Robotic oceanography: Revealing ocean-scale biochemical structure with a deep-diving autonomous vehicle

John Breier [1], Michael Jakuba [1], Mak A. Saito [1], Gregory Dick [1], Sharon Grim [1], Eric Chan [1], Matthew R. McIlvin [1], Dawn Moran [1], Brianna Alanis [1], Andrew Allen [1], Chris Dupont [1], and Rod Johnson [1]

[1]Affiliation not available

October 30, 2023

Abstract

This manuscript reports on a robot called *Clio* that we developed to facilitate basin-scale studies of ocean microbial communities and their biochemistry, to better understand how marine microorganisms regulate ocean and Earth system environmental cycles. *Clio* is designed to facilitate global-scale studies of ocean biochemistry, to move vertically through the water column with high precision, and specifically to return sensor data and samples from large swaths of the ocean ranging in depths from the surface to 6,000 m. Clio is capable of flexible, precise vertical motion that few other ocean robots can perform, and none to our knowledge over this depth range. We tested Clio extensively over several years, six cruises, and 26 dives, it is now fully operational and this manuscript describes all that we did to convince ourselves this was so. In June 2019, it completed its first large-scale ocean survey, and for which this manuscript will be the first data presentation.

Science Robotics: Research Article

Title:
Full: Robotic oceanography: Revealing ocean-scale biochemical structure with a deep-diving autonomous vehicle.
Short: A biochemical deep-diving autonomous vehicle.

Authors: John A. Breier[1*], Michael V. Jakuba,[2] Mak A. Saito[2], Gregory J. Dick[3,4,5], Sharon L. Grim[3,6], Eric W. Chan[1], Matthew R. McIlvin[2], Dawn M. Moran[2], Brianna A. Alanis[1], Andrew E. Allen[7], Chris L. Dupont[7], Rod Johnson[8]
Corresponding author: John.Breier@utrgv.edu

Affiliations:
[1]School of Earth, Environmental, and Marine Sciences, The University of Texas Rio Grande Valley, Edinburg, Texas, 78539, USA.
[3]Department of Earth and Environmental Sciences, the [4]Center for Computational Medicine and Bioinformatics, and the [5]Department of Ecology and Evolutionary Biology,
University of Michigan, Ann Arbor, Michigan, 48109, USA.
[6]now at Space Science and Astrobiology Division, NASA Ames Research Center, Moffett Field, California, 94035, USA.
[7]Microbial and Environmental Genomics, J. Craig Venter Institute, San Diego, California, 92121, USA.
[8]Bermuda Institute of Ocean Sciences, St. George's, GE 01, Bermuda.

ABSTRACT: Vast and diverse microbial communities exist within the ocean. To better understand the global influence of these microorganisms on Earth's climate, we developed a robot capable of sampling biochemistry across ocean basins while still capturing the fine-scale biogeochemical processes therein. Carbon and other nutrients are acquired and released by marine microorganisms as they build and break down organic matter. The vast scale of the ocean makes these processes globally relevant and, at the same time, challenging to fully characterize. Microbial community composition and ocean biochemistry vary across multiple physical scales up to that of the ocean basins. Other autonomous underwater vehicles are optimized for moving continuously and, primarily, horizontally through the ocean. In contrast, *Clio*, the robot we describe, is designed to efficiently and precisely move vertically through the ocean, and integrate with research vessel operations to map large horizontal scales to a depth of 6,000 m. We present results that show how *Clio* conducts high-resolution sensor surveys and sample-return missions, including a mapping of 1,144 kilometers of the Sargasso Sea to a depth of 1,000 m. We further show how the samples obtained enable genomic and proteomic measurements not possible through *in situ* sensing. These results demonstrate a robotic oceanography approach for global-scale surveys of ocean biochemistry.

SUMMARY: Autonomous sample-return and high-resolution depth control enable global-scale biochemical mapping of marine genomics and proteomics.

MAIN TEXT

INTRODUCTION

On Earth, the cycling and availability of carbon, nitrogen, phosphorus, and oxygen, as well as other essential nutrients and cofactors, are strongly influenced by processes that occur within the ocean over basin-scales*(1-3)*. Planktonic marine microorganisms, which collectively account for close to half of planetary primary production, drive these processes*(4)*. Redfield described the reciprocal, and chemical, relationship between ocean biology and the environment over 60 years ago, suggesting this relationship was the result of plankton communities building their cellular biochemical constituents*(5)*. Geochemists and ecologists have incorporated this relationship into large-scale conceptual models of nutrient flows, generally referred to as biogeochemical cycles. These models, by necessity, must attempt to simplify, aggregate, and summarize the net interactions between planktonic cellular biochemical processes and the large-scale marine biogeochemical environment. However, it has always been challenging to identify the essential processes to model and validate their representation and predictions. While our understanding of the processes connecting planktonic microbiology to ocean chemistry has grown substantially over the years*(6-8)*, our appreciation of the complexities involved in these relationships has grown too*(9-11)*. However, the recent inclusion of genomic*(12, 13)*, transcriptomic*(14)*, proteomic*(15)*, and metabolomic*(16)* measurements in ocean studies is creating new potential to understand the specific biochemical mechanisms coupling ocean life and environmental chemistry. Currently, there is keen interest in better understanding these natural processes*(17)*, and urgency to do so because anthropogenic processes are altering the physical and chemical environment of the ocean, with the potential to reduce the ocean's ability to buffer large-scale climatic shifts*(18)*.

To improve our understanding of Earth system biogeochemical cycles, we developed a robotic vehicle to facilitate ocean mapping of microbial and biochemical data (Fig. 1). This robot, called *Clio*, collects samples and measurements for microbial community proteomics, genomics, transcriptomics, organic and inorganic chemistry, and related physical data. These data types include biochemicals, sparsely mapped at present, and supporting parameters that must be measured synoptically. Sensors can measure only a subset of the necessary data; thus, sample-return for shore-based analyses is essential. *Clio* collects samples, and sensor data in the form of (i) high-resolution ocean-depth profiles, (ii) ocean-basin scale transects of the same, and (iii) targeted data collection from biochemical hotspots hidden within the ocean. While we can make some of the biochemical and chemical measurements with small sample volumes (milliliters to liters), we require large volumes for other measurements (e.g., proteomics, organic carbon, and trace metals). Therefore, *Clio* is capable of processing several hundreds of liters of seawater per sample to return material for subsequent high-throughput laboratory analyses. *Clio* is designed to dive to depths of 6,000 m and can, therefore, access most of the ocean. *Clio* operates autonomously without a tethered connection to a surface vessel; this frees the robot from the restricted motion a tether imposes, and it frees the surface vessel to carry out synoptic

science activities (Fig. S1). Before *Clio*, an ensemble of sequential techniques, involving the repeated lowering and recovery of cabled equipment from research vessels, was required to collect similar full ocean-depth datasets.

No existing vehicle design was well suited to diving deep but also systematically crossing an ocean in short order, with or without aid, or processing large volumes of water for sample-return (Fig. 1F). We addressed this challenge by developing a new robot for the most unfulfilled and currently limiting aspect of these objectives while making use of existing ocean surface vessels where they were most capable. Specifically, we optimized *Clio* to execute single vertical ocean data collection dives as effectively as possible. Because crossing the ocean surface is something research vessels do well, and Langragian motion is better for sampling, the *Clio* design foregoes any lateral control. During deployment, *Clio* is launched from a surface vessel and conducts an autonomous sample and data collection mission over a predetermined depth range. *Clio* is recovered by the surface vessel at the end of each dive, typically lasting 7 to 14 hours, and transported to the next dive location where this process is repeated (Fig. 2, Fig.S1). In this way, *Clio* can acquire data and samples within weeks that span ocean basins, enabling the creation of ocean-scale cross-sections of biochemical measurements and associated parameters, in other words, slices of vertical data that cut through the oceans.

Importantly, to achieve the widespread use required to conduct a global survey, the robot and its sampling systems must be practical and reproducible. It must be operable by small teams that can sustain regular back to back dives over several weeks. It must also be reliable enough to complete many dives to the bottom of the ocean during that time frame. It must be cost-effective to build, maintain, and operate so that it can be used repeatedly and replicated when necessary.

Global-scale ocean studies

In 1934, Redfield*(19)* hypothesized that the ratios of inorganic nutrients in seawater were in a reciprocal self-regulating relationship with marine microorganisms based on the analysis of global datasets. This was not obvious and only exists in the oceans because phytoplankton dominate marine biomass and the ocean is constantly mixing, but this only became apparent through the examination of global data sets. Subsequent global scale studies (e.g. WOCE and GEOTRACES) have also transformed our understanding of ocean processes (Supplemental Information: Global Ocean Studies)*(20-22)*.

An important part of these studies has been the systematic collection of data for the creation of ocean sectional maps. The field methods for creating ocean sectional maps have been generally consistent: (i) collect water samples or particulate material from multiple depths in a vertical profile from the surface to the deepest depth, (ii) repeat this process at a series of stations spread along a transect that crosses an ocean basin, and (iii) repeat these transects over the course of as many research cruises in order to create a network of sectional maps that cover the Earth. To date, ocean sectional studies rely on wire-based water column sampling technology

(Fig. S1B, Supplemental Information: Sectional Cruise Methodology) and typically require 30 to 50 days to complete.

The biochemistry of the oceans, including microbial community genomics, transcriptomics, proteomics, and metabolomics have yet to be widely measured. Global genomics surveys to date, such as the Global Ocean Sampling program*(23)* and the Tara Expedition*(24)*, have focused almost entirely on the sunlit euphotic zone. The biochemistry of the rest of the ocean is largely unmapped. This fact, and the results of previous ocean basin studies, are inspiring efforts to launch new global microbiology and biochemistry studies*(25)*, such as the proposed Biogeoscapes effort.

Clio is designed to enable global studies of this kind by speeding the collection of this type of data and the creation of sectional maps. *Clio* collects a full suite of samples during single vertical profile deployments in a fraction of the time required by current methods (Supplemental Materials: Background); and enables autonomous sample collection concurrent with existing wire-based techniques to augment existing ship-based methods. *Clio* is also designed to target features in the water column with higher spatial resolution than possible with wire-based techniques. *Clio* applies robotic underwater vehicle technology to ocean water column chemical studies.

Robotic oceanography for biogeochemical studies

Though robotic oceanography techniques have not been directly applied to the task of sample acquisition for global ocean sectional studies, they have been used to study ocean biogeochemical processes (Fig. 1F). Profiling floats, and more recently autonomous underwater gliders, are used to collect sensor data to support biogeochemical studies, and profiling floats were part of the WOCE study*(26)*. Remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs) equipped with sensors and samplers have been used to study biological and chemical processes in niche environments throughout the ocean. Their applications span a continuum between (i) collecting a core dataset at global scales to (ii) collecting intensive observations in niche environments.

Profiling floats and gliders are used for large-scale observations of specific core ocean parameters*(27-29)*. They are designed for long-term deployments lasting months to years. They control their depth by changing their buoyancy using variable ballast. They typically measure salinity, temperature, dissolved oxygen, chlorophyll, and optical proxies of particle concentration, but are used for other measurements as well*(30, 31)*. ARGO floats are one of the most common profiling float designs; over 4000 ARGO floats have been deployed throughout the ocean. They are typically deployed once for a lifetime duration of 4 or more years. During that time they drift in a Lagrangian manner. They only control their depth to a limited number of set points and most designs are limited to the upper 2,000 m of the ocean. Profiling gliders have wings, control pitch and roll, in addition to buoyancy, and can move horizontally as well as vertically*(32-34)*. These platforms typically have small payload and battery capacities that cannot accommodate large

sample processing systems. Both platforms are optimized for travel over great distances but over long time-scales that would (i) limit the sample collection rate even if payload constraints were not limiting and (ii) raise greater concern regarding sample preservation. At present, profiling floats and gliders are not currently well suited to the type of sample return necessary for biochemical sectional studies.

ROVs have been used as ocean observing platforms since the 1980s*(35-37)*. Among ocean vehicles, they have the highest payload capacities and are the most reconfigurable; as a result, they are excellent platforms for a wide range of intensive observing tasks as well as instrument development. They have enabled the biogeochemical study of seafloor chemosynthetic communities, including hydrocarbon seeps and hydrothermal systems. ROVs have been equipped with a wide variety of dedicated sample-return systems to support these studies including the Hydrothermal Fluid and Particulate Sampler, the Autonomous Microbial Sampler, the GeoMicrobe and Mobile Pumping System Instrument, and the Suspended Particulate Rosette Sampler*(38-41)*. The precision navigation capabilities of ROVs enable targeted sampling in and around seafloor seep and hydrothermal biological “hotspots” where the concentration of microbial life and chemical nutrients can be high, as a result, processing large samples in many cases is not necessary. However, collecting material emitted from these systems up into the water column as it mixes and dilutes with seawater does require processing large samples; the sampling systems developed for *Clio* inherit design principles from instruments developed to return samples from deep-sea hydrothermal plumes*(41, 42)*. Nevertheless, ROVs require cables, cables require winches, and cables and winches typically constrain the supporting research vessel to one activity at a time. Moreover, the design features that make ROVs versatile can also make them operationally and logistically complex. For ocean sectional studies, therefore, ROVs provide little benefit over existing cabled sampling technologies.

AUVs have been used as ocean observing platforms since the 1990s*(43-45)*. Optimized for longer distance surveys and more task-specific observations than ROVs, AUVs trade smaller payload capacities for lower drag and increased range. Fully autonomous, they can operate independently from surface ships. Most AUV designs have been multi-purpose. They are used widely for mapping operations, particularly at the seafloor*(45)*, but also under ice*(46)*. AUVs have also been adapted to the study of surface plankton processes*(47, 48)*. AUVs have been equipped with dedicated sample-return systems to support these studies, including the GULPER*(49)* and the Environmental Sample Processor*(50)*. AUV volume and payload constraints require a significantly higher degree of instrument customization and accommodation than for ROVs. In the coastal surface ocean, phytoplankton concentrations can be high; as a result, processing large samples is not necessary, just as it is not typically necessary at seafloor “hotspots.” Therefore, AUV sample-return systems designed for these surface ocean studies typically focus on the collection and processing of water volumes on the order of 1 to 2 L per sample. A few AUV applications have made use of deep-sea sampling system designs and are capable of collecting larger volumes for a wider range of biochemical analyses*(51, 52)*.

Moreover, to date, most AUV designs are based on the familiar torpedo shape and optimized for long-range lateral motion. For example, the Tethys Long Range AUVs have a range of at least 1800 km*(53)*; and, the long-range Autosub version is designed for a range of 6,000 km*(54)*. These AUVs and others like them can also dive deep, but they are not well suited to repeatedly profiling the full depth of the ocean nor doing so in a single deployment.

Clio is a reimagination of AUV design applied to vertical profiling. It builds upon sample-return instrumentation and techniques developed for the deep-sea, where concentrations are low, and applies them to the full-water column. It is designed and optimized for a single purpose and a mode of science operations orthogonal to previous AUV applications. Instead of collecting core datasets at global scales or intensively observing regional processes, we designed *Clio* to complete an intensive study of global scale biochemistry.

Experimental setting

We tested *Clio* extensively over several years, six cruises, and 26 dives. Initial sea trials were in 2017 near the shelf break of the eastern Atlantic coast of North America, just north of the Gulf Stream and the Sargasso Sea. During 2018 and 2019, we deployed *Clio* repeatedly, on a semi-quarterly basis, at the Bermuda Atlantic Time-series Study (BATS) station (31.66667° N, -64.16667° W) within the Sargasso Sea. The Sargasso Sea is an oligotrophic region of the North Atlantic Ocean contained within an ocean gyre bounded by the Gulf Stream, the North Atlantic Current, the Canary Current, and the North Atlantic Equatorial Current. The Sargasso Sea is one of the regions where Redfield made some of his initial observations in 1933*(19)*.

Contributions

Clio is unique from other ocean robots. It is designed to facilitate global-scale studies of ocean biochemistry, to move vertically through the water column with high precision, and specifically to return samples from large swaths of the ocean ranging in depths from the surface to 6,000 m. *Clio's* range encompasses full ocean depths everywhere except for hadal trenches. *Clio* is capable of flexible, precise vertical motion that few other ocean robots can perform, and none to our knowledge over this depth range. We demonstrate that *Clio* can move through the water column at a controlled rate as slow as desired or up to 0.8 m s^{-1} and settle at desired depths to within 50 cm in less than a minute. It can hold a vertical station within a depth window of less than 5 cm for multiple hours while sampling and sensing. *Clio* is isolated from the influence of ship heave, unlike for tethered instruments. Unlike other AUVs, *Clio* can hold station vertically while drifting in a lagrangian manner laterally. This allows high resolution, vertical profiling using both sensors and samplers that are ideal for the fine chemical gradients that exist within the ocean water column. Importantly for its design application, *Clio* has unique sample processing capabilities and can acquire and return a wider range of sample types, from larger water volumes, than any other ocean robot. From samples collected in the North Atlantic Ocean and the Sargasso Sea, we demonstrated that *Clio* can be used to make novel metagenomic and proteomic measurements as well as collect sensor data and samples for more routine nutrient

measurements. In June 2019, *Clio* completed its first sectional cruise in a series of 9 dives spanning a 1,144 km transect between Bermuda and Woods Hole, Massachusetts, USA, through the Sargasso Sea and the Gulf Stream. This transect focused on the upper 1,000 m of the water column, but also included *Clio*'s deepest dive to date, to 4,100 m. This demonstrated *Clio*'s ability to conduct ocean crossing sectional studies.

