. Undergraduate VIP research teams – At a deeper level of analysis, undergraduates will be engaged in interdisciplinary research teams following the model of Vertically Integrated Projects (VIP)[89]. VIP involves teams of ~12 students in research that is cross-year (equally divided between 2nd, 3rd, and 4th year students) and cross-disciplinary (spanning multiple majors). C-CoMP faculty will offer VIP courses at their institutions that are germane to our science objectives, at the level of 1-2 credits per semester that count toward degree completion. For example, a systems biology analysis of bacterial host/virus systems will involve the team in designing and conducting lab experiments and integrating data using genome-scale metabolic modeling (ST2a, 2c). Initially, faculty and early career researchers in C-CoMP research groups (graduate students, postdocs, experienced undergraduates) will help students develop domain knowledge and techniques. As the team becomes established, experienced VIP students assume some mentorship responsibility. Students will explore research as a career path, acquire technical expertise, develop professional skills (mentoring, communication), and collaborate in research. This near-peer mentoring structure fosters a sense of community and belonging in STEM, especially among students from traditionally marginalized backgrounds[90]. VIP teams have the added benefit of providing an entry point for transfer students if they missed the opportunity to experience research through CUREs.

ET2c. Undergraduate summer research – A deep engagement of undergraduates in ocean science research will occur through application-based 10-week C-CoMP summer internships. This geographically distributed cohort of undergraduates will work on independent projects in C-CoMP laboratories. At the beginning of the summer, the cohort will participate in a week-long online course in data-enabled oceanographic research following the successful models of the Rosetta Commons NSF Research

Experience for Undergraduates (REU) site[91] and the University of Chicago course in data literacy[92] (*Eren*). The course will cover data analysis in the ocean sciences, with learning assessed through assignments using online BATS and MVCO data. At the end of the summer, the cohort will participate in a week-long field- and ship-based capstone experience at BIOS collecting data for C-CoMP science objectives (*Bates*). These flanking activities will develop a sense of community and scientific identity, two predictors that students will continue in a STEM research career[89,93]. Priority will be placed on selecting interns from backgrounds that have traditionally been excluded from the sciences (e.g., racial and ethnic minorities, first-generation college students, students with physical disabilities) and from colleges and universities with limited research infrastructure (e.g., two-year colleges). Local activities and programming for undergraduate researchers will enrich the experiences of the cohort. All partner institutions host NSF-funded REU programs that offer social and professional development activities and training tailored to students' interests (e.g., modeling workshops, coding camp).

ET2d. Professional development for education and outreach – In Year 1, faculty, graduate students, and postdoctoral researchers will participate in 8 h of professional development facilitated by *Dolan* based on the established *Entering Mentoring* curriculum[94]. Continued mentorship education will be embedded in the C-CoMP culture, achieved through professional development activities during annual project meetings and online training[95]. Furthermore, senior personnel involved in CUREs and VIP teams will participate alongside teaching-intensive faculty collaborators in an additional 20 h of professional development on CURE design and instruction, based on the successful CUREnet model[87].

(ET3) ***Education Theme 3: Post-baccalaureate transitions to graduate school***

Transition into ocean science graduate programs will be supported for post-baccalaureate students through technical research training and professional development at partner institutions in a Bridge-to-PhD (B2P) program. Each year, 4-5 new B2P fellows will be placed in C-CoMP research groups. The 2-year fellowship program provides sufficient time to establish core competencies in their field and benefit from tailored mentoring and professional development. Fellows will be chosen through an application process that prioritizes applicants from backgrounds traditionally excluded from the sciences. They will be placed in science disciplines ranging from marine chemistry to microbial ecology to data science by matching student interest with C-CoMP faculty, and their research experiences will be aligned with science objectives of the Center. Through virtual platforms, B2P Fellows will engage in peer mentoring and professional development, including: (1) fellowship and proposal writing, (2) ethics in research, and (3) science communication. Fellows will engage in C-CoMP outreach/education activities, participate in the C-CoMP annual meeting, travel to national meetings, and join undergraduate summer interns at the BIOS capstone week for leadership and team-building.

Our Bridge-to-PhD program is modeled after the successful Columbia University program[96], in which *Dyhrman* has participated and for which she serves on the advisory board. It also builds on the minority-focused Woods Hole Partnership Education Program through which *Doney* and *Saito* have mentored research technicians who successfully continued to STEM graduate programs. To attract a diverse pool of applicants, C-CoMP will publicize widely, leverage personal connections from CUREs and collaborators at minority-serving institutions, correspond with graduate school advisors at these institutions, and use digital outreach (social media, blog posts by existing fellows). The most effective tools will be personal contacts, through which qualified applicants can be encouraged to apply. For example, C-CoMP Diversity Coordinator *Dolan* has been working with minority-serving institutions through CUREnet (ET2a, above) and will engage with their faculty to identify, attract, and support students through the application process. C-CoMP collaborator Breier is based at U-Texas Rio Grande Valley, one of the largest Hispanic-serving institutions in the US, where students with interest in ocean sciences research can be made aware of the B2P program.

(ET4) ***Education Theme 4: Graduate students and postdoctoral researchers***

ET4a. Graduate students – Graduate students supported through C-CoMP faculty research allotments are integral to Center research and education missions. A C-CoMP graduate student organization will support interaction and provide a forum for graduate student-focused programming (on-line seminar series, grant writing courses, coding tutorials). Graduate students will be invited to participate in Center research meetings, online science forums, and field programs for testing data integration and visualization platforms. Graduate students will participate in C-CoMP professional development opportunities and engage in education and outreach activities, including mentoring and field-based training of undergraduate summer interns and B2P fellows at BATS and MVCO, and developing K-12 ocean-based curricula.

ET4b. Postdoctoral researchers – C-CoMP funding will support 2 new interdisciplinary postdoctoral researchers per year for 2-year positions, and an additional 2-3 post-doctoral researchers will be supported annually by C-CoMP faculty through their research allotment. The Center-selected positions will enhance the interdisciplinary research foci of C-CoMP, bridging between two or more host laboratories. Interdisciplinary topics in Years 1-2 will be 'transition from cell models to chemical-microbe network models' (ST2c; working with modelers *Braakman*, *Covert*, and *Segrè*) and 'building data integration and visualization platforms' (ET&T2; working with *Saito* and *Eren*). In subsequent years, the Executive Committee will assess emerging needs for enhancing interdisciplinarity in recruiting Center postdoctoral researchers. All postdoctoral associates will be entrained into mentoring and career development training opportunities (see Postdoctoral Mentoring Plan). They will present their research at C-CoMP annual meetings and at national and international meetings.