RESULTS

In this section, we describe our AUV design for biochemical ocean mapping. We also describe the design of this robot's sample-return systems, which it uses to manipulate seawater to extract samples for return. We then report on the results of field trials where the robot was used to conduct high precision sampling within the water column.

Designing a biochemical AUV for global ocean mapping

Clio's unique H-shape design is a synthesis of vehicle control strategy, deck-side operational considerations, and data collection criteria (Fig. 1, Materials and Methods: Vehicle design). We designed this robot to efficiently conduct single vertical profiles to 6,000 m, stopping and loitering at many arbitrary depths to collect data and samples for return. More specifically, we designed *Clio* to autonomously collect the observations made at sectional cruise stations (*e.g.*, GEOTRACES) but in a fraction of the time and surpassing their spatial resolution. These operations have to be repeated many times, so the robot has to be logistically simple to stage, deploy, recover, and ship worldwide. Importantly, the data, and the returned samples, must be collected in a form that is acceptable to the oceanographic research community and, to the maximum extent possible, readily inter-comparable with samples collected with other existing methods.

Clio's streamlined hull is symmetric about the horizontal plane such that drag is approximately equal during ascent and descent. Stability is achieved, for this otherwise unstable hull form, by using a large separation between the vehicle's centers of mass and buoyancy. This configuration is stable up to the speed at which the destabilizing hydrodynamic moment exceeds the hydrostatic stabilizing moment. Scale model tow tank tests validated this design prior to manufacturing, and the full-scale field tests confirmed it *(55)*. Operation above the critical speed does not endanger the vehicle but results in a limit cycle consisting of large (20° - 30°) oscillations in pitch and roll that decrease transit efficiency.

Clio is positively buoyant at all depths. Positive buoyancy mitigates the risk of vehicle loss in the event of a control system or actuator failure. Vertical thrusters control its depth. Variable-density ballast control was also considered; but, *Clio* must move rapidly between, and settle precisely at, target depths, and this is easier to achieve with active thrusting over this depth range. *Clio* must thrust downward to hold depth while sampling, and it spends most of its time during a deployment holding station and must expend energy to do so, but the energy expended for propulsion is comparable to the power it expends pumping seawater for sample collection. Unlike

other AUVs designed for deep-sea applications, *Clio* does not use expendable ballast to speed descent or for emergency ascents; the metals typically used for these weights represent cleanliness concerns, and their use also complicates vehicle operations and logistics. *Clio* uses ambient pressure air bladders to increase positive buoyancy at the surface where it is most valuable: to loft antennas above the sea surface. Their volume decreases rapidly with depth so that their impact on propulsion power consumption is negligible below approximately 100 m depth. To minimize the increase in positive buoyancy with depth as seawater density increases, we favored the selection of materials with lower bulk moduli to more closely match the average compressibility of seawater *(56)*.

Thrusters enable high-performance depth control for minimal complexity. *Clio* employs conventional proportional, integral, and derivative position control to command the desired thruster winding current and a Paroscientific Series 8000 Digiquartz pressure sensor running at 6 Hz to provide depth feedback. DC brushless motor controllers run their own internal control loop to regulate winding current. The desired thrust is saturated as a means of controlling speed between stations; however, the result is dependent on ballast condition. Closed-loop speed control will be implemented in the future.

Clio and its sample-return systems are sized and shaped in large part in order to use the same filter membranes commonly used for comparable oceanographic studies using existing wire-based sampling systems. The orientation of the sampler containers, such that filter membranes are parallel to the horizontal, is made to match that of existing systems. Samples collected in this way produce a radially uniform pattern across the filter surfaces, which is essential to achieve because the filters are subdivided based on this assumption; this is part of the methodology used with existing wire-based systems, and we match this with *Clio*. An alternative filter arrangement, e.g., parallel to the vertical, would require new method development, validation, and inter-calibration with existing methods. *Clio* is also constructed of materials selected to maximize sample integrity and cleanliness. In particular, ferrous alloys were all but eliminated, iron is a trace and limiting nutrient in seawater, even minute concentrations of anthropogenic iron introduced from ferrous alloys on the vehicle could influence the trace metal measurements or the biochemical measurements made with this system. In place of ferrous alloys, aluminum, titanium, and nickel alloys were used. For similar reasons, the use of oil-filled electrical junction boxes, standard on some ROVs and AUVs, were minimized and where possible liquid fluorocarbons were used in their place. When the mineral oils commonly used for these applications inevitably leak out of their reservoirs, they coat the outer surfaces of vehicle components with a residue that can remain over many dives and become entrained into the sampling systems, potentially biasing them for some organic analyses. The fluorocarbon substitutes are more volatile than mineral oils, and residues quickly evaporate, minimizing this concern*(57)*. The location of the thrusters and the shape of the vehicle are designed to minimize the influence of propeller wash on sensor data or the collected samples. The thrusters are located within vertical channels on the sides of the vehicles. The sensors are arranged at the bottom of the vehicle, which permits the collection of undisturbed sensor profile data during

descent. The intakes for the sample-return system are colocated with the sensors on sides orthogonal to the thrusters.

Manipulating seawater

Clio's primary task is to extract material from seawater and return it for analyses that can not be measured by *in situ* sensors. To this end, *Clio* has four large sampling payload bays (each 25 cm by 25 cm by 115 cm), each capable of containing independent, custom, high-volume *in situ* water filtration sampling systems as well as whole water and filtrate subsampling systems (Fig. 3A). Each sampler can collect up to 19 samples depending on media size. For operations to date, we outfitted *Clio* with sampling systems in just two payload bays and configured them to collect nine sample-sets each consisting of (i) particulate material filtered onto 142 mm diameter, 0.2 µm pore size media, (ii) 200 mL of filtrate, and (iii) 200 mL of unfiltered water (Supplemental Information: Methods). Particulate samples can be preserved *in situ* by preservative administration. Depending on biomass concentration, this configuration typically filters 40 to 300 L of water per sample and 1,500 to 3,500 L of water total per dive; even so, *Clio* has significant unused payload for future instrumentation (Table S1).

The sample processing and return systems are the most essential and unique tools the robot has for interacting with the environment (See Materials and Methods). They are effectively this robot's manipulators. They manipulate seawater to separate, concentrate, and extract different size classes of particulate material from the ocean. Cleanliness was a design concern throughout the vehicle, but especially so in the sample-return systems. Apart from electronics and motor parts contained within the liquid fluorocarbon filled stepper housing, nearly all other parts are plastics: high-density polyethylene, acetal copolymer, polycarbonate, polytetrafluoroethylene, and stereolithography resin (Durable, Formlabs). Titanium and nickel alloys were used selectively where strength was essential. Each sample container unit has an independent intake to preclude cross-contamination. The custom sample container units are designed to interconnect between the sampling valve and all other individual sample units in the set, and the valve, without the need for tubing (Fig. 3B); which is time-consuming to interconnect and difficult to clean*(42)*. The sample containers themselves are designed for critical cleaning and to facilitate sample extraction in shipboard fabricated cleanrooms (Fig. 2E&F, 3C).

Experimental setup

A variety of dive types were conducted during the field trials described here. During initial sea trials, a series of five dives were executed to increasing depths the last of which was to 2,000 m in a water depth of 4,100 m. During the dives at BATS station, we tested a number of vehicle configurations, diving to depths of 2,000 m. The most frequent dive profile to date were 1,000 m dives completed at BATS and during the Bermuda to Woods Hole transect. Near the end of the Bermuda to Woods Hole transect *Clio* completed a dive to 4,100 m. The configuration of the sample-return systems for these deployments is described in Supplemental Information: Biochemical Methods.

AUV performance, endurance, and power efficiency

The typical 1,000 m AUV profile took 16 hours to complete (*e.g.*, Fig. 4). While diving *Clio* was subject to Lagrangian horizontal transport dependent on the prevailing vertical currents. This is intentional; Lagrangian motion results in better samples that are more representative of the water mass being studied. As a result of Lagrangian transport, the AUV typically surfaced 2 to 15 kilometers from the launch site; in the most extreme case, during a profile at the edge of the Gulf Stream Current, the vehicle surfaced 34 kilometers from the launch site. The supporting research vessel typically transits at 15 km hr^{-1} and can typically reach the AUV in <1 hr, recovering the AUV back on to deck was completed in <1 hr depending on sea-state. This operational tempo is in keeping with our design objectives. Meeting these objectives required careful attention to understanding and optimizing the vehicle's ballasting and power demands.

Energy expended during dives consists of: (i) baseline power consumption including profiling sensors and sampling pumps, and (ii) propulsion power spent to overcome buoyancy while sampling, and to overcome drag, in addition, while transiting. As with all vehicles, this represents a tradeoff and every vehicle has an optimal speed at which range is maximized*(58)*. Slow transits reduce the energy needed to overcome drag, but increase the energy needed to sustain baseload and overcome buoyancy. The energy consumed overcoming buoyancy during the descent comes close to, but does not completely cancel, the energy negated by buoyancy on the ascent, due to asymmetry in the propellers. Fast transits reduce the energy needed to sustain baseload and overcome buoyancy but increase the energy needed to overcome drag. *Clio's* mission calls for relatively rapid transits where the power to overcome drag losses dominate. For most missions, the vehicle expends the majority of its battery energy while holding depth and sampling. Vehicle buoyancy represents a purely parasitic load while sampling, and therefore the accurate prediction of propulsion power while holding depth, which varies as a function of pressure, temperature, and *in situ* seawater density, is necessary to maximize endurance while ensuring vehicle safety. The 26 dives to date have been used to test our understanding of the vehicle's drag, buoyancy, and coefficients of thermal expansion and bulk modulus; in these dives thrust, power, and ballasting conditions were measured using the approach described in Supplemental Information: Engineering Methods.

Transit propulsion power depends on buoyancy and speed. Fig. 5 illustrates the propulsion power consumed while transiting as a function of transit speed for various ballasts. Conservative ballasting (vehicle very buoyant) increases descent power but decreases ascent power. Descent power typically exceeds ascent power, a consequence of the additional power required to overcome buoyancy when descending, but at sufficiently high speed the opposite holds, a consequence of asymmetry in the propellers, which are designed for high efficiency while thrusting down to hold depth. Below the critical speed power approximately scales as the velocity cubed. Above the critical speed of about 0.75 m s^{-1} the power required increases dramatically; however, the precise speed at which this transition occurs varies, perhaps as a result of trim. The critical speed is also somewhat different for ascent versus descent because of

the relative positions of the center of mass and aerodynamic center change. We have observed stable transit at speeds as high as 0.9 m s. The round-trip transit energy consumption including baseload, expressed as % battery km^{-1} descent and ascent varies between 3% and 5% near the critical speed across the practical range of ballasts. For an aggressively ballasted vehicle, this figure holds throughout the water column despite the increase in buoyancy with depth.

Clio's buoyancy changes with depth as the density of seawater and all of the materials comprising the vehicle respond to differences in temperature and pressure. *Clio's* deepest dive, to 4,100 m, afforded an opportunity to test our understanding of the vehicle's net coefficient of thermal expansion and net bulk modulus. Analysis of calibrated thruster data from *Clio*'s dive to 4,100 m indicates a best fit coefficient of thermal expansion of 3.65e-05 m m^{-1} K^{-1} and a net bulk modulus of 3.3 GPa. This is 1.5 times less compressible than typical seawater (2.3 GPa) and favorable compared to most deep-diving AUVs that are typically 2 to 5 times less compressible than seawater *(59)*. For a vehicle of *Clio's* size the typical range corresponds to 90 N to 150 N in added buoyancy at 6000 m, i.e. a significant increase in parasitic propulsion load. The value for *Clio* is 50 N.

Fig. 5 illustrates the power consumed while holding depth to collect samples. Background vehicle processes and sensors consume 1.2% of battery capacity hr^{-1}. Each additional pump used while sampling consumes another 0.4% of battery capacity hr^{-1}. The energy consumed by the thrusters varies with depth and depends on the surface ballast condition. For conservative ballasting, *i.e.*, 100 N buoyant or 1.4% of vehicle displacement, energy consumption is 1.5% to 3% of battery capacity hr^{-1}. For a more aggressive ballasting state, i.e., 50 N buoyant or 0.7% of vehicle displacement, energy consumption is reduced to 1% to 2% of battery capacity hr^{-1}. Propulsion energy consumption is lowest at intermediate depths 100 m to 1,000 m where the air bladder volume has been compressed and vehicle buoyancy has yet to be significantly influenced by pressure or increasing background density. Fig. 5 shows that with aggressive ballasting, the energy expended by propulsion while holding depth is commensurate with the energy required for sampling throughout *Clio's* 6,000 m depth range, thus validating a key assumption made during the design process that chose thrusters as the sole means of depth control.

Baseline cleanliness and genomic resolution

Baseline cleanliness was assessed using genomic analyses of samples returned from a *Clio* profile. cDNA sequences were scanned for putative contaminant taxa identified by Salter et al.*(60)* and the interpretation of Sheik et al.*(61)* for low-biomass sequenced samples from the Census of Deep Life. Nodes that match in taxonomy to these were flagged as putative contaminants (Fig. S2, Table S2). In this profile, we assessed the background cDNA sequence contamination to be between 0.2% to 4.5%, and 1% on average.

Genomic analyses also show the microbial community structure and genomic resolution achieved by *Clio* (Fig. 6, Table S2). Fig. 6 shows the results from nine samples collected by *Clio* from a depth of 30 m to 2,000 m depth and analyzed for microbial DNA and RNA measured as

cDNA (Supplementary Materials: Biochemical Methods). Up to 38,280 DNA sequences and 24,493 cDNA sequences were obtained from filtered volumes of 18 to 57 L, using 1/16 split portions of the filtered biomass. In two cases, insufficient biomass or extraction methods, or both, resulted in low cDNA yields (Table S2: samples D5 & D17); in all other cases, filter biomass was sufficient to generate sequencing libraries with good resolution that were representative of the water column community. These results demonstrate core functionality, baseline necessary volumes, and illustrate, to some degree, the amount of data returned from these biochemical measurements and contained within just one profile.

The genomic data in Figure 6A&B is categorized at the phyla level into 27 groups, but this is just a fraction of the data. A deeper characterization is possible within each of these levels. Figure 6C&D shows the classification of the proteobacteria phyla into a further 37 clades. Moreover, the cDNA is itself representative of only a fraction of the RNA transcriptome that can be measured for these groups. Nor does this include the proteomic and metabolomic datasets that can be acquired from the same samples.

High-resolution proteomic measurements

The proteomic analysis of samples returned from a *Clio* profile collected at BATS station shows the novel biochemical measurements that can be made by this method (Figures 7-8, Table S3). Proteomic analysis can only be performed on returned samples, as is true of metabolomic analysis as well. Figure 7A&B shows the results from ten samples of microbial biomass collected by *Clio* on April 14, 2018, at BATS station, and analyzed using metaproteomic methods (Supplementary Materials: Biochemical Methods) (Tables S3&S4). From these samples, 1,577 proteins were identified using high-resolution liquid chromatography-mass spectrometry from filtered volumes of 56 - 135 L per depth (similar to Figure 4), using 1/2 split portions of the filtered biomass*(15, 62)*. The vertical distributions of these proteins follow the niche depth range for the various microbial groups, as illustrated by a subset of proteins taxonomically attributed to *Prochlorococcus, Synechococcus*, and *Pelagibacter* (Fig. 7B).

Clio can also conduct higher resolution sampling than existing wire-based techniques. To demonstrate this, on October 24, 2018, at BATS station, *Clio* collected five samples across a chlorophyll maximum at 5 m increments while holding depth to a precision of <5 cm (Fig. 7C&D, Tables S5-S8). In comparison, simultaneously deployed wire-deployed *in situ* pumps varied in depth keeping by more than 15 m due to high winds and waves. Metaproteomic analysis of these targeted, high resolution *Clio* samples revealed distinct depth differences within the 5 m sampling increments, with numerous proteins at the top (Fig. 7E) of the chlorophyll feature being significantly more abundant than those at the center, and vice versa (Fig 7F). These results are consistent with prior observations, including for several *Prochlorococcus* proteins: urea and phosphate transporters were more abundant above the maximum where nitrogen is scarcest (Fig 7E&G), and the iron-free electron transport protein flavodoxin was most abundant at the chlorophyll maximum where iron scarcity is known to occur (Fig 7F&G). Nutrient stress biomarkers such as these are valued for their ability to diagnose the environmental stresses

faced by each major species present and how their biochemical responses change in time and space*(15)*. In addition, different isoforms of key biochemical proteins such as glyceraldehyde 3-phosphate dehydrogenase were observed with depth corresponding to high and low light *Prochlorococcus* ecotypes. These results demonstrate how *Clio*'s unique depth keeping capabilities can lead to new observations that are otherwise difficult if not impossible.

Sectional transect

During June 2019, in a series of 9 dives spanning the 1,144 km transect between Bermuda and Woods Hole, Massachusetts, USA, we used *Clio* to collect (i) 158 sample sets for proteins, metals, and nutrients (Table S9), and (ii) complementary sensor data (Table S10). *Clio*'s optical beam transmission data indicates how low the concentration of suspended material was in this region (Fig. S3). Consequently, *Clio* had to filter 20,878 L of seawater to collect these samples. The biochemical data from this cruise is being analyzed and will be reported elsewhere. Fig. 8A shows the level 3, monthly averaged, chlorophyll-a estimates from the MODIS-Aqua spectroradiometer for June 2019 (see Materials and Methods), as well as the *Clio* stations for this transect. Fig. 8 also shows the results for chlorophyll-a, measured by *Clio*'s fluorometer, as well as dissolved inorganic nitrogen measured from samples collected by *Clio*. A comparison of the remotely sensed MODIS-Aqua chlorophyll-a to the *in situ Clio* observations clearly shows the presence of a deep chlorophyll maxima (DCM) layer that occurs at depths between 80 to 150 meters throughout this region (Fig. 8B). This feature indicates a region of active primary productivity that can not be observed by remote sensing techniques and must be observed from within the water column.

During this transect, we also implemented an adaptive sampling algorithm to better target discrete features like the DCM. During *Clio* dives 024 and 025, the depths of a subset of stations were left unspecified with instructions to sample the DCM. *Clio* autonomously determined the depth of the DCM based on its chlorophyll fluorometer during descent. After completing several preprogrammed stations at deeper depths, *Clio* ascended to distribute the remaining stations at depths across the DCM. This adaptive sampling approach is similar to that taken by Zhang et al. *(63)* to sample thin phytoplankton layers using a conventional AUV for lateral surveys. This capability, coupled with high-resolution sampling, will enable even finer sampling resolution of discrete features like the DCM in the future.

The sectional cruise was completed in 14 days. If the same set of samples were collected without *Clio*, we estimate the cruise would have taken twice as long. Moreover, this transect is a relatively short distance for ocean sectional studies; the absolute time savings would typically be greater. During this section, *Clio* also completed a profile to a depth of 4,100 m. Collectively this data and the reliable performance of the robot indicate that *Clio* can complete the sectional cruises we designed it to perform.

Biochemical significance of trials

During the BATS deployments and the sectional cruise, we used *Clio's* sampling abilities to better understand the vertical structure of plankton communities and their relationship to the available nutrients in the environment. The thousands of functional proteins observed represent a tremendous amount of information, but the characteristics of the ecosystem can be discerned from a few example proteins. Key proteins increase in response to extreme nutrient scarcity for nitrogen, phosphorus, and iron, for example, vertical distributions of a urea transporter, alkaline phosphatase, and flavodoxin from the cyanobacterium *Prochlorococcus* reflect adaptations to scarce N, P, and Fe (Fig. 7G). Similarly, proteins can be indicative of key metabolic processes such as photosynthetic carbon fixation (PsbO and CsoS2; Figs. 7F-G) from *Prochlorococcus* and the use of the organic molecule taurine as a substrate for respiration in the bacterium *Pelagibacter* (Fig. 7B). Together these signals can provide a powerful assessment of the metabolic capabilities of each major microbial taxon present in the community, indicating what is controlling their growth and their contribution to the marine carbon cycle at each point in space and time.

DISCUSSION

These results demonstrate that *Clio* can be used to observe ocean biochemistry across large sections of the ocean. We showed that the vehicle can efficiently transit the water-column while stopping and loiter precisely at depths as needed. We showed how the vehicle, its sample processing systems, and the methods for using them, produce samples rich in genomic information with low background contamination levels. We also showed that *Clio* returns samples for novel metaproteomic analysis that can only be realized from the processing of large volumes of water, significantly larger than can be collected by other current AUVs. We demonstrated how *Clio* maps chemical and physical properties across sections of the ocean using both (i) data collected from its sensors, e.g., chlorophyll, salinity, temperature, and optical backscatter, and (ii) data produced from analysis of the samples it collected. In this paper, we showed, as an example, how the sectional map of chlorophyll-a, measured *in situ* by *Clio's* sensor, reveals the deep chlorophyll maximum environment hidden from remote satellite sensors. We also showed, using nitrate plus nitrite data, how the analysis of returned samples collected by this method can reveal the distribution of biochemically relevant ocean parameters that no sensor can measure. The genomic and proteomic measurements both illustrate the richness of these datasets; however, the data we showed is only a limited fraction of what can be measured in these analyses. Further, the data we showed represents only a partial set of the important biochemical compounds that could be measured with this method. The datasets that *Clio* enables, therefore, are in a class of "big data" that is challenging to represent. In the simplest case, if a distinct sectional map were created for every biochemical compound so measured, the result would be many 10,000s of mapping products. The oceanographic community is now only at the beginning of the process of studying this wide range of compounds and understanding their role in the cycling of nutrients between the environment and living systems. Our goal is that *Clio*, and robots like it, will contribute to a global-scale biochemical mapping of the ocean *(64)*, allowing researchers to answer fundamental questions about the limiting constraints on ocean life and the role of ocean life in regulating Earth's chemical environment. Moreover, our goal is that *Clio* will

also accelerate the pace of this research so that we can all be better informed as to how the ocean and Earth's climate are changing and what we can do to mitigate the negative consequences of these changes.

Limitations and future steps

A notable limitation to *Clio* is that it is not suited to under-ice applications. We designed *Clio* to control its position solely in the vertical dimension and intentionally to allow its lateral movement to be Lagrangian, because this results in better samples. If used under-ice, *Clio* could not return to the launch point. However, the sample return systems developed for *Clio* can be used on other ocean vehicles, such as the hybrid ROV NUI *(65)*, that do have lateral control and are well matched to under-ice operations. Also, we intentionally designed this version of *Clio* with a maximum depth of 6,000 m; therefore, it can not be used within hadal trenches. Hadal trenches are proportionally insignificant in terms of the ocean's influence on Earth system nutrient cycles, however, they are interesting environments for studying the influence of extreme pressures on life processes; a hadal version of *Clio* is possible.

MATERIALS AND METHODS

Vehicle design

Clio is similar in size to CTD-water rosettes (Fig. 1F), which are common to oceanographic research vessels, and can be deployed, recovered, and shipped with similar equipment. On deck, the vehicle stands vertically on integral landing skids and can be launched and recovered from this position without a cradle. The vehicle has an aluminum structural chassis that supports all components including the: (i) two vertical thrusters, (ii) three pressure housings for electronics and batteries below, (ii) syntactic foam flotation above, (iv) four sample-return payload bays around the perimeter of the mid-section, and (v) sensors near the bottom (Fig. 1). The vehicle's free flooded internal volume, including the sample-return payload bays, is enclosed in high-density polyethylene shells. These shells are removable, which provides operators easy access to the sample-return systems and the power and communication cables (Fig. 2B). At the top of the vehicle is an acoustic modem that allows operators to acoustically range on the vehicle and receive regular vehicle status updates. Also on top are three independently-powered depth-activated beacons, a GPS/Iridium unit to provide an over-the-horizon location, a strobe for line-of-sight location, and a VHF radio beacon, which alert operators to the vehicle's position when it surfaces. The computer systems, batteries, and thrusters leverage designs inherited from previous deep-sea AUVs: *ABE (66)*, *Sentry (67)*, *Nereus (68)*, and *NUI (65). Clio* carries a Paroscientific pressure sensor, a Seabird 49 Fastcat CTD, and typically also a Wetlabs 25 cm pathlength transmissometer, a Wetlabs ECO FL chlorophyll-a fluorometer, and an Aanderra dissolved oxygen optode. All AUV components are pressure rated to 6,000 m or greater.

Sampling system design

The *Clio* sampling systems inherit several design features and principles from previous deep-sea ROV and AUV sampling systems*(41, 42, 51)*; but their overall design, and many features, have

been uniquely developed for this open-ocean water column sampling. Specifically, they employ custom high flowrate valves to speed sample collection and minimize power consumption while pumping; the diameter of the fluidic path within the valves and the rest of the fluidic system is never smaller than 9.5 mm. Each valve has 19 ports and is actuated by a direct drive NEMA 42 stepper motor with a bipolar torque of 40 N-m. The stepper motors are equipped with absolute encoders to ensure their position is readily known at all times. The sample containers are first, followed by the control valve, then a deep-sea positive displacement pump (8 L min^{-1} pump head, Mclane Research Laboratories) that pulls seawater through the system, and last a digital flow meter (SPX model, Seametrics) that records sample flowrate and volume. The flow rate through the system depends on the filter media used and amount of suspended particulate material in the sample but typically the average sampling flowrate achieved is 1-2 L min^{-1}. The control and motor power electronics for the samplers are pressure tolerant designs located within the fluorocarbon filled stepper motor housings.

For *Clio*, we developed custom sample container units to maximize sample capacity and cleanliness (Fig. 3B). These units contain three compartments: a compartment within the upper housing that retains unfiltered seawater, a compartment that contains a filter membrane, and a lower compartment that retains filtered seawater. A fourth compartment can be added at the intake to separate and retain course material (e.g., zooplankton), or this material can be separated using an additional filter mesh in the filter compartment. Each of these compartments is separated from each other by one-way silicone valves; the intake also has one-way silicone valves. These valves are normally closed when there is no differential pumping pressure, and the compartments are isolated from each other and the environment. When a sample container is connected to the pump by the valve and pumping starts, the valves open passively, and water can flow through the system. In the case of time-sensitive, biochemical measurements, particularly for RNA, the filter compartment can be filled with a preservative by back-injection from a bladder contained in the lower compartment by pumping slowly in reverse. Smaller one-way valves in the back-flow direction allow for displacement of seawater with the preservative.

MODIS-Aqua satellite data

Monthly averaged MODIS-Aqua level 3, mapped, chlorophyll-a data for June 2019 was retrieved from NASA*(69)* and the colormap in Fig. 8, used for this data, was made with the ESA SNAP software tool.

REFERENCES

1. K. R. Arrigo, Marine microorganisms and global nutrient cycles. *Nature.* **437**, 349-355 (2005).
2. M. J. Follows, S. Dutkiewicz, S. Grant, S. W. Chisholm, Emergent biogeography of microbial communities in a model ocean. *Science.* **315**, 1843-1846 (2007).
3. G. J. Herndl, T. Reinthaler, Microbial control of the dark end of the biological pump. *Nat. Geosci.* **6**, 718-724 (2013).
4. D. A. Bryant, The beauty in small things revealed. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9647-9649 (2003).
5. A. C. Redfield, The biological control of chemical factors in the environment. *Am. Sci.* **46**, 205-221 (1958).
6. J. B. Waterbury, S. W. Watson, R. R. Guillard, L. E. Brand, Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature.* **277**, 293-294 (1979).
7. S. J. Giovannoni, M. S. Rappé, K. L. Vergin, N. L. Adair, 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the green non-sulfur bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 7979-7984 (1996).
8. A. K. Hawley, H. M. Brewer, A. D. Norbeck, L. Paša-Tolić, S. J. Hallam, Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11395-11400 (2014).
9. M. V. Brown, M. Ostrowski, J. J. Grzymski, F. M. Lauro, A trait based perspective on the biogeography of common and abundant marine bacterioplankton clades. *Mar. Genomics.* **15**, 17-28 (2014).
10. Z. I. Johnson *et al.*, Niche partitioning among prochlorococcus ecotypes along ocean-scale environmental gradients. *Science.* **311**, 1737-1740 (2006).
11. G. Rocap *et al.*, Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature.* **424**, 1042-1047 (2003).
12. S. Ganesh, D. J. Parris, E. F. DeLong, F. J. Stewart, Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *Isme J.* **8**, 187-211 (2014).
13. N. Kashtan *et al.*, Fundamental differences in diversity and genomic population structure between Atlantic and Pacific Prochlorococcus. *Isme J.* **11**, 1997-2011 (2017).
14. K. Anantharaman, J. A. Breier, C. S. Sheik, G. J. Dick, Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 330-335 (2013).
15. M. A. Saito *et al.*, Needles in the blue sea: Sub‐species specificity in targeted protein biomarker analyses within the vast oceanic microbial metaproteome. *Proteomics.* **15**, 3521-3531 (2015).
16. C. L. Fiore, K. Longnecker, M. C. Kido Soule, E. B. Kujawinski, Release of ecologically relevant metabolites by the cyanobacterium S ynechococcus elongatus CCMP 1631. *Environ. Microbiol.* **17**, 3949-3963 (2015).
17. V. Marx, When microbiologists plunge into the ocean. *Nature Methods.*, 1548-7105 (2020).
18. S. C. Doney, The growing human footprint on coastal and open-ocean biogeochemistry. *Science.* **328**, 1512-1516 (2010).
19. A. C. Redfield, On the proportions of organic derivatives in sea water and their relation to the composition of plankton *in James Johnstone Memorial Volume*, R. J. Daniel, Ed. (Liverpool University Press, Liverpool, 1934), pp. 176-192.
20. A. Ganachaud, C. Wunsch, Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature.* **408**, 453-457 (2000).
21. W. S. Broecker, Thermohaline circulation, the Achilles heel of our climate system: will

man-made CO_2 upset the current balance? *Science.* **278**, 1582 (1997).
22. M. J. R. Fasham *et al.*, A new vision of ocean biogeochemistry after a decade of the Joint Global Ocean Flux Study (JGOFS). *Ambio.* **30**, 4-30 (2001).
23. D. B. Rusch *et al.*, The Sorcerer II global ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
24. P. Bork *et al.*, *Tara* Oceans studies plankton at planetary scale. *Science.* **348**, 873 (2015).
25. A. E. Santoro, Crystal ball: the microbial map of the ocean. *Environ. Microbiol. Rep.* **11**, 35-37 (2019).
26. A. Ganachaud, C. Wunsch, Oceanic nutrient and oxygen transports and bounds on export production during the World Ocean Circulation Experiment. *Global Biogeochem. Cycles.* **16**, 1057-14 (2002).
27. M. A. Spall, Circulation in the Canary Basin: A model/data analysis. *J. Geophys. Res. Oceans.* **95**, 9611-9628 (1990).
28. L. Wu, Z. Jing, S. Riser, M. Visbeck, Seasonal and spatial variations of Southern Ocean diapycnal mixing from Argo profiling floats. *Nat. Geosci.* **4**, 363-366 (2011).
29. Y. Ohno, T. Kobayashi, N. Iwasaka, T. Suga, The mixed layer depth in the North Pacific as detected by the Argo floats. *Geophys. Res. Lett.* **31**(2004).
30. N. L. Williams *et al.*, Calculating surface ocean pCO2 from biogeochemical Argo floats equipped with pH: An uncertainty analysis. *Global Biogeochem. Cycles.* **31**, 591-604 (2017).
31. J. K. B. Bishop, T. J. Wood, Year-round observations of carbon biomass and flux variability in the Southern Ocean. *Global Biogeochem. Cycles.* **23**, GB2019-n/a (2009).
32. M. J. Perry, B. S. Sackmann, C. C. Eriksen, C. M. Lee, Seaglider observations of blooms and subsurface chlorophyll maxima off the Washington coast. *Limnol. Oceanogr.* **53**, 2169-2179 (2008).
33. D. C. Webb, P. J. Simonetti, C. P. Jones, SLOCUM: an underwater glider propelled by environmental energy. *IEEE J. Ocean. Eng.* **26**, 447-452 (2001).
34. S. Glenn *et al.*, Glider observations of sediment resuspension in a Middle Atlantic Bight fall transition storm. *Limnol. Oceanogr.* **53**, 2180-2196 (2008).
35. D. Yoerger, J. Newman, J. -. Slotine, Supervisory control system for the JASON ROV. *IEEE J. Ocean. Eng.* **11**, 392-400 (1986).
36. D. M. Karl, A. M. Brittain, B. D. Tilbrook, Hydrothermal and microbial processes at Loihi Seamount, a mid-plate hot-spot volcano. *Deep Sea Res. Part A Oceanogr. Res. Pap.* **36**, 1655-1673 (1989).
37. B. I. Bergström, J. Larsson, J. -. Pettersson., Use of a remotely operated vehicle (ROV) to study marine phenomena: I. Pandalid shrimp densities. *Mar. Eco. Prog. Ser.* **37**, 97-101 (1987).
38. J. P. Cowen *et al.*, Advanced instrument system for real-time and time-series microbial geochemical sampling of the deep (basaltic) crustal biosphere. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **61**, 43-56 (2012).
39. C. D. Taylor *et al.*, Autonomous Microbial Sampler (AMS), a device for the uncontaminated collection of multiple microbial samples from submarine hydrothermal vents and other aquatic environments. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **53**, 894-916 (2006).
40. D. A. Butterfield *et al.*, Mixing, reaction and microbial activity in the sub-seafloor revealed by temporal and spatial variation in diffuse flow vents at Axial Volcano. *The Subseafloor Biosphere at Mid-Ocean Ridges.* **144**, 269-289 (2004).
41. J. A. Breier *et al.*, A suspended-particle rosette multi-sampler for discrete biogeochemical sampling in low-particle-density waters. *Deep Sea Res. Part I Oceanogr. Pap.* **56**, 1579-1589 (2009).