Enabling effective education practices in enhancing workforce diversity and ocean literacy

The C-CoMP team is rounded out by education researchers (*Dolan, O'Dwyer*) who will conduct empirical research on the impacts of C-CoMP education and outreach activities. Despite likely differences in the types, duration, and intensity of programs between the target audiences (US society, K-12, undergraduate and post-baccalaureate communities), analogous approaches can be adopted to create a cohesive body of research on ocean literacy and career-related interests and motivation. Research methodologies will examine the impacts of both formal curriculum interventions and research education experiences. Given the aim to broaden participation among racial, socioeconomic, sexual, and gender groups, all educational research studies will examine the experiences and outcomes of sub-populations of individuals with attention to intersectionality (i.e., experiences of individuals with multiple marginalized identities). Educational research activities will be framed in terms of relevant learning, motivation, and career development theory, as well as the logical connections between the stated goals of each activity, the fidelity of its implementation, and its hypothesized outcomes.

The development and implementation of C-CoMP education activities will evolve over the course of the funding period, and therefore goals for educational research will vary accordingly. During the development phase, we will follow the core principles of "improvement science" advocated by the Carnegie Foundation for the Advancement in Teaching. This approach guides ongoing development and modification of activities by focusing on what works, for whom, and under what conditions, leading to activities that are scalable to other locations and populations[97]. When educational activities are optimized later in the funding period, more conventional research approaches will be used to examine the impact of activities on learner outcomes and their mechanisms of action.

For K-12 communities, consider the Artistic Oceanographer Program (AOP)[81] (ET1a), aimed at improving students' ability to conduct autonomous inquiry. Educational research will be conducted with teachers and their students in the NYC metropolitan area to examine changes in content knowledge and beliefs

relating to oceanographic concepts as a result of engaging with the AOP. For example, we will conduct a randomized experiment whereby teachers who receive training and implement AOP activities with their students are compared to teachers and their students in business-as-usual classrooms. In addition to administering content knowledge assessments that are aligned with NGSS, we will conduct surveys and scales to measure teacher and student attitudes and beliefs about the relevance of ocean science.

Likewise, an improvement science approach followed by research on targeted outcomes will be conducted for the undergraduate education activities. All undergraduate activities will be evaluated both formatively and summatively to make year-to-year improvements and examine effectiveness in terms of developing student interest, motivation, and integration into the scientific community[93,98,99]. Evaluation will address success in: (1) engaging all C-CoMP senior personnel in outreach and/or undergraduate education and training; (2) providing experiences that promote learning and development; (3) providing inclusive access to research (that is, offering multiple entry points into research, setting mentor-to-student ratios tuned to the difficulty of the research challenge); and, (4) integrating data science competencies at all program levels. Undergraduate activities will also be studied to address novel questions in undergraduate science education, such as the amount of time students must engage in research and the types of research engagement necessary for students to begin to identify as scientists, to develop a sense of belonging with the scientific community, to make decisions regarding whether to pursue a science research career, and to be successful in their graduate education and career pursuits[100]. Students' experiences will be compared to examine differential impacts on career interests and motivation as well as their integration into the scientific community[88,93], including differential experiences or impacts for students from traditionally marginalized backgrounds.

For more informal science education experiences, novel methods for examining impacts are needed. NSF's seminal framework for evaluating the impacts of informal science education projects[98] and the work conducted under the auspices of The Center for Advancement of Informal Science Education (CAISE)[101,102] highlight opportunities and challenges in discerning the impacts of informal science interventions. Leveraging work-to-date in the field, C-CoMP will implement rigorous designs to examine the impact of exhibits, youth and community programs, learning technologies, and collaborations among various institution types. These will include short duration experimental designs, survey research designs, and designs that incorporate rich qualitative narratives. We will capitalize on recent work[103] that synthesizes available and shared measures for examining common outcomes of information science education experiences. Research activities conducted during C-CoMP will advance both formal and informal science education fields.

Collectively, the education and outreach elements of C-CoMP are grounded in research and theory on student learning development and public literacy, led by experts in education and outreach programming and research, and designed to contribute to the knowledge base in STEM education and outreach. Furthermore, these elements will build capacity among Center personnel and enhance infrastructure that broadens participation in ocean science and diversifies our workforce.

IV. BROADENING PARTICIPATION OBJECTIVES

Broadening participation in the ocean sciences to achieve racial, sexual, and gender diversity is central to C-CoMP research and education goals. Despite concerted effort over four decades, underrepresented minority (URM; typically Black/African American, Latinx/Hispanic and Indigenous/Native American individuals) representation in the ocean sciences is woefully inadequate. People of color account for 3.8% of tenured or tenure-track faculty in the top 100 geoscience departments, and only 4% of doctorates from graduate training programs in oceanography[104] in 2016. A substantially smaller share of URM bachelor's degree recipients in the US enter graduate school compared to white- and Asian-identifying counterparts[105]. The

higher proportion of first-generation URM students may be a contributing factor; in 2011–2012, 42% of Black students and 48% of Latinx students were first-generation, compared with 28% of white students[106]. New Americans and first in family to high school or college encounter barriers from family expectations for familiar career paths with established financial security (e.g., physician, lawyer), and misconceptions about the cost (or lack thereof) of STEM Ph.D. programs and resulting job prospects[93]. Well-documented disparities in writing reference letters for women and URM[107] also contribute to retention failures.

A culture of diversity and inclusion in C-CoMP will embrace best practices for hiring, planning, and implementing meetings and field experiences. C-CoMP participants at all levels will complete training in Responsible Conduct of Research and be informed of avenues for reporting concerns within the Center and at partner institutions (see C-CoMP Ethics Plan). Participants involved in hiring will be aware of potential bias in reference letters and best practices guidelines for the evaluation of applications[107-109]. Annual meetings and workshops will be planned in accordance with recommendations for inclusive scientific meetings[110].

Two top priorities recognized by the National Academy of Sciences[111,112] for broadening participation in the STEM workforce are: (1) improving undergraduate retention and completion; and (2) improving the transition to graduate school. The first priority is addressed with C-CoMP undergraduate research programs that integrate academic, social, and professional development (ET2) and provide opportunities to explore research as a career regardless of student social identities[85,87,113]. The second priority is addressed with our Bridge-to-PhD program (ET3) that supports students at a transition typified by large losses in diversity in STEM training[96]. Research-focused employment in a C-CoMP laboratory will provide B2P Fellows with near-peer mentoring, role models, technical experience, and an established scientist mentor for graduate student applications and beyond. Further, the cohort nurtured in B2P programs improves long-term URM retention and advancement in STEM careers[96]. Authentic research experiences for B2P participants that address problems of societal relevance (e.g., Earth's carbon cycle and climate system) also increase URM engagement and retention[93,114].

Recruitment to C-CoMP opportunities will occur through existing education networks (see ET4) as well as outreach to local chapters of national advancement organizations such as the Society for Advancing Chicanos/Hispanics and Native Americans in Science (SACNAS). In recognition of their value, the American Geophysical Union (AGU) has recently started a Bridge-to-PhD Program that solicits applications from post- baccalaureate URM students and places them in positions with partners at host institutions; C-CoMP senior personnel will participate in the AGU program. For both the undergraduate summer intern and B2P programs, C-CoMP education research (*Dolan, O'Dwyer*) will measure success in recruitment, experiences, and retention of URMs in STEM, both to communicate outcomes to the education research community and to improve our own programs.

KNOWLEDGE TRANSFER OBJECTIVES

C-CoMP science and education efforts will improve understanding of the drivers and vulnerabilities of the ocean carbon cycle and expand ocean literacy. Along this journey, we will generate new knowledge, methodologies, and technologies to advance science and education. Science accomplishments will be transferred to experts with complementary knowledge through targeted workshops, graduate student and postdoctoral forums, and social media interactions (e.g., Twitter (@microbialplanet)) and newsletters. Stakeholders in other microbiome fields, "open science" proponents, teachers, and policy-makers will be engaged through Center communications and targeted outreach. Developing cross-disciplinary platforms and hosting round-tables at national meetings will further our own science through ideas exchange while engaging a larger community committed to improved understanding of chemical-microbe networks. We

will learn from knowledge transfer strategies of previous STCs, and evaluate our success in both exporting C-CoMP products and importing advanced academic and industry products (Fig. 9).

Our science will expand the lexicon of molecules known to mediate microbial functions in ecosystems. Newly identified molecules, proteins, and genes will populate databases, providing better opportunities for addressing microbial interactions mediated through chemical transfers (Fig. 9). These functional annotations will be relevant to microbiome studies in freshwater, soils, mammalian hosts, agriculture – essentially all environments and hosts that harbor microbial consortia. Advances in metabolic modeling will bring contextual information to metabolite ecology and establish a foundation of microbial function and metabolism relevant to other systems. Technological innovations for sampling and identifying metabolites in seawater are relevant for other fields where metabolomics is challenging and key proteins and genes elude discovery. We will both contribute to and learn from expertise in science and industry. The modifications to Clio will make it the first AUV sampler available to the scientific community with capacity to link metabolite inventories with gene, protein, and transcript data in time and space.

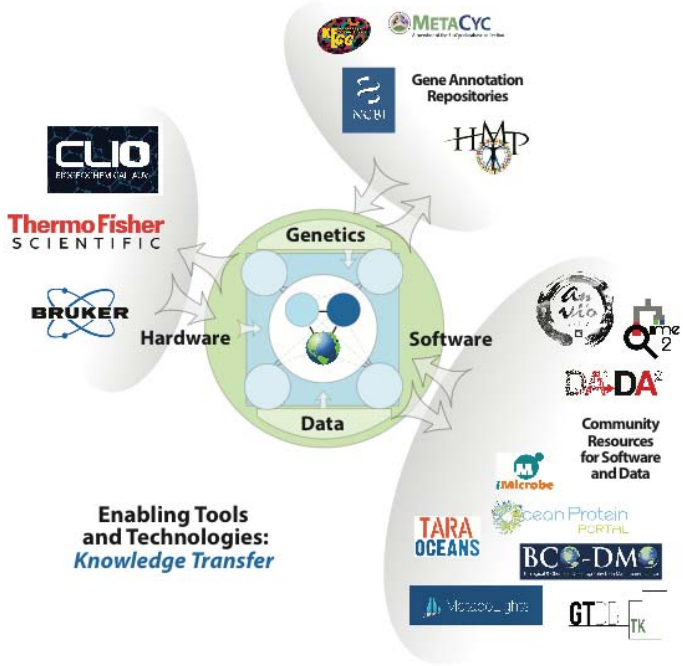

Fig. 9. C-CoMP will exchange knowledge with scientific entities and industries to develop new chemistry capabilities, advance computation, populate databases, and improve data analysis.

An open science environment is a central element of C-CoMP. Rapid data release, continuous community engagement, and fully reproducible data science will be one of our most effective avenues for transferring knowledge generated through NSF's considerable investment. C-CoMP will enable effective data integration by the development of novel tools to merge heterogeneous types of new and existing chemical and molecular data with environmental parameters, leveraging our expertise in computer science, informatics, analytical chemistry, chemical oceanography, microbiology, and ecology. The Center will provide the structure necessary for availing these tools to a wide-ranging community of scientists and educators that will benefit from shared data, software platforms, genetic tools, and chemical methods (Fig. 9).

To promote maximum data access (Fig. 10), C-CoMP laboratories will store raw data in repositories with digital object identifier (DOI) assignments. Metadata will be stored and version-tracked with Google sheets, for which application programmer interfaces are available. As intermediate data products emerge, they will be shared within and external to C-CoMP via the Google Drive platform. As data products harden, they will be deposited on GitHub and Zenodo, which accept version-controlled intermediate data products for DOI assignments. A project page at BCO-DMO (see below) will collate field data and links to all repositories. C-CoMP will maintain a blog on its web page, promoted by social media, to share data analysis news and intermediate findings. Our goal is to reach researchers and students promptly, benefit from community exchange, and create networking opportunities for young scientists. We will share fully reproducible bioinformatics analyses using Docker containers and Jupyter notebooks for

platform independence and ease-of-use. We will build cohesion through in-person meetings, workshops and hackathons, and digital communications (Fig. 10).

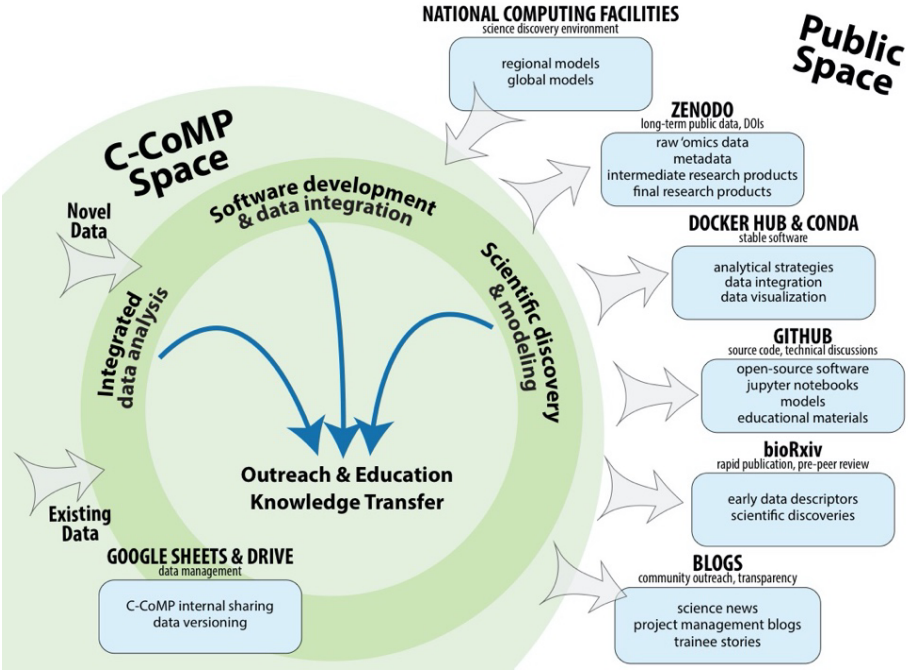

Fig. 10. Strategies for C-CoMP data integration, access, and serving.