42. J. A. Breier *et al.*, A large volume particulate and water multi-sampler with *in situ* preservation for microbial and biogeochemical studies. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **94**, 195-206 (2014).
43. B. Allen *et al.*, REMUS: a small, low cost AUV; system description, field trials and performance results. *OCEANS '97 MTS/IEEE Conference Proceedings.* **2**, 994-1000 vol.2 (1997).
44. H. Schmidt *et al.*, Real-time frontal mapping with AUVs in a coastal environment. *OCEANS '96 MTS/IEEE Conference Proceedings.* **3**, 1094-1098 vol.3 (1996).
45. M. A. Tivey, H. P. Johnson, A. Bradley, D. Yoerger, Thickness of a submarine lava flow determined from near-bottom magnetic field mapping by autonomous underwater vehicle. *Geophys. Res. Lett.* **25**, 805-808 (1998).
46. K. W. Nicholls, E. P. Abrahamsen, K. J. Heywood, K. Stansfield, S. Østerhus, High-latitude oceanography using the autosub autonomous underwater vehicle. *Limnol. Oceanogr.* **53**, 2309-2320 (2008).
47. T. O. Fossum *et al.*, Toward adaptive robotic sampling of phytoplankton in the coastal ocean. *Sci. Robot.* **4**, eaav3041 (2019).
48. K. M. Yamahara *et al.*, *In-situ* autonomous acquisition and preservation of marine environmental DNA using an autonomous underwater vehicle. *Front. Mar. Sci.* **6**, 373 (2019).
49. L. E. Bird, A. Sherman, J. Ryan, Development of an active, large volume, discrete seawater sampler for autonomous underwater vehicles. *OCEANS '07 MTS/IEEE Conference Proceedings.*, 1-5 (2007).
50. Y. Zhang *et al.*, Autonomous tracking and sampling of the deep chlorophyll maximum layer in an open-ocean eddy by a long-range autonomous underwater vehicle. *IEEE J. Ocean. Eng.*, 1-14 (2019).
51. D. L. Valentine *et al.*, Autonomous marine robotic technology reveals an expansive benthic bacterial community relevant to regional nitrogen biogeochemistry. *Environ. Sci. Technol.* **50**, 11057-11065 (2016).
52. A. F. Govindarajan, J. Pineda, M. Purcell, J. A. Breier, Species- and stage-specific barnacle larval distributions obtained from AUV sampling and genetic analysis in Buzzards Bay, Massachusetts, USA. *J. Exp. Mar. Biol. Ecol.* **472**, 158-165 (2015).
53. B. W. Hobson *et al.*, Tethys-class long range AUVs - extending the endurance of propeller-driven cruising AUVs from days to weeks. *2012 IEEE/OES Autonomous Underwater Vehicles.*, 1-8 (2012).
54. D. T. Roper *et al.*, Autosub long range 1500: An ultra-endurance AUV with 6000 km range. *OCEANS '17 MTS/IEEE Conference Proceedings.*, 1-5 (2017).
55. M. V. Jakuba, J. A. Breier, D. Gomez-Ibanez, K. Tradd, M. A. Saito, *Clio*: an autonomous vertical sampling vehicle for global ocean biogeochemical mapping. *2018 IEEE/OES Autonomous Underwater Vehicles.*, 1-8 (2018).
56. S. Levitus, *Climatological atlas of the world ocean* (US Department of Commerce, National Oceanic and Atmospheric Administration, Princeton, N.J., 1982).
57. J. Bell, J. Betts, E. Boyle, MITESS: A moored in situ trace element serial sampler for deep-sea moorings. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **49**, 2103 (2002).
58. A. M. Bradley, M. D. Feezor, H. Singh, F. Yates Sorrell, Power systems for autonomous underwater vehicles. *IEEE J. Ocean. Eng.* **26**, 526-538 (2001).
59. C. Eriksen, Systems and methods for compensating for compressibility and thermal expansion coefficient mismatch in buoyancy controlled underwater vehicles. U.S Patent 8,726,827, May 20 (2014).
60. S. J. Salter *et al.*, Reagent and laboratory contamination can critically impact

sequence-based microbiome analyses. *BMC Biol.* **12**, 87 (2014).
61. C. S. Sheik *et al.*, Identification and removal of contaminant sequences from ribosomal gene databases: lessons from the Census of Deep Life. *Front. Microbiol.* **9**, 840 (2018).
62. M. A. Saito *et al.*, Progress and challenges in ocean metaproteomics and proposed best practices for data sharing. *J. Proteome Res.* **18**, 1461-1476 (2019).
63. Y. Zhang, R. S. McEwen, J. P. Ryan, J. G. Bellingham, Design and tests of an adaptive triggering method for capturing peak samples in a thin phytoplankton layer by an autonomous underwater vehicle. *IEEE J. Ocean. Eng.* **35**, 785-796 (2010).
64. M. A. Saito, C. Breier, M. Jakuba, M. McIlvin, D. Moran, Envisioning a chemical metaproteomics capability for biochemical research and diagnosis of global ocean microbiomes *in The chemistry of microbiomes* (National Academies Press, Washington, D.C. ,2017), pp. 29-36, chap. 5.
65. M. V. Jakuba *et al.*, Teleoperation and robotics under ice: Implications for planetary exploration. *2018 IEEE Aerospace Conference.*, 1-14 (2018).
66. D. R. Yoerger, A. M. Bradley, B. B. Walden, H. Singh, R. Bachmayer, Surveying a subsea lava flow using the Autonomous Benthic Explorer (ABE). *Int. J. Syst. Sci.* **29**, 1031-1044 (1998).
67. C. L. Kaiser *et al.*, The design and 200 day per year operation of the autonomous underwater vehicle Sentry. *2016 IEEE/OES Autonomous Underwater Vehicles.*, 251-260 (2016).
68. L. L. Whitcomb *et al.*, Navigation and control of the Nereus hybrid underwater vehicle for global ocean science to 10,903 m depth: preliminary results. *2010 IEEE International Conference on Robotics and Automation.*, 594-600 (2010).
69. NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group, Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua chlorophyll data; 2019 reprocessing. https://oceancolor.gsfc.nasa.gov/l3/, accessed on (1 May, 2020).
70. H. Craig, K. K. Turekian, The GEOSECS Program: 1973–1976. *Earth Planet. Sci. Lett.* **32**, 217-219 (1976).
71. K. A. Fanning, Nutrient provinces in the sea: Concentration ratios, reaction rate ratios, and ideal covariation. *J. Geophys. Res. Oceans.* **97**, 5693-5712 (1992).
72. M. J. McPhaden *et al.*, The Tropical Ocean-Global Atmosphere observing system: A decade of progress. *J. Geophys. Res. Oceans.* **103**, 14169-14240 (1998).
73. L. D. Talley *et al.*, Changes in ocean heat, carbon content, and ventilation: A review of the first decade of GO-SHIP global repeat hydrography. *Ann. Rev. Mar. Sci.* **8**, 185-215 (2016).
74. E. Mawji *et al.*, The GEOTRACES Intermediate Data Product 2014. *Mar. Chem.* **177**, 1-8 (2015).
75. W. J. Jenkins *et al.*, An intermediate-depth source of hydrothermal 3He and dissolved iron in the North Pacific. *Earth Planet. Sci. Lett.* **539**, 116223 (2020).
76. C. Measures, M. Hatta, J. Fitzsimmons, P. Morton, Dissolved Al in the zonal N Atlantic section of the US GEOTRACES 2010/2011 cruises and the importance of hydrothermal inputs. *Deep Sea Res. Part II Oceanogr. Res. Pap.* **116**, 176-186 (2015).
77. J. A. Breier *et al.*, Sulfur, sulfides, oxides and organic matter aggregated in submarine hydrothermal plumes at 9°50′N East Pacific Rise. *Geochimica Et Cosmochimica Acta.* **88**, 216-236 (2012).
78. A. McCarthy, E. Chiang, M. L. Schmidt, V. J. Denef, RNA preservation agents and nucleic acid extraction method bias perceived bacterial community composition. *PLoS ONE.* **10**, e0121659 (2015).
79. J. G. Caporaso *et al.*, Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *Isme J.* **6**, 1621-1624 (2012).

80. J. J. Kozich, S. L. Westcott, N. T. Baxter, S. K. Highlander, P. D. Schloss, Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbio.* **79**, 5112-5120 (2013).
81. A. M. Seekatz *et al.*, Fecal microbiota transplantation eliminates clostridium difficile in a murine model of relapsing disease. *Infect. Immun.* **83**, 3838-3846 (2015).
82. A. M. Eren *et al.*, Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* **4**, 1111-1119 (2013).
83. A. M. Eren *et al.*, Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *Isme J.* **9**, 968-979 (2014).
84. S. M. Huse *et al.*, Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet.* **4**(2008).
85. P. D. Schloss *et al.*, Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537-7541 (2009).
86. C. S. Hughes *et al.*, Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757-n/a (2014).
87. A. Bankevich, P. A. Pevzner, TruSPAdes: barcode assembly of TruSeq synthetic long reads. *Nat. Methods.* **13**, 248-250 (2016).
88. A. Bankevich *et al.*, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455-477 (2012).
89. D. Antipov, A. Korobeynikov, J. S. McLean, P. A. Pevzner, hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics (Oxford, England).* **32**, 1009-1015 (2016).
90. A. D. Prjibelski *et al.*, ExSPAnder: a universal repeat resolver for DNA fragment assembly. *Bioinformatics (Oxford, England).* **30**, i293-i301 (2014).
91. M. Rho, H. Tang, Y. Ye, FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* **38**, e191 (2010).
92. C. L. Dupont *et al.*, Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *Isme J.* **9**, 1076-1092 (2015).
93. L. I. Gordon, J. C. Jennings Jr, A. A. Ross, J. M. Krest., *"A suggested protocol for continuous flow automated analysis of seawater nutrients (phosphate, nitrate, nitrite and silicic acid) in the WOCE Hydrographic Program and the Joint Global Ocean Fluxes Study, Methods Manual WHPO 91-1,"* (WOCE Hydrographic Program Office, , 1993).
94. F. Armstrong, C. R. Stearns, J. Strickland, The measurement of upwelling and subsequent biological process by means of the Technicon Autoanalyzer® and associated equipment. *Deep Sea Res.* **14**, 381-389 (1967).
95. K. P. Koltermann, V. Goretski, K. Jancke, Eds., *Volume 3: Atlantic Ocean* (International WOCE Project Office, Southampton, UK, 2011).

Acknowledgments

We thank the crews of R/V *Neil Armstrong* and R/V *Atlantic Explorer* for assistance and support with the ocean deployments. We thank the NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group for the Moderate-resolution Imaging Spectroradiometer (MODIS) Aqua Chlorophyll Data, NASA OB.DAAC, Greenbelt, MD, USA, https://oceancolor.gsfc.nasa.gov/l3/, accessed on 1 May 2020, and the ESA and developers of the SNAP tool used to illustrate this data. **Funding:** The development of *Clio* was supported by US National Science Foundation grants OCE-1333212 to Breier, Jakuba, Saito, and OCE-1334727 to Dick. The Sargasso Sea study was supported by US National Science Foundation grants OCE-1658030 to Saito and Jakuba, OCE-1657885 to Johnson, and OCE-1658067 to Breier. The Gordon and Betty Moore Foundation grant GBMF3782 provided support to Saito. The NOAA Educational Partnership Program with Minority-Serving Institutions Cooperative Agreement Award #NA16SEC4810009 provided support to Breier and Alanis; the contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the U.S. Department of Commerce, National Oceanic and Atmospheric Administration. **Author contributions:** J.B. designed the sampling systems. M.J. designed the vehicle. M.S., M.M., D.M., A.A., C.D. developed the proteomic methodology and associated metagenomic libraries and performed the analyses. G.D. and S.G. developed the DNA/RNA methodology and performed the analyses. R.J., M.J., M.S., M.M., S.G., E.C., B.A., and J.B. developed the field oceanographic methods and performed the fieldwork. J.B. wrote the manuscript with contributions from all. **Competing interests:** The authors declare they have no competing interests. **Data and materials availability:** Hydrographic, nutrient, and protein datasets and associated metadata are provided as supplemental tables and also have been submitted to the Biological and Chemical Oceanography Data Management Office (www.bco-dmo.org). Raw mass spectrometry and DNA sequence datasets have been deposited at the EBI PRIDE and NCBI repositories, respectively.

FIGURES

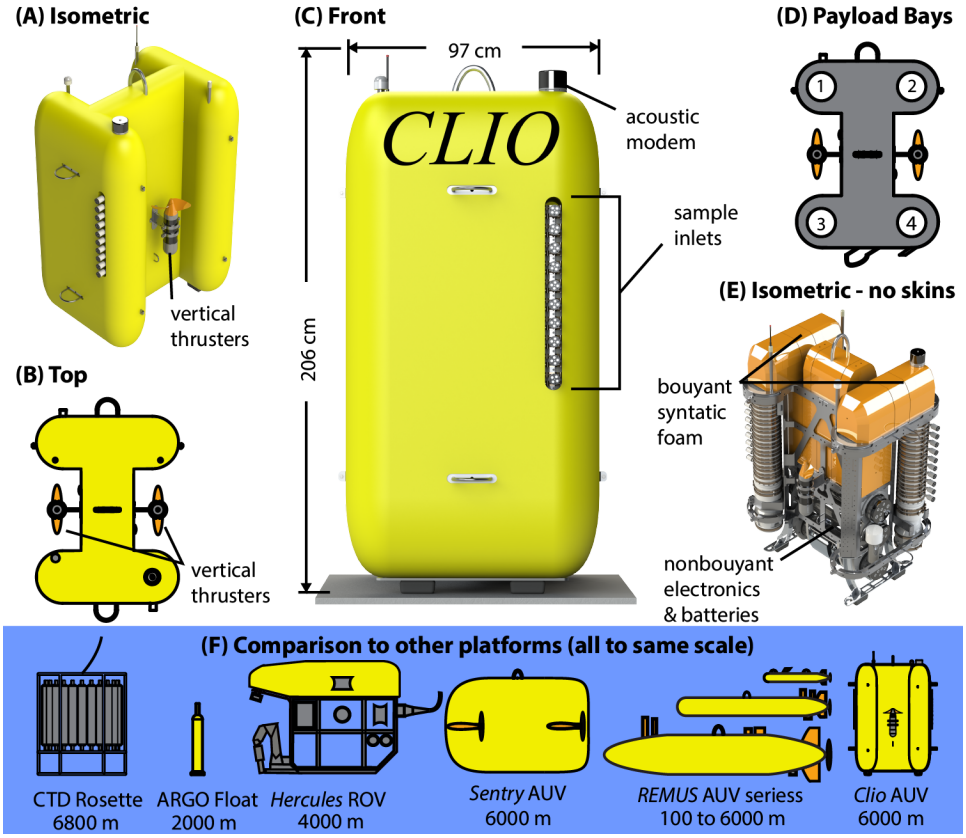

Fig. 1. *Clio* vertical AUV for biochemical ocean profiling. *Clio* is designed to vertically profile the ocean water-column to 6,000 m and drift horizontally in a Lagrangian manner. The (**A**) unique H shape of its hull protects the (**B**) thrusters and creates payload bays for sensors and sample-return systems, the inlets of which extend out of the (**C**) side of the vehicle. The (**D**) four payload bays are arranged at the corners of the H shape. The (**E**) isometric view with the vehicle skins removed shows the general arrangement of floatation in the upper part of the vehicle with the heaviest components near the bottom; this view also shows the two sample-return systems occupying payload bays 1 and 4 in the current configuration. *Clio* inherits component-level technology developed for other ocean robots and is (**F**) comparable in size to other deep-diving vehicles and sampling platforms, but its design and purpose are unique and directed at facilitating the goal of mapping global ocean biochemistry.

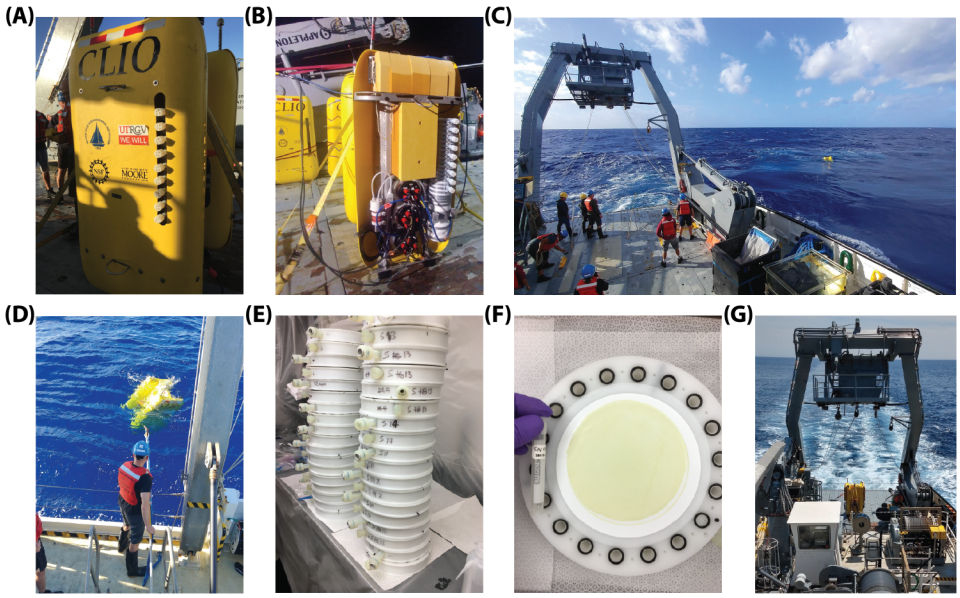

Fig. 2. Ocean biochemical profiling with *Clio*. (**A**) *Clio* is an autonomous robot designed for vertically profiling the ocean water-column to 6,000 m and collecting biochemically relevant data and samples. (**B**) *Clio* onboard R/V *Atlantic Explorer*, with its front and back skins removed to access the sampling bays for dive preparation prior to launch. (**C**) *Clio* at the surface after a successful sampling dive, waiting for recovery by R/V *Atlantic Explorer*. (**D**) *Clio* being hooked for recovery prior to being brought back on board by crane. (**E**) One dive's worth of 18 samples from the Sargasso Sea returned to a shipboard clean room for post-dive subsampling and processing. (**F**) One of these particulate samples filtered by *Clio* at 70 m depth from 47.7 L of oligotrophic seawater being prepared for splitting and storage. (**G**) *Clio* secured to the aft deck of R/V *Atlantic Explorer* being transported to the next dive on this 1,144 km transect through the upper water-column of the Sargasso Sea.