C-CoMP will partner with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) in an informatics research collaboration, based at WHOI, to enable synergy in developing data integration platforms for the broader science community. BCO-DMO is a data management office that facilitates deposition and serving of oceanography datasets for NSF grantees. With nearly 10,000 curated datasets ingested to date (complete with metadata, methodologies, and links to raw data repositories), BCO-DMO leverages emerging informatics research to maximize data discovery and re-use. The C-CoMP project page on BCO-DMO will hold field data and links to raw data, intermediate data products, and bioinformatics packages in other repositories. BCO-DMO uses standard vocabularies for oceanographic parameters, leveraging curation efforts by the British Oceanographic Data Centre. For environmental parameters not yet in the curated vocabulary (e.g., metabolites and peptide biomarkers), the Digital Coordinator will work with BCO-DMO staff to incorporate new parameters. C-CoMP research on integration of rich multi-omics datasets in collaboration with BCO-DMO will benefit both C-CoMP and the continuing development of BCO-DMO infrastructure and practices. C-CoMP will invest early-on in the design and implementation of data integration platforms, supporting a programmer and a postdoctoral researcher to work with *Eren*, *Saito* and the Digital Coordinator in Years 1 and 2 to gather relevant legacy data and establish a working prototype for data integration and visualization. Efforts will build on experience from Anvi'o, BCO-DMO, and the Ocean Protein Portal[19,26].

Findings from our science efforts will be echoed in the Center's education efforts, broadening the appreciation for marine microbes in K-12 and undergraduate curricula. We will ground our K-12 curricular efforts in the NRC Framework for science education to ensure broad applicability to the large proportion of US school children taught with these standards. Integrating curricular development with rigorous education research ensures that we export effective curricula to appropriate venues such as NGSS-affiliated teacher resources. At the undergraduate level, we have strong connections with CUREnet, ensuring that ocean-centered undergraduate course materials will be available to faculty across the US. Through rigorous education research, our CUREs will be continuously evaluated and modified for success in broadening participation in STEM. Best practices will be exported to the education community through publications, presentations, and the CUREnet and VIP communities. Data integration platforms developed within C-CoMP will be incorporated into all education and training initiatives (see Fig. 7).

On the broadest societal level, C-CoMP seeks understanding of the factors that control cycling of 25% of global carbon fixation and their sensitivity to a changing climate. Improvements to global carbon models through inclusion of microbially-mediated climate-carbon feedbacks in the surface ocean will be transferred to scientists and policy-makers involved in climate change mitigation[115]. We will interface with these communities through the Climate Model Intercomparison Project (CMIP)[116], the Community Earth System Model (CESM) workshop, the CESM Advisory Board (CAB) that connects all the major U.S. climate centers, and the science-policy bridging organization COMPASS. In the policy arena, *Doney* has previously briefed administration and congressional staff and given congressional testimony, and these avenues will be pursued to communicate C-CoMP advances.

MANAGEMENT PLAN

Leadership. The C-CoMP administration and management plan (Fig. 11) establishes points of interaction across scientists, educators, and administrators. **Elizabeth Kujawinski** and **Mary Ann Moran** will lead C-CoMP with assistance from Managing Director **Sofia Ibarrarán-Viniegra**. The leadership team will work closely with the External Advisory Board, Executive Committee, external Evaluator, and Ethics Panel. Kujawinski will assume overall leadership of C-CoMP, guiding science, education, broadening participation, and knowledge transfer initiatives. She will represent C-CoMP within WHOI, at NSF, and at public events. As co-Director and Research Coordinator, Moran will oversee research programs and work with the Managing Director to coordinate science working groups. She will assume duties of the Director should Kujawinski be unavailable for any reason. Ibarrarán-Vineigra will have primary responsibility for Center administration, overseeing fiscal matters, annual report preparation, and annual meeting and workshop logistics. Additional administrative support is provided through WHOI's Marine Chemistry & Geochemistry Department for overhead-supported functions.

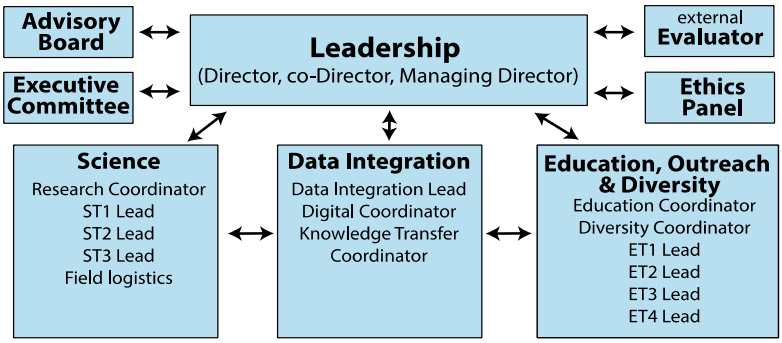

Figure 11. Organizational chart for C-CoMP leadership and management. Senior personnel leading science, data integration, and education/ diversity efforts (Table 1, below) will use video- and in-person meetings to connect across activities. ST = Science Theme; ET = Education Theme.

Kujawinski is a Senior Scientist at WHOI with experience leading interdisciplinary lab- and field-based research on molecule-microbe interactions and chemical methods development. She leads the WHOI FT-MS facility and served on advisory boards for UNOLS and the National High Magnetic Field Laboratory.

Moran is a Regents' Professor at UGA with experience leading research teams in interdisciplinary marine microbial ecology. Moran holds elected leadership roles in the American Academy of Microbiology and AAAS, was a GBMF Investigator in Marine Microbiology, and is a Simons Investigator in Life Sciences.

Ibarrarán-Viniegra is a biology Ph.D. with experience in project management. She has served as a project coordinator for the WHOI-based Ocean Twilight Zone and will join C-CoMP full-time, if funded. She is experienced in coordinating multi-institutional research teams and conversant in WHOI administrative and financial policies.