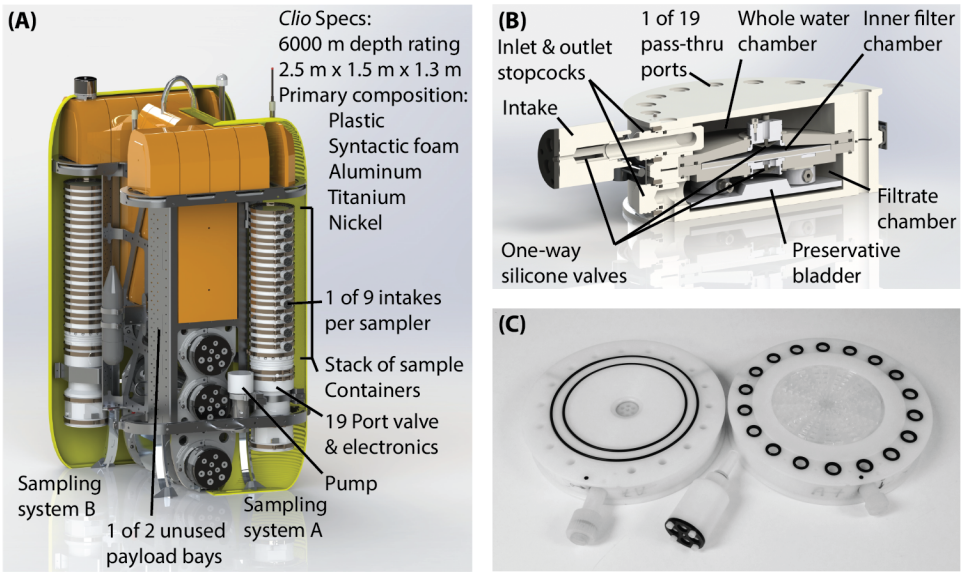

Fig. 3. *Clio* sampling system. (**A**) *Clio* is shown here with most of its outer skins removed to reveal three of its four payload bays; one of which is empty. The two used bays in this configuration are identical large volume sample collection systems as they were configured in these applications. The other two bays are latent payload space that has yet to be used. Additional sensors including fluorometers, transmissometers, oxygen sensors, and sensors for conductivity, temperature, and depth can be mounted external to the skins or internal to the skins when using a pump. A wide variety of filter holders or sample collection devices can be integrated with the *Clio* collection systems, but to maximize sample return when using larger filters (i.e. 142 mm diameter filter membranes) we designed (**B**) custom sample containers that combine the collection of 200 mL of whole water, 200 mL of filtrate, and particulate material collected by filtration all into a single compact unit. These individual sample units contain one-way silicone valves that allow for back injection of preservative, contained in a bladder, into the filter chamber immediately after collection to preserve RNA, for example, until (**C**) shipboard sample extraction can be performed.

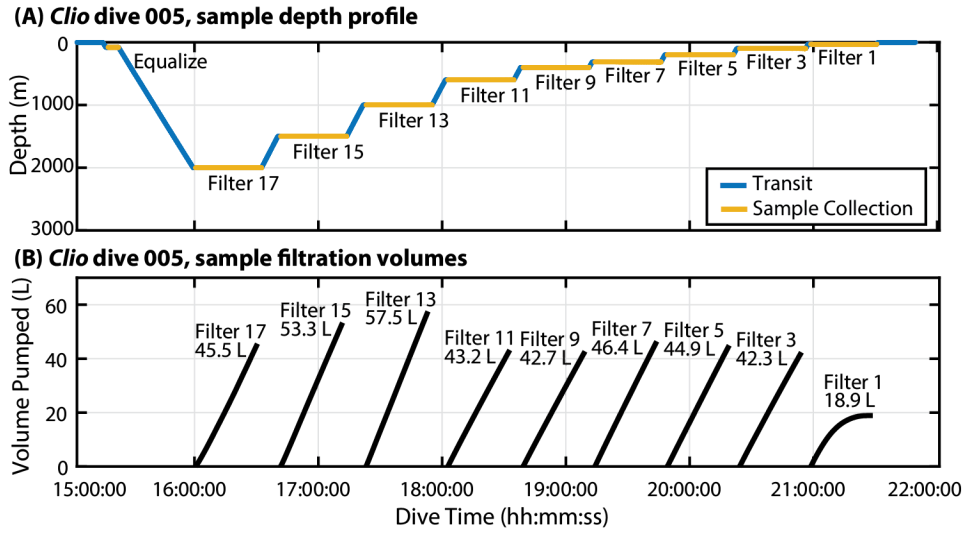

Fig. 4. Typical *Clio* dive profile. (**A**) In applications to date, a typical *Clio* sampling dive involves the vehicle briefly holding a shallow depth to ensure all air is purged, then driving down to its deepest sampling depth, after which it holds depth at each target location while collecting samples, before advancing to the next depth as it returns to the surface. (**B**) At each sample depth, a sample is collected by filtration for a predetermined maximum amount of time and the filtered volume is tracked in real-time and digitally recorded. Samples collected in the surface ocean (e.g. Filter 1) collect biomass more rapidly due to the higher particulate concentration, which is reflected in the nonlinearity of the sample flow curve, and therefore these samples require less filtration. Dives such as these to 2,000 m in the oligotrophic ocean typically take as little as seven hours to complete, though frequently *Clio* will be tasked to hold in place in order to return to the surface at a later time convenient for recovery by the research vessel.

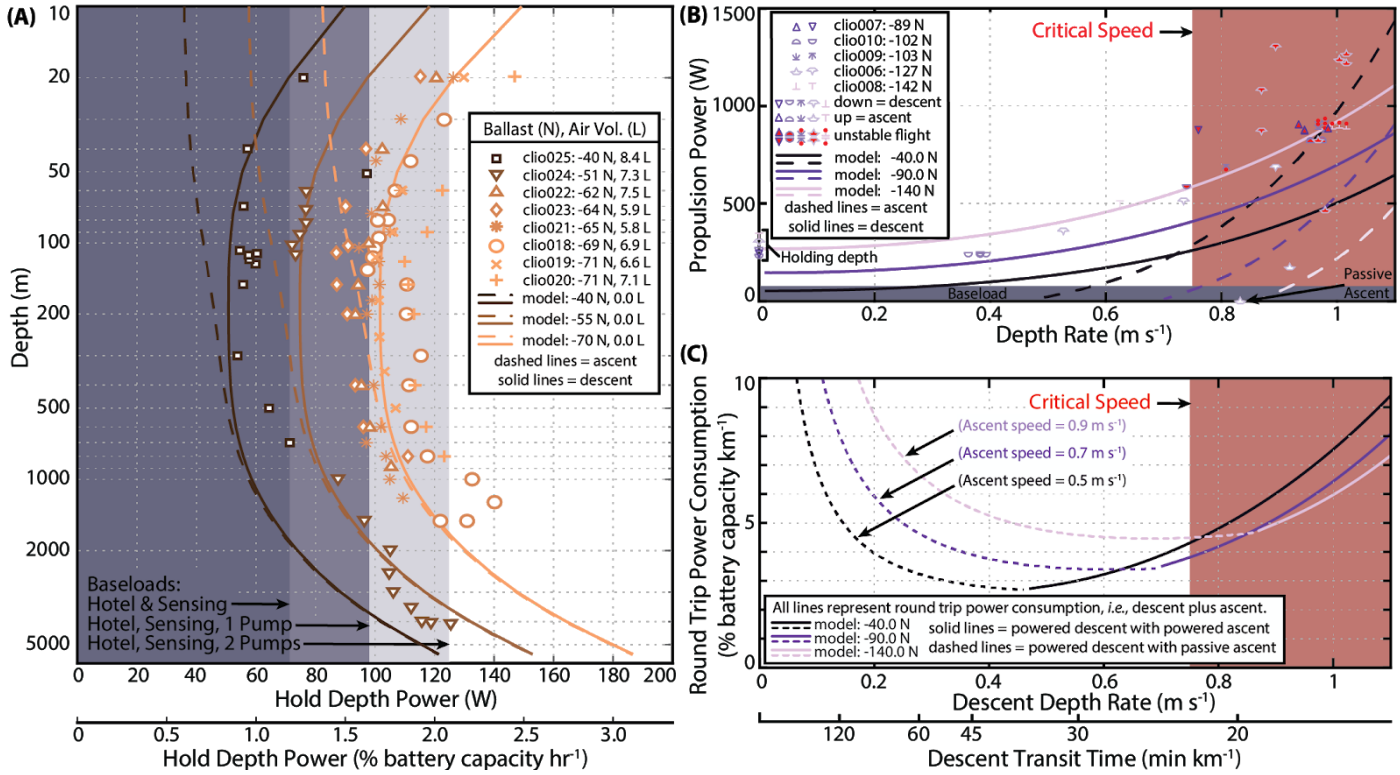

Fig. 5. Power consumption while holding depth and transiting. Power consumption while holding depth (**A**) consists of baseload, the power to run pumps, and propulsion power to overcome vehicle buoyancy (including air bladders). Vehicle buoyancy varies with depth for any given background water column profile. The solid lines are predictions for propulsion power consumption assuming a global mean temperature and density profile*(56)*, whereas the markers represent data for *in situ* profiles measured by *Clio*. Propulsion power during transits (**B**) depends on speed and vehicle buoyancy, cumulative with baseload. Above the critical speed at which *Clio* transitions to unstable flight, the effective drag rises markedly resulting in a concomitant increase in propulsion power. The model predictions apply for stable flight only (sometimes observed above the critical speed). (**C**) The propulsion energy consumed to execute a mid-water round-trip descent plus ascent for a 1 km profile. The dashed portions of the model curves indicate power that would be expended during passive ascent to slow the vehicle; and therefore reflect energy consumed by base load only during a passive ascent at the terminal speed indicated. The model curves in (**A**,**C**) allow accurate prediction of total energy consumption for mission planning.

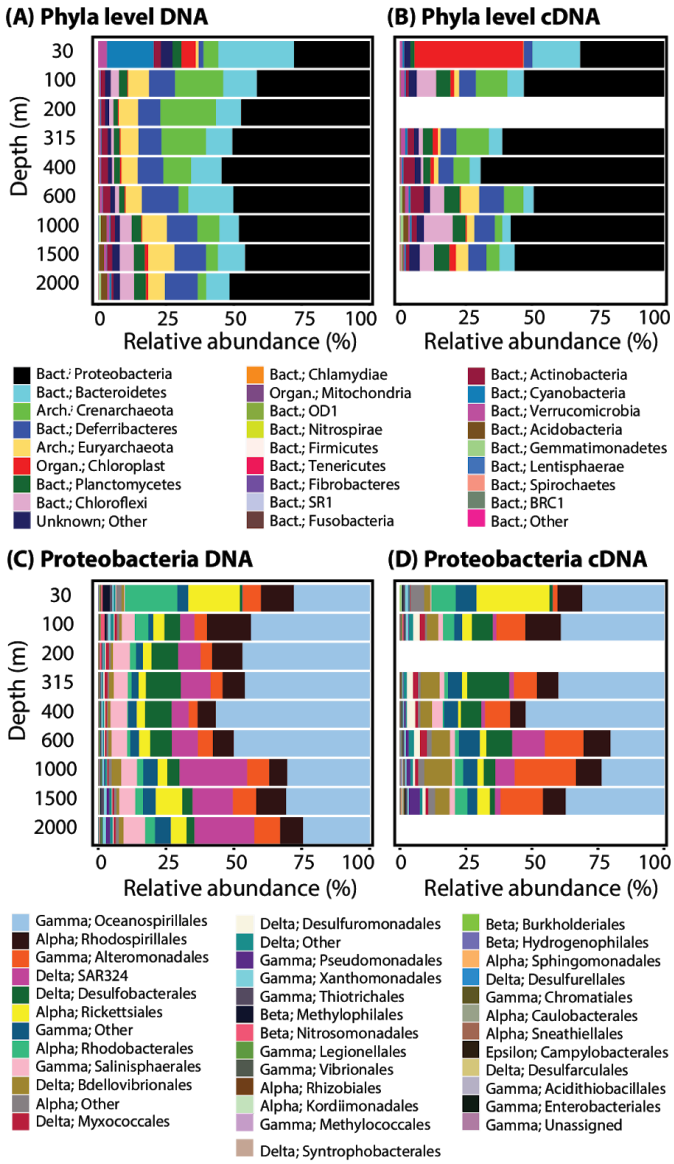

Fig. 6. Composition of *Clio* DNA and cDNA samples.
The relative abundances of phyla in *Clio* (**A**) DNA and (**B**) cDNA samples; and the relative abundance within proteobacteria in *Clio* (**C**) DNA and (**D**) cDNA samples. Samples are organized by depth on the y-axis, and relative abundance (out of 100%) on the x-axis. Two cDNA samples, AR20-5-D5 at 200m and AR20-5-D17 at 2,000m, do not show on the graph due to either or both insufficient biomass and extraction methods resulting in low cDNA yields.

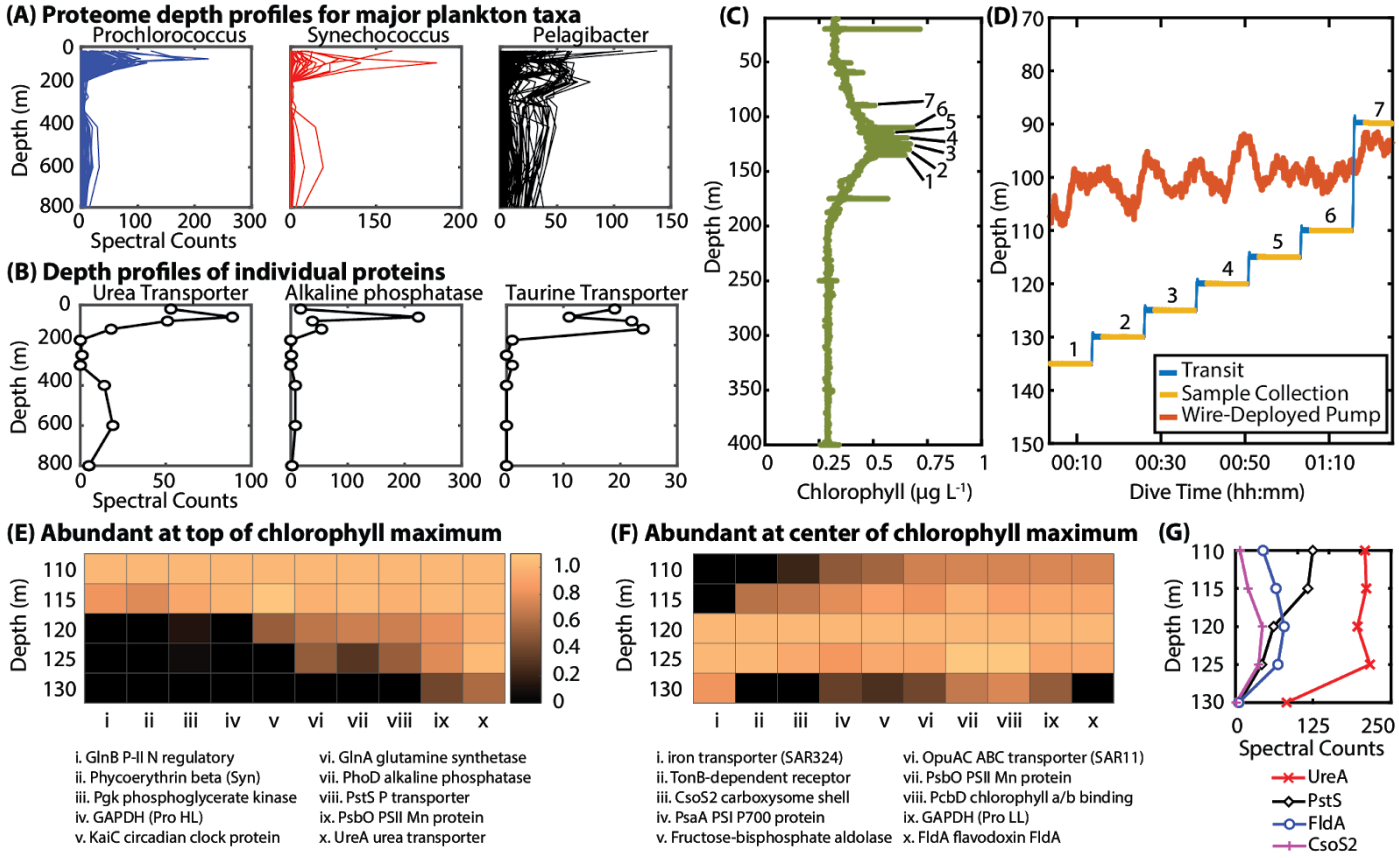

Fig. 7. *Clio* high-resolution vertical sampling. Profiles of (**A**) proteins categorized by major microbial taxa and (**B**) selected nutrient response proteins from *Prochlorococcus* (urea transporter and alkaline phosphatase) and *Pelagibacter* (taurine transporter) from dives Clio-007/008 from BATS station. (**C-G**) Higher-resolution sampling of the deep chlorophyll maximum on dive Clio-014. (**C**) *Clio*'s chlorophyll profile; sample locations numbered. (**D**) *Clio* depth versus time, with the same locations labeled, shows holding station to <1 m, while depth versus time for a concurrent wire-deployed pump shows the limitation ship heave imposes on depth precision for cabled instruments. Significant differences in proteins (Fisher's Exact Test $p<0.05$) were observed in (**E**) 10 selected proteins more abundant at the top (110 m) or (**F**) center (120 m) of the chlorophyll maximum (ratio of total exclusive spectral counts, normalized and log-transformed; see Supplemental Tables 5-8) All proteins are *Prochlorococcus* unless labeled otherwise (Syn is *Synechococcus*, Pro is *Prochlorococcus*; HL and LL refer to high- and low-light ecotypes). (**G**) Profiles of selected proteins: phosphate transporter PstS, urea transporter UreA, flavodoxin FldA (iron-free substitute for ferredoxin), and CsoS2 carboxysome shell protein demonstrate observation of biologically meaningful differences in P, N, Fe, and C metabolisms with 5 m resolution.