External Advisory Board (EAB). An External Advisory Board (EAB) of leading scientists in chemistry, oceanography, microbial sciences, data science, and STEM education will provide guidance (at annual meetings and by video) twice each year. EAB membership will be for staggered 3-year terms to balance new perspectives with programmatic continuity. We will invite 6 individuals who have demonstrated

scientific vision and a track record in leading interdisciplinary programs, balanced across career stages and including those with experience building diverse and inclusive research teams.

Executive Committee. The Executive Committee will oversee Center science and education activities. Committee members include the Director, co-Director/ Research Coordinator, Science Lead, Education Lead, Data Integration Lead, Diversity Coordinator, and Knowledge Transfer Coordinator. The Executive Committee will meet with the Managing Director and Education Coordinator via web-conference bi-weekly to guide research and education activities and address challenges as they arise. Coordination with the Ethics Panel will occur every six months or as needed.

Ethics Panel. The Ethics Panel will address policies and arising concerns related to ethics and responsible conduct of research (see Ethics Policy). The Panel will be balanced by career stage, gender, and research expertise, rotating among C-CoMP personnel (initially: **Sonya Dyhrman (chair),** *Braakman*, graduate student, postdoc). Issues will be resolved a timely manner in consultation with the Executive Committee.

External evaluator. **Karen DeMeester** (UGA) will provide external evaluation of C-CoMP research goals. She will develop an evaluation schedule and instruments and collect and analyze data for semi-annual reports to the executive committee. Evaluation instruments will monitor Center progress, document effectiveness and impact, and prepare reports to NSF. Evaluations will be conducted in accordance with the Joint Committee on Standards for Educational Evaluation. Evaluations will be embedded in Center activities whenever possible to reduce participant burden.

Science coordination. Working groups will be established along science themes to facilitate synergies in research and technology. Each working group (along with Research Coordinator Moran) will meet by video every six months and by the digital communication platform Slack (prototyped successfully during proposal preparation) frequently. Working group membership (Table 2) explicitly includes non-domain scientists in each group for novel perspectives and cross-communication among themes. Data will be shared and versioned through Google Drive and other digital platforms (see Knowledge Transfer). Emerging topics will be explored with 2-3 targeted workshops and/or hackathons per year.

Table 2. Roles for Senior Personnel. Working groups are linked to science and education themes; leads are in bold.

Name, Institution	**Expertise**	**Role in C-CoMP; working group membership**
Elizabeth Kujawinski, WHOI	Marine metabolomics	Director; ST1a, ST1b, ST2b, ST3b, **ET3**
Mary Ann Moran, UGA	Marine microbiology	Co-dir/Res. Coord; **ST1b**, ST1c, ST2a, ST3c, ET2
Scott Doney, UVA	Ocean ecosystem modeling	Exec Comm; ST1c, ST2c, **ST3b,c**; ET4
Sonya Dyhrman, Columbia	Marine microbiology	Exec Comm; Ethics panel chair; ST1b, **ST2a,** ST3b, ET1
A. Murat Eren, U-Chicago	Data integration	Exec Comm; Data Integration; **ST1c**, ST2c, ST3a; ET2
Nicholas Bates, BIOS	Nutrient biogeochemistry	ST1c, **ST2b**, ST3b, ET2
Rogier Braakman, MIT	Metabolic evolution	**ST1c,** ST2c, ST3c, ET3
Markus Covert, Stanford	Metabolic modeling	ST1a, **ST2c,** ST3a, ET4
Erin Dolan, UGA	STEM education	Diversity Coordinator; ST1b, **ET2**
Arthur Edison, UGA	Metabolomics	**ST1a**, ST2b, ST3a, ET2
Laura O'Dwyer, BC	K-12 education methods	ST1c, **ET1**, ET4
Mak Saito, WHOI	Marine proteomics	Knowledge Transfer Coordinator; ST1c, ST2b, ST3c, ET3
Daniel Segrè, BU	Microbial network models	ST1b, ST2c, **ST3a,** ET2
Matthew Sullivan, OSU	Marine viruses	ST1c, ST2a, ST3a, **ET4**

Data integration. Integration of data streams through science and education will be led by **A. Murat Eren**, who has extensive experience in open-source data merging and visualization. The Digital Coordinator (housed with BCO-DMO staff at WHOI) will manage C-CoMP data streams (except confidential education data managed by education researchers) and support compatibility and visualization efforts.

Mak Saito will serve as the Knowledge Transfer Coordinator, leveraging his experience as Science PI of BCO-DMO and expertise in data analysis platforms in the ocean sciences.

Education, outreach and diversity coordination. We will recruit a Ph.D.-level Education Coordinator with responsibility for logistics associated with education and outreach initiatives, including coordinating development of K-12 curriculum materials, supporting diversity initiatives with the Diversity Coordinator, and overseeing the Bridge-to-PhD Fellows program. **Erin Dolan** will serve as Diversity Coordinator, tapping into extensive experience with STEM initiatives for broadening participation. Education working groups will support coordination and cross-fertilization (Table 2).

Center Cohesion. The C-CoMP philosophy is to engage an integrated team of researchers, including those whose fields do not naturally intersect with carbon cycle research but bring novel perspectives to this fundamental challenge. We will focus on the seams where divergent scientific disciplines overlap; we will ensure equal voices in the topics considered by our team; we will promote interactions that are fully respectful and inclusive of all perspectives. Open exchange of research ideas and methodological and conceptual challenges will integrate the science and education. Ideas and accomplishments in science, education, and broadening participation will be communicated to the C-CoMP community through a video conferencing series that engages graduate student and post-docs; science working groups that initiate discussion threads through Slack channels; and data-mining workshops that collectively generate hypotheses and design experiments. Face-to-face all-project meetings, workshops, video conferencing, and real time web-based communication (e.g., using Slack) will engage and integrate the C-CoMP team.

Organizations. C-CoMP will be based at the Woods Hole Oceanographic Institution (WHOI), a leading center for research and education in ocean sciences and engineering. Collaboration is the norm at WHOI, across scientific disciplines and between scientists and engineers, creating an environment fostering cooperation for advanced technological innovation and scientific discovery. C-CoMP will benefit from WHOI's leadership and participation in high-impact projects with complementary research agendas, such as innovations from The Ocean Twilight Zone, a privately-funded study of the ecology of ocean midwaters, the NASA/NSF-supported EXPORTS program, and the NSF-supported Ocean Observatories Initiative (OOI). WHOI is the current home of OOI, the NSF-funded Biological and Chemical Oceanography Data Management Office (BCO-DMO), the Ocean Carbon and Biogeochemistry project office (OCB), and the newly-funded New England Long Term Ecological Research site (NE-LTER), which includes MVCO, one of our proposed field sites.

Each partner institution brings additional capabilities to C-CoMP (see Facilities statements). Facilities in advanced chemistry and biology are based at the University of Georgia, WHOI, Columbia University, MIT and The Ohio State University. Computational expertise for modeling and data mining are based at the University of Virginia, the University of Chicago, Boston University, WHOI, and Stanford University. WHOI and BIOS will facilitate field research at MVCO and BATS, and have experience leveraging time-series data for educational and outreach activities. The University of Georgia and Boston College will lead the educational research and coordinate programming with partner institutions.

Budget management. The 14 senior personnel have annual research budgets of approximately \$150K y^{-1} to support primary research activities, augmented with salary and education funds for administrative roles, summer undergraduate interns, B2P Fellows, and Center postdoctoral fellows in their laboratories. Initial supported efforts for key tools and technologies include: (1) building data merging platforms; (2) constructing mutant libraries for model marine bacteria; (3) developing and deploying a metabolomics-focused module for AUV Clio. Priority elements in future budgets will be set by the Executive Committee, as will adjusting funding priorities as needed to insure optimal use of NSF resources.

References

1 Braakman, R., Follows, M. J. & Chisholm, S. W. Metabolic evolution and the self-organization of ecosystems. *P Natl Acad Sci USA* **114**, E3091-E3100 (2017).

2 Howard-Varona, C. *et al.* Multiple mechanisms drive phage infection efficiency in nearly identical hosts. *ISME J* **12**, 1605 (2018).

3 Howard-Varona, C. *et al.* Regulation of infection efficiency in a globally abundant marine Bacteriodetes virus. *ISME J* **11**, 284 (2017).

4 Johnson, W. M., Soule, M. C. K. & Kujawinski, E. B. Evidence for quorum sensing and differential metabolite production by a marine bacterium in response to DMSP. *ISME J* **10**, 2304 (2016).

5 Kujawinski, E. B. *et al.* Phosphorus availability regulates intracellular nucleotides in marine eukaryotic phytoplankton. *Limnol Oceanogr Lett* **2**, 119-129 (2017).

6 Landa, M., Burns, A. S., Roth, S. J. & Moran, M. A. Bacterial transcriptome remodeling during sequential co-culture with a marine dinoflagellate and diatom. *ISME J* **11**, 2677-2690 (2017).

7 Moran, M. A. *et al.* Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* **432**, 910 (2004).

8 Read, B. A. *et al.* Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* **499**, 209 (2013).

9 Pachauri, R. K. *et al. Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. (IPCC, 2014).

10 Siegel, D. A. *et al.* Global assessment of ocean carbon export by combining satellite observations and food-web models. *Global Biogeochem Cycles* **28**, 181-196 (2014).

11 Honjo, S. *et al.* Understanding the role of the biological pump in the global carbon cycle: an imperative for ocean science. *Oceanography* **27**, 10-16 (2014).

12 Hansell, D. A. Recalcitrant dissolved organic carbon fractions. *Annu Rev Mar Sci* **5**, 421-445, doi:10.1146/annurev-marine-120710-100757 (2013).

13 Cavicchioli, R. *et al.* Scientists' warning to humanity: microorganisms and climate change. *Nature Rev Microbiol*, 1 (2019).

14 Hertkorn, N. *et al.* Characterization of a major refractory component of marine dissolved organic matter. *Geochim Cosmochim Acta* **70**, 2990-3010 (2006).

15 Azam, F. *et al.* The ecological role of water-column microbes. *Mar Ecol Prog Ser* **10**, 257-263 (1983).

16 Pomeroy, L. R. The ocean's food web, a changing paradigm. *Bioscience* **24**, 499-504 (1974).

17 Durham, B. P. *et al.* Cryptic carbon and sulfur cycling between surface ocean plankton. *P Natl Acad Sci USA* **112**, 453-457, doi:10.1073/pnas.1413137112 (2015).

18 Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**, 465 (2016).

19 Saito, M. A., Breier, C., Jakuba, M., McIlvin, M. & Moran, D. in *The Chemistry of Microbiomes: Proceedings of a Seminar Series.* 29-36 (The National Academies Press).

20 CIFAR-Workshop. (Canadian Institute for Advanced Research, Toronto, CA, 2018).

21 MMI-Workshop. (Gordon and Betty Moore Foundation, New York, NY, USA, 2015).

22 Holderman, N. *et al.* in *Experimental Nuclear Magnetic Resonance Conference* (Pacific Grove, CA, 2019).

23 Widner, B., Kido Soule, M., Ferrer-González, F., Moran, M. A. & Kujawinski, E. B. in *ASLO 2019 Aquatic Sciences Meeting* (ASLO, San Juan, Puerto Rico, 2019).

24 Sogin, E. M., Puskas, E., Dubilier, N. & Liebeke, M. Marine metabolomics: measurement of metabolites in seawater by gas chromatography mass spectrometry. *bioRxiv*, 528307 (2019).

25 Seymour, J. R., Amin, S. A., Raina, J.-B. & Stocker, R. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nat Microbiol* **2**, 17065 (2017).