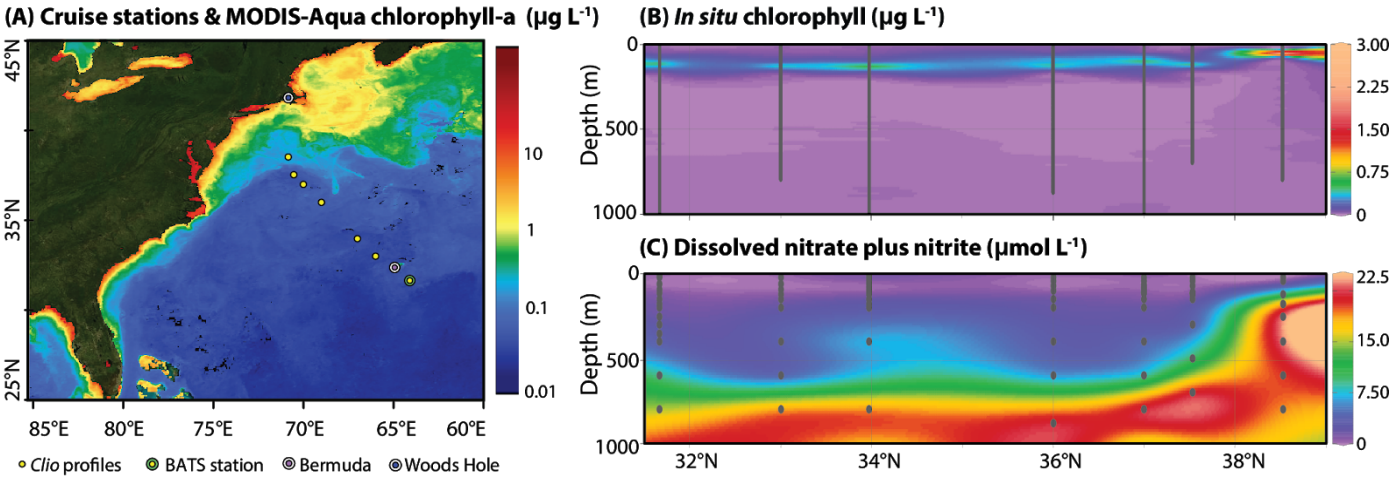

Fig. 8. Ocean biogeochemical structure revealed using *Clio*. (**A**) The first *Clio* sectional cruise stations superimposed over the level 3, monthly averaged chlorophyll-a estimates from the MODIS-Aqua spectroradiometer for June 2019. (**B**) *In situ* chlorophyll-a measured by *Clio*'s fluorometer: grey dots indicate the location of point data that is spatially interpolated, and (**C**) dissolved nitrate plus nitrite measured from samples collected by *Clio*.

SUPPLEMENTAL INFORMATION:

BACKGROUND:

Global ocean studies

The Geochemical Ocean Sections Study (GEOSECS, field phase: 1972-1978), which included the Atlantic, Pacific, and Indian Oceans, was designed to improve our understanding of ocean mixing and circulation*(70)*. GEOSECS was the first modern global mapping of ocean chemical properties and from it came many of the methods used in subsequent studies as well as a broad understanding of chemical distributions in the ocean and their relationships with the global thermohaline circulation system*(21, 71)*. Following GEOSECS, was the Tropical Ocean Global Atmosphere program (TOGA: field phase 1985-1994), which focused on the study of coupled ocean and atmosphere interactions in the tropical Pacific. From TOGA, we gained the knowledge to better predict El Niño and Southern Oscillation events*(72)*. Then came the World Ocean Circulation Experiment (WOCE, field phase: 1990-1998), which was designed to develop and test global ocean models that could better inform climate studies*(20, 26)*, and contemporaneously, Joint Global Ocean Flux Study (JGOFS, field phase: 1989-1998), which is the modern basis of our understanding of the ocean carbon cycle*(22)*. More recently, the Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP: 2009 to present), seeks to understand how ocean properties vary, particularly with respect to climate change, by systematically repeating ocean basin scale observations*(73)*, and, the GEOTRACES program (2010 to present) seeks to improve our understanding of global ocean biogeochemical cycles through the study of trace elements and their isotopes*(74-76)*.

Sectional cruise methodology

Sectional studies to date have relied on the collection of samples by equipment deployed by wired cables from surface ships. Water samples are collected by Niskin and GO-Flo bottles on water carousels and suspended particulates are collected by wire-deployed large volume in situ stand-alone pumps or by ship-board filtration of water samples (e.g., Fig. S1B). GEOTRACES cruises are good examples of the modern wire-based approach to creating an ocean sectional map. All sample types are collected as vertical profiles at stations spaced along the cruise track. The time required to collect these samples is determined by the combination of the wire-speed for the winch (typically 0.5 to 0.7 m s-1), cast depths, pumping time, processing time, and the number of sampling casts, which must be sequential to avoid entangling wires.

The U.S. GEOTRACES North Atlantic Leg 2 sectional cruise is a useful example of the wire-based field techniques used for global ocean studies and involves methodology put into practice during GEOSECS. Sectional cruises vary in length but the U.S. GEOTRACES North Atlantic Leg 2 section (Cape Verde to Bermuda to Woods Hole, USA), for example, had 24 stations. At each station multiple sampling casts are required, these include Super-station, Full-station, and Demi-station. They differ in the number and type of samples collected. Full- and Super- stations represent the core sampling plan: 14 of the North Atlantic section stations were Full- or Super- stations. Full-stations require 5 water sampling casts and 2 pumping casts and require ~27 hrs to complete. Super stations are Full stations plus 3 additional water sampling casts and are 30 to 39 hrs long. Demi-stations consist of just 2 shallow water sampling casts and are only ~3 hrs long. During Full- and Super- stations the two pumping casts collect suspended particulate samples at up to 16 water depths. During these pumping casts as many

as 8 stand-alone pumps are strung at fixed points along a wire, lowered to their target depths, filter water for a fixed preprogrammed amount of time, and are brought back to the surface and removed from the wire one at a time. Each pump is set to filter simultaneously onto a Whatman® QMA quartz fiber filter (GE Healthcare) for organic carbon analyses and a Supor® polyethersulfone membrane (Pall Corporation) filter for trace metal analyses. Typically 1100 L is filtered through the QMA and 500 L is filtered through the Supor® filter. The shallow pump cast takes ~5 hrs and the deep pump cast takes ~9 hrs.

ENGINEERING METHODS

Field measurements of thrust, power, and ballast conditions

Thrust was computed from logged thruster current while holding depth using a calibrated winding-current-to-thrust coefficient (bollard pull) determined from dock tests. Power consumption was recorded directly by the vehicle. Ballast and air bladder volume were determined for each dive via least-squares after first adjusting the average thrust at each station for the dependence of buoyancy on pressure, temperature and background density using the model for vehicle displacement as a function of temperature and pressure fit to data from *Clio* dive number 024 to 4,100 m. Thrust during transits was determined by fitting models to data from a series of dives (clio006-010) that included ascents and descents at various speeds. Data from each ascent or descent were classified as corresponding to stable or unstable flight depending on the magnitude of pitch and roll oscillations observed (<5° or >20°, respectively). Only data from stable descents was used to fit models. Three physics-based models were fit to the data and then solved to arrive at a prediction for power as a function of depth-rate. First, logged depth-rate and thruster rotational speed, along with ballast determined as above, were used to fit a thrust model parameterized by a quadratic drag coefficient and linear thrust coefficient (with respect to advance ratio in the first quadrant). Second, logged depth-rate and thruster current and rotational speed were used to fit a torque model parameterized by a lumped linear torque coefficient (with respect to advance ratio in the first quadrant and lumped with a motor torque constant). Third, logged thruster current and rotational speed and total power consumption with known baseload were used to fit a second-order power consumption model parameterized by mechanical efficiency, motor winding resistance, and known baseload. For a given ballast and depth rate model (i) predicts thruster rotational speed, which feeds into (ii) to predict thruster current, which feeds into (iii) to predict total power. There was insufficient data to permit the same approach for the ascent. Instead, first and third quadrant thruster and propeller performance data from towing tank tests were used to scale the descent thrust and torque coefficients proportionally.

BIOCHEMICAL METHODS

Sample system configuration

Each of these sample-return systems was configured to collect nine sample-sets each consisting of (i) particulate material filtered onto 142 mm diameter filter media (ii) 200 mL of filtrate, and (iii) 200 mL of unfiltered water. During the 2017 sea trials, both of these two systems were loaded with 0.2 µm pore size Supor® polyethersulfone (PES) filter membranes (Pall Corporation), and one was configured to preserve samples in situ with RNALater for subsequent

RNA analysis. RNA analyses were not part of the subsequent studies and preservative was not used. During the semi-quarterly deployments at BATS station, both of these two systems were loaded with 0.2 μm pore size PES filter membranes for both proteomic analyses and trace metal analyses. During the sectional cruise through the Sargasso Sea, one of the two systems was loaded with 0.2 μm pore size filter media (PES) for both proteomic analyses and trace metal analyses, and the other was loaded with Whatman® QMA quartz fiber filter media (GE Healthcare) for organic carbon analyses.

Sample system purging

The *Clio* sampling systems are designed to operate at ambient pressures. A choice then has to be made about what fills the fluidic spaces prior to the AUV launch. From a purely operational perspective, the most expedient choice is air. From a cleanliness choice, air is not the best choice. If air initially fills the fluid space then seawater will be forced into the fluidic system as the vehicle dives and air is compressed. This would contaminate all the samples with surface seawater. Instead, we prefill the fluidic systems with ultrapure analytic grade deionized water prior to launch as part of the method of operation developed for previous sample-return systems*(42)*. This water is displaced while sampling, and its volume is flushed many times with seawater being filtered. As the fluidic path contains filters, multiple chambers, passive seals, and torturous lengths that pass through the valve, current methods for prefilling with DI may still leave some residual air bubble traps in the system, to further control how seawater fills these small remaining air voids, the AUV spends 4 minutes holding depth at 50 m, while the valves are cycled, to briefly connect to each sample line and allow seawater to backfill any remaining trapped air by flowing through the pump, the pump is not actuated during this process. In this way, we guide to some degree the depth where residual trapped air is filled with seawater and choose a consistent depth that minimizes this contamination pathway.

Sample handling

All sampling system components were thoroughly cleaned prior to use by a sequence of 5% ethanol and trace-metal clean 5% hydrochloric (Seastar) acid baths*(77)*. All plasticware was rinsed with ultrapure water and leached with 10% hydrochloric (Seastar) acid at pH 2, and sample storage bottles were rinsed with the sample prior to sample collection. All pre- and post-deployment filter handling and water subsampling were performed under laminar flow within a clean environment with positive pressure constructed shipboard from plastic sheeting. Acid-cleaned (10% hydrochloric, Seastar), 142 mm diameter, 0.2 μm poresize Supor® polyethersulfone membranes (Pall Corporation) were used for DNA and RNA, and proteomics measurements. The filters for DNA/RNA analysis were preserved in situ with RNAlater (Ambion Life Technologies). In addition, to limit protein collection from multicellular organisms, 51 μm acid-cleaned (10% hydrochloric, Seastar) nylon mesh prefilters were added before each of the 142 mm 0.2 μm poresize filters during the Bermuda Atlantic Time-series Study (BATS) station deployments (AE1810, AE1817, AE1829, and AE1906) and the 2019 Sargasso Sea sectional cruise (AE1913). The 142 mm 0.2 μm poresize filters from *Clio* were split shipboard and stored at -80 C, with RNAlater (Ambion Life Technologies), for subsequent analyses. For the *Clio* sea trial cruise (AR20), 1/16 of the filter was used for DNA/RNA analysis. For the BATS station deployments (AE1811 and AE1829) and the 2019 Sargasso Sea sectional cruise (AE1913), 1/2 of the 0.2 μm filter was used for proteomics analysis. During the 2019 Sargasso Sea sectional cruise, aliquots of filtrate collected by *Clio* were also subsampled shipboard for shore-based nutrient analysis; these aliquots were stored in acid-cleaned high-density polyethylene bottles and frozen until analysis.

DNA/RNA

DNA and RNA were co-extracted using a Qiagen AllPrep RNA/DNA/miRNA kit. Extracted RNA was reverse transcribed into ss cDNA, and submitted alongside DNA to the University of Michigan Microbiome sequencing core for 16S rRNA gene V4 sequencing on the Illumina MiSeq with a 2x250 cycle kit. They were sequenced and multiplexed ~380 on the entire run.

DNA and RNA from samples were extracted using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen) as described in McCarthy *et al*.*(78)* with modifications. Unless otherwise specified, all centrifugations were conducted at 20,000 x g for 15 seconds at room temperature. Briefly, 1/16 of a 142 mm filter, folded cell side in, was rinsed with 1X phosphate-buffered saline (pH 7.4) to remove residual RNAlater. The rinsed filter was placed into a 2 mL tube with 6 µL of 2-mercaptoethanol and 600 µL RLT lysis buffer and incubated on a vortex shaker (speed 5) at room temperature for 90 minutes. After incubation, the tube was vortexed at maximum speed for 10 minutes. Lysate was transferred to a QiaShredder column (Qiagen) and centrifuged. Homogenized lysate from the column was transferred to the AllPrep DNA Mini Spin column and centrifuged to capture DNA on the membrane. The column was transferred to a new collection tube and stored on ice while RNA was processed from the flow-through of the QiaShredder column. The remainder of the protocol followed manufacturer's instructions, with two elution steps of 30 µL each for RNA, and two elutions of 50 µL for DNA that were combined. RNA elutions were flash-frozen on dry ice and stored at -80˚C until further processing. DNA samples were stored at 4˚C for short-term handling, or -20˚C for long-term storage.

From the first elution, RNA and DNA concentrations were assayed using Ribogreen and Picogreen (Invitrogen), and cDNA was synthesized from RNA using High Capacity Reverse Transcription Kit (Applied Biosystems) as described in the manual with modifications. Briefly, in a 0.2 mL nuclease-free tube, the following components were added: 3.2 µL of nuclease-free water, 2.0 µL 10X RT Buffer, 2.0 µL 10X random primers, 0.8 µL 25X dNTP mix, 1.0 µL reverse transcriptase, 1.0 µL RNase inhibitor, and 5-10 µL of RNA (corresponding to 0.49-14 ng of RNA). cDNA synthesis thermocycling was 25˚C for 10 minutes, 37˚C for 120 minutes, and 85˚C for 5 minutes. Samples were stored up to 24 hours at 4˚C before verifying cDNA synthesis with PCR for the bacterial 16S v4 region (515F-806R).

Extracted DNA and synthesized cDNA were submitted to the University of Michigan Center for Microbial Systems for bacterial and archaeal 16S v4 sequencing*(79-81)* using the Illumina MiSeq 2x250 (500) cycle kit. Raw reads (FastQ files) were submitted to NCBI (project: PRJNA607796) and processed using illumina-utils*(82)* (https://github.com/merenlab/illumina-utils).

Sequence data was analyzed using Minimum Entropy Decomposition (MED)*(83)*. MED analysis was conducted with version 2.1 of MED pipeline using the following parameters: (i) 1,939,015 merged high-quality reads from the V4 region with 5 or fewer mismatches, representing 76 samples; (ii) maximum of 4 discriminants for decomposition; (iii) the abundance of the most abundant unique read in a node was 5 reads; (iv) attempt to relocate reads from failed nodes to final topology; (v) a maximum variation of 3 nucleotides, or 1% of average sequence length, allowed in each node; (vi) the minimum entropy parameter for a component to be picked for decomposition was 0.0965. After resolving nodes, a representative sequence from each node was queried against a curated SILVA119 database using Global Alignment for Sequence

Taxonomy GAST*(84)*; this database contains 19,258 archaeal sequences, 50,197 eukarya sequences, and 386,827 bacterial sequences. The presence of chimeras was evaluated using the sample sequences as a reference, via uchime in Mothur v.1.36.1*(85)*. The representative sequence was queried from all putative chimeric nodes against SILVA123 using BLAST, and only those reads that did not have taxonomy assigned via both GAST and BLAST were considered chimeric and removed from the study. Sequences were also scanned for putative contaminant taxa using the list provided in Salter *et al*.*(60)* and the interpretation from Sheik *et al*.*(61)* of low-biomass sequenced samples from the Census of Deep Life; nodes that match in taxonomy to these were flagged as putative contaminants.