26 Eren, A. M. *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
27 Harcombe, W. R. *et al.* Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep* **7**, 1104-1115 (2014).
28 Weyman, P. D. *et al.* Inactivation of *P haeodactylum tricornutum* urease gene using transcription activator-like effector nuclease-based targeted mutagenesis. *Plant Biotechnol J* **13**, 460-470 (2015).
29 Peters, J. M. *et al.* Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. *Nature Microbiol* **4**, 244 (2019).
30 Price, M. N. *et al.* Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* **557**, 503 (2018).
31 Kehe, J. *et al.* Massively parallel screening of synthetic microbial communities. *P Natl Acad Sci USA*, 201900102 (2019).
32 Wetmore, K. M. *et al.* Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *Mbio* **6**, e00306-00315 (2015).
33 Mutalik, V. K. *et al.* Dual-barcoded shotgun expression library sequencing for high-throughput characterization of functional traits in bacteria. *Nat Commun* **10**, 308 (2019).
34 Ramaswamy, V. *et al.* Development of a ^{13}C-optimized 1.5-mm high temperature superconducting NMR probe. *J Mag Reson* **235**, 58-65 (2013).
35 Clendinen, C. S. *et al.* An overview of methods using ^{13}C for improved compound identification in metabolomics and natural products. *Front Plant Sci* **6**, 611 (2015).
36 Judge, M. T. *et al.* Continuous in vivo metabolism by NMR. *Front Mol Biosci* **6** (2019).
37 Flynn, K. J., Clark, D. R. & Xue, Y. Modeling the release of dissolved organic matter by phytoplankton 1. *J Phycol* **44**, 1171-1187 (2008).
38 Ankrah, N. Y. D. *et al.* Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J* **8**, 1089 (2014).
39 Carlson, C. A. & Hansell, D. A. in *Biogeochemistry of Marine Dissolved Organic Matter* 65-126 (Elsevier, 2015).
40 Gifford, S. M., Sharma, S., Booth, M. & Moran, M. A. Expression patterns reveal niche diversification in a marine microbial assemblage. *ISME J* **7**, 281-298, doi:10.1038/ismej.2012.96 (2013).
41 McCarren, J. *et al.* Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc. Nat. Acad. Sci. USA* **107**, 16420-16427, doi:10.1073/pnas.1010732107 (2010).
42 Lidbury, I., Murrell, J. C. & Chen, Y. Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. *Proc. Nat. Acad. Sci. USA* **111**, 2710-2715, doi:10.1073/pnas.1317834111 (2014).
43 Moran, M. A. & Durham, B. P. Sulfur metabolites in the pelagic ocean. *Nat Rev Microbiol* **17**, 665-678 (2019).
44 Mühlenbruch, M., Grossart, H. P., Eigemann, F. & Voss, M. Phytoplankton-derived polysaccharides in the marine environment and their interactions with heterotrophic bacteria. *Environ Microbiol* **20**, 2671-2685 (2018).
45 Obernosterer, I. & Herndl, G. J. Phytoplankton extracellular release and bacterial growth: dependence on the inorganic N: P ratio. *Mar Ecol Progr Ser* **116**, 247-257 (1995).
46 Bertrand, E. M. & Allen, A. E. Influence of vitamin B auxotrophy on nitrogen metabolism in eukaryotic phytoplankton. *Front Microbiol* **3**, 375, doi:10.3389/fmicb.2012.00375 (2012).
47 Amin, S. A. *et al.* Photolysis of iron-siderophore chelates promotes bacterial-algal mutualism. *P Natl Acad Sci USA* **106**, 17071–17076 (2009).

48 Casey, J. R., Mardinoglu, A., Nielsen, J. & Karl, D. M. Adaptive evolution of phosphorus metabolism in *Prochlorococcus*. *mSystems* **1**, e00065-00016 (2016).
49 Gianchandani, E. P., Chavali, A. K. & Papin, J. A. The application of flux balance analysis in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **2**, 372-382 (2010).
50 Pacheco, A. R., Moel, M. & Segre, D. Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat Commun* **10**, 103 (2018).
51 Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotech* **34**, 828 (2016).
52 Erbilgin, O. *et al.* MAGI: A method for metabolite, annotation, and gene integration. *ACS Chem Biol* (2019).
53 Djoumbou-Feunang, Y. *et al.* BioTransformer: a comprehensive computational tool for small molecule metabolism prediction and metabolite identification. *J Cheminformatics* **11**, 2 (2019).
54 Fu, H., Uchimiya, M., Gore, J. & Moran, M. A. Ecological drivers of bacterial community assembly in synthetic phycospheres. *P Natl Acad Sci USA* (in press).
55 Howard-Varona, C. *et al.* Phage-specific metabolic reprogramming of virocells. *ISME J*, 1-15 (2020).
56 Ryan, J. *et al.* Harmful phytoplankton ecology studies using an autonomous molecular analytical and ocean observing network. *Limnol Oceanogr* **56**, 1255-1272 (2011).
57 Nowinski, B. *et al.* Microbial metagenomes and metatranscriptomes during a coastal phytoplankton bloom. *Sci Data* **6**, 1-7 (2019).
58 Bernstein, D. B., Dewhirst, F. E. & Segre, D. Metabolic network percolation quantifies biosynthetic capabilities across the human oral microbiome. *eLife* **8**, e39733 (2019).
59 Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol* **3**, 804-873 (2018).
60 DiMucci, D., Kon, M. & Segre, D. Machine learning reveals missing edges and putative interaction mechanisms in microbial ecosystem networks. *mSystems* **3**, e00181-00118 (2018).
61 Lima-Mendez, G. *et al.* Determinants of community structure in the global plankton interactome. *Science* **348**, 1262073 (2015).
62 Levering, J. *et al.* Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLoS One* **11**, e0155038 (2016).
63 Zhao, Q., Stettner, A. I., Reznik, E., Paschalidis, I. C. & Segrè, D. Mapping the landscape of metabolic goals of a cell. *Genome Biol* **17**, 109 (2016).
64 Siegel, D. A. Resource competition in a discrete environment: Why are plankton distributions paradoxical? *Limnol Oceanogr* **43**, 1133-1146 (1998).
65 Saifuddin, M., Bhatnagar, J. M., Segrè, D. & Finzi, A. C. Microbial carbon use efficiency predicted from genome-scale metabolic models. *Nat Commun* **10**, 3568 (2019).
66 Szul, M. J., Dearth, S. P., Campagna, S. R. & Zinser, E. R. Carbon fate and flux in *Prochlorococcus* under nitrogen limitation. *mSystems* **4**, e00254-00218 (2019).
67 Nicholson, D. P., Stanley, R. H. & Doney, S. C. A phytoplankton model for the allocation of gross photosynthetic energy including the trade-offs of diazotrophy. *J Geophys Res: Biogeosciences* **123**, 1796-1816 (2018).
68 Bonan, G. B. & Doney, S. C. Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science* **359**, eaam8328 (2018).
69 Goldford, J. E., Hartman, H., Marsland, R. & Segrè, D. Environmental boundary conditions for the origin of life converge to an organo-sulfur metabolism. *Nat Ecol Evol* **3**, 1715-1724 (2019).
70 Friedrichs, M. A. *et al.* Assessment of skill and portability in regional marine biogeochemical models: Role of multiple planktonic groups. *J Geophys Res Oceans* **112** (2007).