Proteomics

Protein sampling: Samples for a vertical profile at the BATS station were collected on dives Clio-007 (depths 20, 60, 80, 120, 175, 250, 400, 600, and 800m), Clio-008 (300m), and Clio-014 (110, 115, 120, 125, and 130m). Pump volumes for dives Clio-007 and Clio-008 are shown in Table S3. Pump volumes for dive Clio-014 are shown in Table S5.

Protein extraction: Proteins were extracted from half sections of 142 mm 0.2 µm Supor® filters (Pall Corporation) using a modified magnetic bead method *(86)*. Filter sections were placed in 5 mL of protein extraction buffer: 50 mM HEPES pH 8.5 (Boston BioProducts), 1% SDS in HPLC grade water. All reagents in this protocol are made with HPLC grade water. Samples were heated at 95 ˚C for 10 minutes and shaken at room temperature for 30 minutes. Supor® filters were removed and protein extracts were filtered through 5.0 µm Millex low protein binding filters (Merck Millipore). Millex filters were rinsed with 1 mL of extraction buffer to ensure no loss of protein. Samples were then spun for 30 minutes at 3,220 rcf in an Eppendorf 5810 centrifuge. Supernatant was removed and transferred to a Vivaspin 5K MWCO ultrafiltration unit (Sartorius Stedim). Protein extract was concentrated to approximately 350 µL, washed with 1 mL of lysis buffer, and transferred to a 2 mL ETOH washed microtube (all tubes from this point on are ETOH washed). Vivaspins were rinsed with small volumes of protein extraction buffer to remove all concentrated protein and all samples were brought up to 430 µL. 30 µl was set aside for total protein quantification and DNA analysis.

Protein quantification: Standard curves were generated using albumin standard (Thermo Scientific). Total protein was quantified after extraction and after purification with 2 µL of sample in duplicate using the BCA method (Thermo Scientific Micro BCA Protein Assay Kit #23235). Absorbance was measured on a Nanodrop ND-1000 spectrophotometer (Thermo Scientific). Protein reduction and alkylation. 50 units (2 µL) of benzonase nuclease (Novagen) was added to each sample and incubated at 37 ˚C for 30 minutes. Samples were reduced by adding 20 µL of 200 mM DTT (Fisher Scientific) in 50 mM HEPES pH 8.5 at 45 ˚C for 30 minutes. Samples were alkylated by adding 40 µL of 400 mM iodoacetamide (Acros) in HEPES pH 8.5 for 30 minutes at 24˚C, occasionally heating to 37 ˚C to prevent precipitation. The reaction was quenched by adding 40 µL of 200 mM DTT in 50 mM HEPES pH 8.5.

Protein clean up and digestion. SpeedBead Magnetic Carboxylate Modified Particles (GE Healthcare #65152105050250 and #45152105050250) were prepared according to Hughes *et al. (86)*. 20 µL (20 µg µL-1) of magnetic beads were added to 400 µL of extracted protein sample. Samples were heated at 37 ˚C periodically to avoid precipitation. Samples were acidified to a pH of 2-3 by adding 50 µL of 10% formic acid. 2X volume (1,100 µL) of acetonitrile

was immediately added. Samples were incubated at 37 ˚C for 15 minutes and then at room temperature for 30 minutes. Samples were placed on a magnetic rack, incubated for 2 minutes, the supernatant was removed and discarded. Samples were washed 2 times removing and discarding supernatants with 1,400 µL of 70% ETOH for 30 seconds on the magnetic rack. 1,400 µL of acetonitrile was added to each sample for 30 seconds on the magnetic rack. Supernatant was removed and discarded. Samples were air-dried for approximately 4 minutes until acetonitrile had just evaporated. Samples were removed from the magnetic rack and beads were reconstituted in 90µL of 50mM HEPES pH 8.0. Purified protein was quantified as described above. Trypsin (Promega) dissolved in HEPES pH 8.0 at a concentration of 0.5 µg µL-1 was added to samples at a 1:25 trypsin to protein ratio and incubated at 37 ˚C overnight. Peptide recovery and preparation. Acetonitrile was added to digested peptides at a concentration of ≥ 95% and incubated for 20 minutes at room temperature. Samples were then placed on the magnetic rack for 2 minutes and supernatant was removed and discarded. 1,400 µL of acetonitrile was added to samples on the magnetic rack for 15 seconds. Supernatant was removed and discarded. Samples air-dried for approximately 4 minutes, just until acetonitrile was evaporated. Beads were reconstituted in 90 µL of 2% DMSO and incubated off of the rack at room temperature for ≥ 15 minutes. Samples were centrifuged slowly and briefly at 900 rcf to remove liquid from the tube walls. Samples were incubated on the magnetic rack for 15 minutes and supernatant containing peptides was transferred to a new ETOH washed 1.5 mL microtube. This step was repeated to ensure removal of all magnetic beads. 1% trifluoroacetic acid was added to samples for a final concentration of 0.1%. Samples were zip tipped with Pierce C18 tips (Fisher Scientific) according to manufacturer's protocol with a final resuspension in 25 µL of 70% acetonitrile, 0.1% formic acid. Samples were evaporated to approximately 10 µL in a DNA110 Speedvac (ThermoSavant). Samples with lower protein concentrations were further evaporated to minimize acetonitrile percentage in final resuspension – zip tip product to be less than 30% of total final buffer B volume. Samples were finally resuspended to a peptide concentration of 1µg µL-1 in buffer B (2% acetonitrile, 01% formic acid).

Global metaproteomics mass spectrometry: Purified peptides were diluted to 0.1 µg µL-1 and 20 µL (2 µg) was injected onto a Michrom Advance HPLC system with reverse phase chromatography coupled to a Thermo Scientific Q-Exactive Orbitrap mass spectrometer with a Michrom Advance CaptiveSpray source. Each sample was concentrated onto a trap column (0.2 x 10 mm ID, 5 µm particle size, 120 Å pore size, C18 Reprosil-Gold, Dr. Maisch GmbH) and rinsed with 100 µL 0.1% formic acid, 2% acetonitrile (ACN), 97.9% water before gradient elution through a reverse phase C18 column (0.1 x 150 mm ID, 3 µm particle size, 120 Å pore size, C18 Reprosil-Gold, Dr. Maisch GmbH) at a flow rate of 500 nL min-1. The chromatography consisted of a nonlinear 180 min gradient from 5% to 95% buffer B, where A was 0.1% formic acid in water and B was 0.1% formic acid in ACN (all solvents were Fisher Optima grade). The mass spectrometer monitored MS1 scans from 380 m/z to 1580 m/z at 70K resolution. MS2 scans were performed at 15K resolution on the top 10 ions with an isolation window of 2.0 m/z and a 10 second exclusion time. Peptide-to-spectrum mass spectra identifications were made using SEQUEST HT and a metagenome sequence database (see below) within Thermo Proteome Discoverer 2.1 software using a parent ion tolerance of 10 ppm and a fragment tolerance of 0.02 Da. Processed files were then loaded into Scaffold 4.8 (Proteome Software Inc.) for an FDR of less 1% on the protein level with 1 peptide minimum, and a custom peptide threshold (Sequest XCorr confidence thresholds of 1.2, 1.9, 2.3, and 2.6 for +1, +2, +3, and +4 respectively, and DeltaCn of at least 0.1). These settings resulted in a 0.7% false discovery rate (FDR) on the protein level and 0.26% FDR on the peptide level against a decoy reversed

database, equivalent to a protein probability of 99.0% and peptide probability of 95% for Clio-007/8 samples. Similarly, FDRs of 0.8% and 0.31% for proteins and peptides respectively were obtained, equivalent to a protein probability of 99.0% and peptide probability of 95% for Clio-014 samples.

Metagenomic database for metaproteomics: Metaproteomics peptide-to-spectrum identifications were made against a database created from combining metagenomic sequencing from two samples collected from the Bermuda Atlantic Time-Series Study station on February 23, 2012, at 80m (sample GOS-843) and August 8th, 2012 at 130m (sample GOS-883). Metagenomic extractions were conducted as described by Rusch *et al. (23)*; which is optimized to obtain high molecular weight DNA (40Kbp+) with minimal sheering. For metagenomic sequencing, mixed DNA/RNA was treated with RNAase and then prepared according to the manufacturer protocols for TruSeq Synthetic Long Read Sequencing (TSLR; Illumina, Inc., San Diego, CA, USA). The short reads from the TSLR sequencing were assembled into synthetic long reads using TruSPAdes*(87)* and the synthetic long reads were subsequently assembled into longer contigs using SPAdes*(88)* (with k-mer sizes of 33, 55, 95 and 127). The raw Illumina reads were not trimmed for adapters or low-quality sequences prior to assembly with TruSPAdes so as to retain as much information as possible for good quality barcode assembly. We further represented each TSLR as a path in the assembly graph, not unlike representing Pacific Biosciences or Oxford Nanopore reads using the hybridSPAdes framework*(89)*. We further applied exSPAnder*(90)* to these paths to construct long contigs from TSLRs. All code is available at http://bioinf.spbau.ru/spades. Open reading frames were called with FragGeneScan*(91)* and predicted proteins were annotated as in Dupont *et al. (92)*.

Dissolved inorganic nitrogen

Analyses for dissolved nitrate plus nitrite and ammonium were performed at Oregon State University using methods developed for the WOCE *(93)*. The method for dissolved nitrate plus nitrite consists of reducing nitrate to nitrite using a cadmium column, followed by reaction of nitrite with sulfanilmanide and N-(1-napthyl)ethylenediamine dihydrochloride, and quantifying the absorbance of the resulting colored compound with a spectrophotometer*(94)*. The method for dissolved ammonium consists of reacting ammonium with alkaline phenol and hypochlorite to form indophenol blue and quantifying the absorbance of the resulting colored compound with a spectrophotometer. A quality control evaluation of nutrient data was performed by comparing results against nutrient samples measured during World Ocean Circulation Experiment meridional cruise A23*(95)* that followed a similar cruise track, as a result, two of the 87 samples were identified as anonymously low given their location and depth and those samples are flagged as such in Supplementary Table 9 and excluded from Fig. 8C.

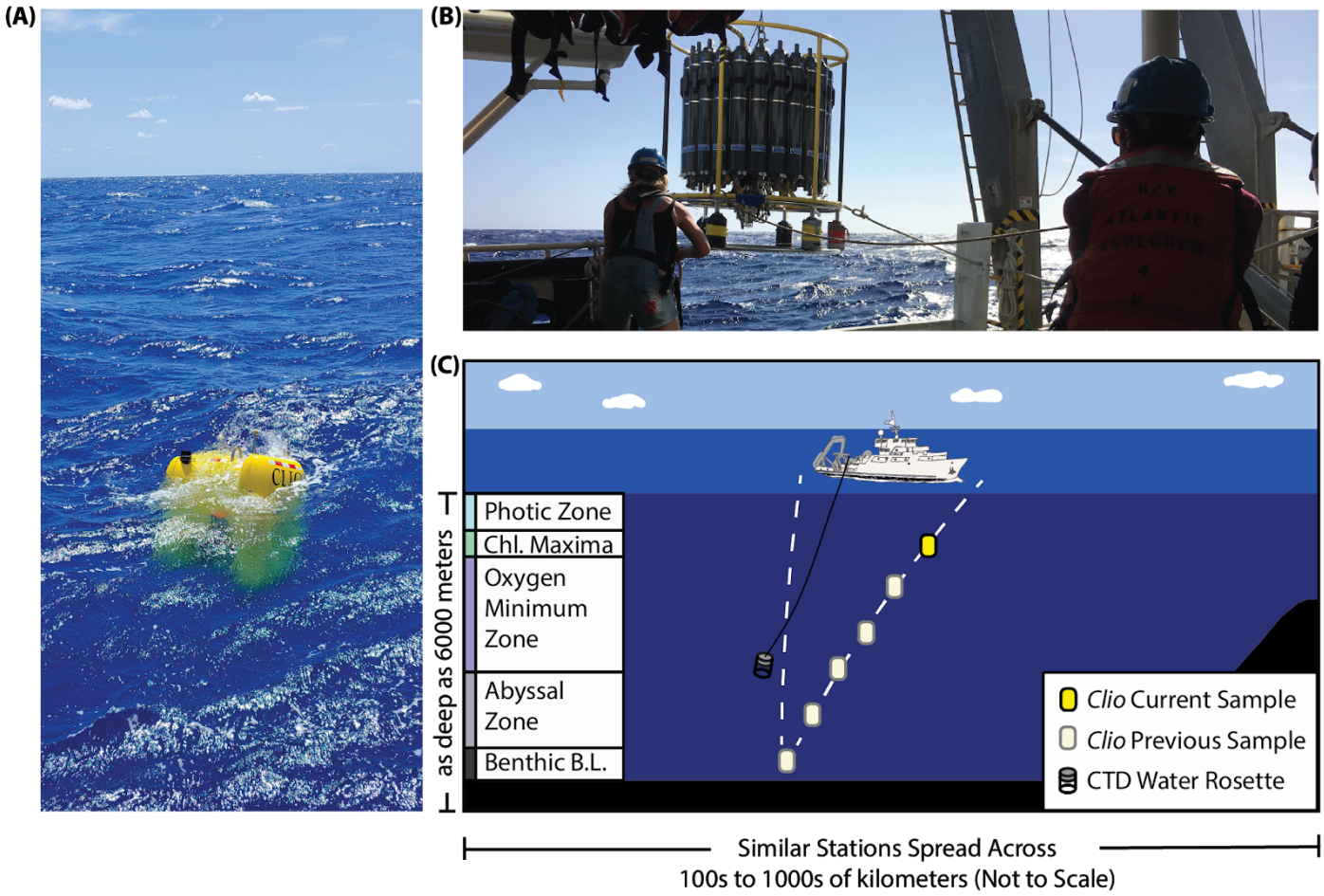

Fig. S1. *Clio* robotic approach to ocean biochemical profiling. (**a**) A key advance of *Clio* is that it is free to navigate itself along data collection profiles independently of its host ship. Comparable existing technology, used for full-depth ocean biochemical studies, including (**b**) Niskin bottle water rosettes, rely on a ship-based winched cable for positioning. Once launched *Clio* is free to operate independently of the host ship: diving to sample depths, drifting in a Lagrangian manner with water-masses being studied and (**c**) conducting targeted sampling of features anywhere in the water-column including at the chlorophyll (Chl.) maxima and within the oxygen minimum zone and the benthic boundary layer (B.L.). This, in turn, enables the host ship science party to carry out additional synoptic scientific observations and sampling with traditional cabled equipment.

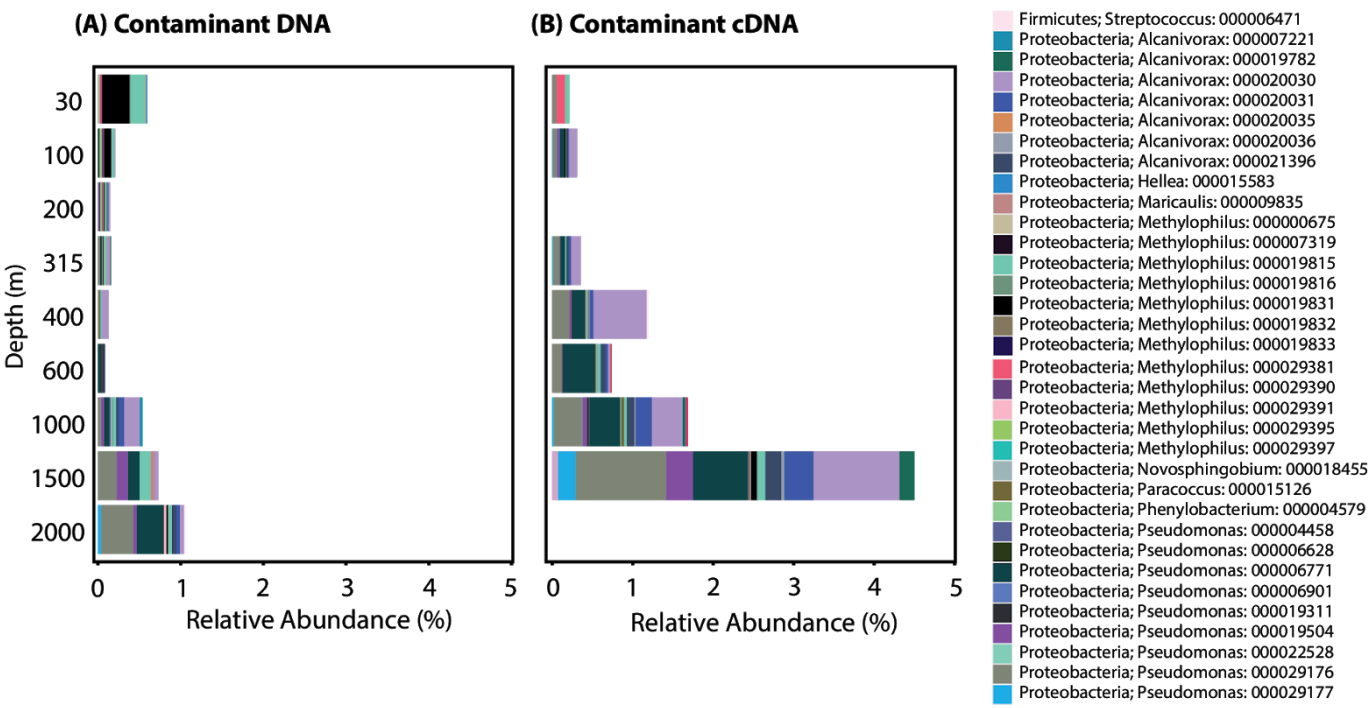

Fig. S2. Abundances of putative contaminant sequences in *Clio* DNA and cDNA samples. The relative abundances of putative contaminant taxa in *Clio* (**a**) DNA and (**b**) cDNA samples.