71 Luo, Y.-W., Friedrichs, M. A., Doney, S. C., Church, M. J. & Ducklow, H. W. Oceanic heterotrophic bacterial nutrition by semilabile DOM as revealed by data assimilative modeling. *Aquatic Microb Ecol* **60**, 273-287 (2010).
72 Schwinger, J. *et al.* Nonlinearity of ocean carbon cycle feedbacks in CMIP5 earth system models. *J Climate* **27**, 3869-3888 (2014).
73 Moore, J. K., Lindsay, K., Doney, S. C., Long, M. C. & Misumi, K. Marine ecosystem dynamics and biogeochemical cycling in the Community Earth System Model [CESM1 (BGC)]: Comparison of the 1990s with the 2090s under the RCP4. 5 and RCP8. 5 scenarios. *J Climate* **26**, 9291-9312 (2013).
74 Hurrell, J. W. *et al.* The community earth system model: a framework for collaborative research. *Bull Am Meteorolog Soci* **94**, 1339-1360 (2013).
75 Follows, M. J., Dutkiewicz, S., Grant, S. & Chisholm, S. W. Emergent biogeography of microbial communities in a model ocean. *Science* **315**, 1843-1846 (2007).
76 Coles, V. *et al.* Ocean biogeochemistry modeled with emergent trait-based genomics. *Science* **358**, 1149-1154 (2017).
77 Litchman, E. *et al.* Global biogeochemical impacts of phytoplankton: a trait-based perspective. *J Ecol* **103**, 1384-1396 (2015).
78 NRC. *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. (National Academies Press, 2012).
79 NGSS. (2019).
80 ONL. (ed Ocean Literacy Network) (Ocean Literacy Network, 2013).
81 Haley, S. T. & Dyhrman, S. T. in *Science and Children* Vol. April/May 31-35 (National Science Teaching Association, 2009).
82 Sullivan, M. https://u.osu.edu/viruslab/outreach/ (2019).
83 Gibbs Jr, K. D. & Marsteller, P. (Am Soc Cell Biol, 2016).
84 James, S. M. & Singer, S. R. From the NSF: The National Science Foundation's investments in broadening participation in science, technology, engineering, and mathematics education through research and capacity building. *CBE—Life Sci Ed* **15**, fe7 (2016).
85 Bangera, G. & Brownell, S. E. Course-based undergraduate research experiences can make scientific research more inclusive. *CBE—Life Sci Ed* **13**, 602-606 (2014).
86 Hanauer, D. I. *et al.* An inclusive Research Education Community (iREC): Impact of the SEA-PHAGES program on research outcomes and student learning. *P Natl Acad Sci USA* **114**, 13531-13536 (2017).
87 Rodenbusch, S. E., Hernandez, P. R., Simmons, S. L. & Dolan, E. L. Early engagement in course-based research increases graduation rates and completion of science, engineering, and mathematics degrees. *CBE—Life Sci Ed* **15**, ar20 (2016).
88 Hidi, S. & Renninger, K. A. The four-phase model of interest development. *Ed Psychol* **41**, 111-127 (2006).
89 Abler, R., Coyle, E., Kiopa, A. & Melkers, J. in *2011 Frontiers in Education Conference (FIE).* T3F-1-T3F-7 (IEEE).
90 Wilson, Z. S. *et al.* Hierarchical mentoring: A transformative strategy for improving diversity and retention in undergraduate STEM disciplines. *J Sci Ed Technol* **21**, 148-156 (2012).
91 Alford, R. F., Leaver-Fay, A., Gonzales, L., Dolan, E. L. & Gray, J. J. A cyber-linked undergraduate research experience in computational biomolecular structure prediction and design. *PLoS Comp Biology* **13**, e1005837 (2017).
92 Eren, M. http://merenlab.org/momics/ (Univeristy of Chicago, 2019).

93 Estrada, M., Woodcock, A., Hernandez, P. R. & Schultz, P. Toward a model of social influence that explains minority student integration into the scientific community. *J Educat Psychol* **103**, 206 (2011).
94 Pfund, C., Handelsman, J. & Branchaw, J. *Entering mentoring*. (WH Freeman, 2014).
95 Byars-Winston, A. & Dahlberg, M. in *The Science of Effective Mentorship in STEMM* (eds A Byars-Winston & ML Dahlberg) (National Academies Press, Washington, D.C., 2019).
96 Rudolph, A. L., Holley-Bockelmann, K. & Posselt, J. PhD bridge programmes as engines for access, diversity and inclusion. *Nat Astron*, 1-6 (2019).
97 Park, S., Hironaka, S., Carver, P. & Nordstrum, L. (Carnegie Foundation for the Advancement of Teaching, 2013).
98 Allen, S. *et al.* (ed A. J. Friedman).
99 Yang, Y. & Barth, J. M. Gender differences in STEM undergraduates' vocational interests: People–thing orientation and goal affordances. *J Vocation Behav* **91**, 65-75 (2015).
100 NAS. *Undergraduate research experiences for STEM students: Successes, challenges, and opportunities*. (National Academies Press, 2017).
101 CAISE. *Center for Advancement of Informal Science Education: About CAISE*, <http://www.informalscience.org/about-caise> (2017).
102 CAISE. *Center for Advancement of Informal Science Education: Evaluation tools and instruments*, <https://www.informalscience.org/evaluation/evaluation-tools-instruments> (2017).
103 Grack Nelson, A., Goeke, M., Auster, R., Peterman, K. & Lussenhop, A. in *New Directions for Evaluation, 2019* Vol. 2019 (eds A. C. Fu, A. Kannan, & R. J. Shavelson) 59-86 (2019).
104 Bernard, R. E. & Cooperdock, E. H. No progress on diversity in 40 years. *Nat Geosci* **11**, 292 (2018).
105 DOE. *Integrated Postsecondary Education Data System*, <https://ncsesdata.nsf.gov/webcaspar/> (2019).
106 Higgins, S. First Generation Students in Higher Education. *UB ScholarWorks* (2017).
107 Dutt, K., Pfaff, D. L., Bernstein, A. F., Dillard, J. S. & Block, C. J. Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nat Geosci* **9**, 805-808 (2016).
108 Morgan, W. B., Elder, K. B. & King, E. B. The emergence and reduction of bias in letters of recommendation. *J Appl Soc Psychol* **43**, 2297-2306 (2013).
109 Houser, C. & Lemmons, K. Implicit bias in letters of recommendation for an undergraduate research internship. *J Furth High Edu* **42**, 585-595 (2018).
110 Bohon, W. *et al.* in *AGU Fall Meeting Abstracts.*
111 NAS. *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. (National Academies Press, 2014).
112 NAS. *Integrating discovery-based research into the undergraduate curriculum: Report of a convocation*. (National Academies Press, 2015).
113 Gentile, J., Brenner, K. & Stephens, A. *Undergraduate Research Experiences for STEM Students: Successes, Challenges, and Opportunities.*, (National Academies Press, 2017).
114 Brown, E. R., Thoman, D. B., Smith, J. L. & Diekman, A. B. Closing the communal gap: The importance of communal affordances in science career motivation. *J Appl Soc Psychol* **45**, 662-673 (2015).
115 Fischhoff, B. Evaluating science communication. *P Natl Acad Sci USA* **116**, 7670-7675 (2019).
116 Orr, J. *et al.* Biogeochemical protocols and diagnostics for the CMIP6 Ocean Model Intercomparison Project (OMIP). *Geosci. Model Dev.* **10**, 2169–2199 (2017).