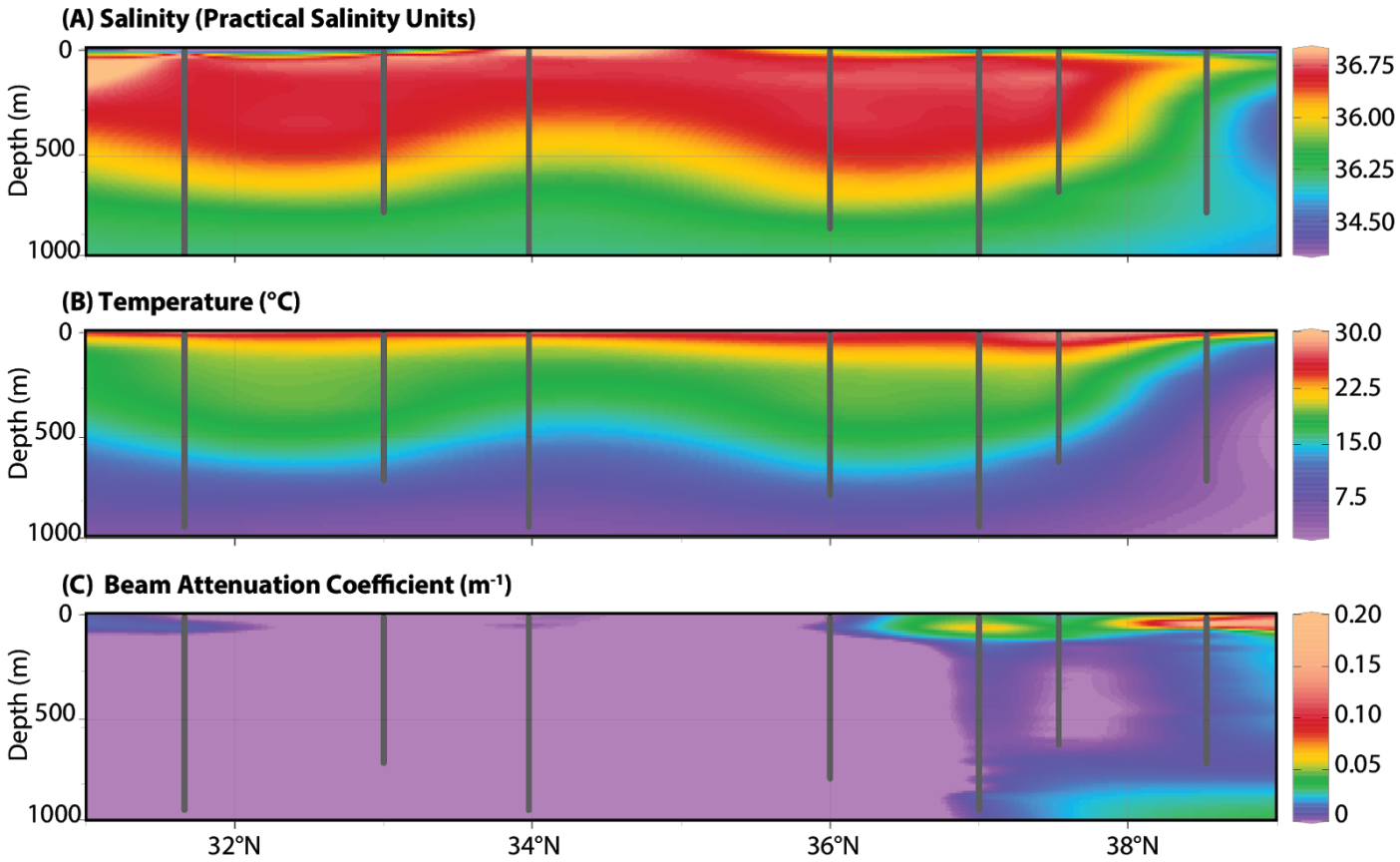

Fig. S3. Ocean environmental structure revealed using *Clio*. Vertical profiles of (**A**) salinity, (**B**) water temperature, and (**C**) optical beam attenuation coefficient (m^{-1}), which is inversely proportional to optical clarity, and a proxy for dissolved and suspended material. Grey dots indicate the location of point data that is spatially interpolated.

Supplemental Tables:

Table S1. *Clio* sampling configuration used in the Sargasso Sea study. Omics filters were 0.2 µm poresize, 142 mm diameter, Supor® polyethersulfone membranes (Pall Corporation). Carbon system filters were 142 mm diameter, precombusted Whatman® QMA quartz fiber (GE Healthcare).

	Sargasso Sea Study							Currently	
	Science Bay 1			Science Bay 2				Unused Capacity	
	Omics			Carbon				Bay 3	Bay 4
	Whole	Particles	Filtrate	Whole	Particles	Filtrate			
	Water			Water					
	(200 mL)	(100 L)	(200 ml)	(200 mL)	(100 L)	(200 ml)			
Media	seawater	Supor	seawater	seawater	QMA	seawater			
Derived		DNA/RNA	Nutrients		POC				
Analyses		Proteins	Metals						
		Metals							
Samples	10	10	10	10	10	10			
Filtered		1000 L			1000 L				

Table S2. North Atlantic shelf break DNA & RNA. This station, occupied during cruise AR20, was at the North Atlantic shelf break. RNAlater® (Qiagen) was used to preserve samples *in situ*. All filters were 0.2 µm poresize, 142 mm diameter, Supor® polyethersulfone membranes (Pall Corporation). Extracted RNA was reverse transcripted into ss cDNA. Putative contaminants based on taxonomy (Supplemental Notes: Methods DNA/RNA).

						DNA	cDNA[d]	DNA	cDNA
Station Location	*Clio* Dive	Sample	*In Situ* Pres.	Depth (m)	Vol. (L)	Number of Sequences		Contaminant Sequences	
39.569° N, 71.230° W	5	D1	yes	30	18.9	32,484	1,879	193	4
39.569° N, 71.230° W	5	D3	yes	100	42.3	30,665	19,652	64	60
39.569° N, 71.230° W	5	D5	yes	200	44.9	34,803	42	51	B.D.*
39.569° N, 71.230° W	5	D7	yes	315	46.4	38,280	11,989	60	42
39.569° N, 71.230° W	5	D9	yes	400	42.7	13,337	17,816	17	212
39.569° N, 71.230° W	5	D11	yes	600	43.2	17,994	23,710	15	174
39.569° N, 71.230° W	5	D13	yes	1000	57.5	20,723	24,493	111	412
39.569° N, 71.230° W	5	D15	yes	1500	53.3	2,193	11,026	16	496
39.569° N, 71.230° W	5	D17	yes	2000	45.5	17,770	70	187	B.D.

*Below detection limits.

Table S3. Bermuda Atlantic time-series study station proteins. The location of this station was 31.66667° N, -64.16667° W, occupied during cruise AE1811-BATS346 on April 14, 2018. All filters were 0.2 µm poresize, 142 mm diameter, Supor® polyethersulfone membranes (Pall Corporation).

					Urea Transporter	Alkaline Phosphatase	Taurine Transporter
Station Location	*Clio* Dive	Sample	Depth (m)	Vol. (L)	Spectral Counts	Spectral Counts	Spectral Counts
BATS Station	007	1-1	20	81.4	53	17	19
BATS Station	007	1-3	60	73.7	89	224	11
BATS Station	007	1-5	80	135.1	51	38	22
BATS Station	007	1-7	120	63.5	18	54	24
BATS Station	007	1-9	175	86.2	0	0	1
BATS Station	007	1-11	250	56.5	1	1	0
BATS Station	008	2-11	300	89.3	0	0	1
BATS Station	007	1-13	400	98.5	14	8	0
BATS Station	007	1-15	600	116.6	19	8	0
BATS Station	007	1-17	800	57.3	5	2	0

Table S4. Spectral counts and annotations of proteins from Clio dive 007 and 008 BATS station. (submitted separately and electronically)

Table S5. High-resolution chlorophyll maximum sampling profile. The location of this station was 31.66667° N, -64.16667° W, occupied during cruise AE1829-BATS352 on October 24, 2018. All filters were 0.2 µm poresize, 142 mm diameter, Supor® polyethersulfone membranes (Pall Corporation), with a 51µm poresize prefilter.

Station Location	*Clio* Dive	Sample	Depth (m)	Vol. (L)
BATS Station	014	B1	110	24.2
BATS Station	014	B3	115	23.8
BATS Station	014	B5	120	26.0
BATS Station	014	B7	125	25.0
BATS Station	014	B9	130	23.3

Table S6. Total exclusive spectral counts (SC) of selected proteins abundant at the top of the chlorophyll maximum from Clio-014 high-resolution sampling from 110m-130m. Selected proteins were significantly different between the top (110m) and center (120m) of chlorophyll maximum based on Fisher's Exact Test (p-values <0.05).

KEGG* Description	SC-110	SC-115	SC-120	SC-125	SC-130	Fisher p-value	best_hit_species
glnB; nitrogen regulatory protein P-II; K04752 nitrogen regulatory protein P-II 2	12	8	0	0	0	< 0.00010	Prochlorococcus marinus str. MIT 9202
C-phycoerythrin class I beta chain; K05377 phycoerythrin beta chain	12	7	0	0	0	< 0.00010	Synechococcus WH 8109
pgk; phosphoglycerate kinase (EC:2.7.2.3); K00927 phosphoglycerate kinase [EC:2.7.2.3]	44	36	6	5	2	< 0.00010	Prochlorococcus marinus str. AS9601
gap3; putative glyceraldehyde 3-phosphate dehydrogenase (EC:1.2.1.12); K00134 glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]	32	31	1	2	1	< 0.00010	Prochlorococcus marinus str. MIT 9301
kaiC; circadian clock protein KaiC; K08482 circadian clock protein KaiC	12	14	4	0	0	0.016	Prochlorococcus marinus str. AS9601
glnA; glutamine synthetase, glutamate--ammonia ligase (EC:6.3.1.2); K01915 glutamine synthetase [EC:6.3.1.2]	142	134	63	46	14	< 0.00010	Prochlorococcus marinus str. AS9601
alkaline phosphatase (EC:3.1.3.1); K01077 alkaline phosphatase [EC:3.1.3.1]	173	172	85	34	0	< 0.00010	Prochlorococcus marinus str. NATL1A
ABC transporter, substrate binding protein, phosphate; K02040 phosphate transport system substrate-binding protein	124	116	61	42	6	< 0.00010	Prochlorococcus marinus str. MIT 9301
psbO; photosystem II manganese-stabilizing protein; K02716 photosystem II	71	71	45	42	17	0.00037	Prochlorococcus marinus str. AS9601

oxygen-evolving enhancer protein 1							
putative UreA ABC transporter, substrate binding protein; K11959 urea transport system substrate-binding protein	208	210	196	216	82	0.0061	Prochlorococcus marinus str. NATL2A

*Kyoto Encyclopedia of Genes and Genomes.

Table S7. Total exclusive spectral counts (SC) of selected proteins abundant at the center of the chlorophyll maximum from Clio-014 high-resolution sampling from 110m-130m. Selected proteins were significantly different between the top (110m) and center (120m) of chlorophyll maximum based on Fisher's Exact Test (p-values <0.05).

KEGG* Description	SC-110	SC-115	SC-120	SC-125	SC-130	Fisher p-value	best_hit_species
iron ABC transporter iron-binding protein, putative; K02012 iron(III) transport system substrate-binding protein	0	0	6	6	4	0.027	SAR324 cluster bacterium JCVI-SC AAA005
TonB-dependent receptor, plug	1	5	12	12	0	0.0047	gamma proteobacterium SCGC AAA076-P13
carboxysome shell protein CsoS2	7	20	43	37	1	< 0.00010	Prochlorococcus marinus str. NATL2A
psaA; photosystem I P700 chlorophyll a apoprotein A1; K02689 photosystem I P700 chlorophyll a apoprotein A1	15	29	48	33	14	0.00036	Prochlorococcus marinus str. NATL1A
fructose-1,6-bisphosphate aldolase (EC:4.1.2.13); K01623 fructose-bisphosphate aldolase, class I [EC:4.1.2.13]	18	37	49	43	9	0.0016	Prochlorococcus marinus str. NATL1A
opuAC; substrate-binding region of ABC-type glycine betaine transport system; K02002 glycine betaine/proline transport system substrate-binding protein	30	37	57	48	14	0.028	Pelagibacter HTCC7211
photosystem II manganese-stabilizing protein; K02716 photosystem II oxygen-evolving enhancer protein 1 || psbO; photosystem II manganese-stabilizing protein; K02716 photosystem II oxygen-evolving enhancer protein 1	38	66	71	108	33	0.018	Prochlorococcus marinus str. NATL1A
chlorophyll a/b binding light harvesting protein PcbD; K08919 chlorophyll a/b binding light-harvesting protein	41	58	75	91	42	0.02	Prochlorococcus marinus str. NATL1A

PcbB							
glyceraldehyde 3-phosphate dehydrogenase(NADP+) (phosphorylating) (EC:1.2.1.13); K00150 glyceraldehyde-3-phosphate dehydrogenase (NAD(P)) [EC:1.2.1.59]	54	79	96	80	31	0.014	Prochlorococcus marinus str. NATL2A
isiB; flavodoxin FldA; K03839 flavodoxin I	44	65	78	68	5	0.026	Prochlorococcus marinus str. NATL1A

*Kyoto Encyclopedia of Genes and Genomes.

Table S8. Spectral counts and annotations of 595 identified proteins from Clio-014 high-resolution sampling 110-130m at BATS station. (submitted separately and electronically)

Table S9. Sargasso Sea dissolved inorganic nitrogen.

Station	*Clio* Dive	Lat.	Lon.	Sample	Depth (m)	Nitrate + Nitrite (μmol L^{-1})	Ammonium (μmol L^{-1})
1	18	31.66667	-64.16667	S1	10	0.40	1.38
1	18	31.66667	-64.16667	S18	150	0.96	0.03
1	18	31.66667	-64.16667	S17	200	1.49	0.04
1	18	31.66667	-64.16667	S16	300	2.72	B.D.*
1	18	31.66667	-64.16667	S15	400	3.50	0.08
1	18	31.66667	-64.16667	S14	600	9.52	0.06
1	18	31.66667	-64.16667	S13	800	15.3	0.12
1	19	31.66588	-64.16638	S10	20	0.15	0.10
1	19	31.66588	-64.16638	S11	60	0.16	0.03
1	19	31.66588	-64.16638	S12	90	0.19	0.22
1	19	31.66588	-64.16638	S13	115	1.50	0.34
1	19	31.66588	-64.16638	S14	175	1.95	0.15
1	19	31.66588	-64.16638	S15	250	1.75	0.26
1	19	31.66588	-64.16638	S16	350	3.42	0.43
2	20	33.00062	-65.99615	S1	20	0.19	0.16
2	20	33.00062	-65.99615	S2	60	0.17	0.05
2	20	33.00062	-65.99615	S3	90	0.15	0.43
2	20	33.00062	-65.99615	S4	120	0.83	0.27
2	20	33.00062	-65.99615	S5	150	1.44	0.07
2	20	33.00062	-65.99615	S6	200	2.18	0.19
2	20	33.00062	-65.99615	S7	400	3.61	0.48
2	20	33.00062	-65.99615	S8	600	3.36	0.03
2	20	33.00062	-65.99615	S9	800	19.8	0.27
3	21	33.97500	-66.99833	S10	10	0.54	0.12
3	21	33.97500	-66.99833	S1	20	2.58	0.14
3	21	33.97500	-66.99833	S11	30	0.45	0.05
3	21	33.97500	-66.99833	S12	45	0.43	0.07

3	21	33.97500	-66.99833	S2	60	0.33	0.04
3	21	33.97500	-66.99833	S13	75	0.41	0.10
3	21	33.97500	-66.99833	S3	90	0.24	0.22
3	21	33.97500	-66.99833	S14	105	0.40	0.25
3	21	33.97500	-66.99833	S4	120	1.47	0.13
3	21	33.97500	-66.99833	S5	150	2.28	0.05
3	21	33.97500	-66.99833	S15	175	3.14	0.08
3	21	33.97500	-66.99833	S6	200	2.59	0.21
3	21	33.97500	-66.99833	S7	400	Q.C.†	Q.C.
3	21	33.97500	-66.99833	S8	600	Q.C.	Q.C.
3	21	33.97500	-66.99833	S16	700	4.25	0.07
3	21	33.97500	-66.99833	S9	800	19.8	0.08
3	21	33.97500	-66.99833	S17	1000	17.5	0.06
5	22	36.00000	-69.00100	S1	20	0.25	0.31
5	22	36.00000	-69.00100	S2	40	0.11	1.45
5	22	36.00000	-69.00100	S3	70	0.09	0.39
5	22	36.00000	-69.00100	S4	100	0.36	0.37
5	22	36.00000	-69.00100	S5	150	2.20	0.18
5	22	36.00000	-69.00100	S6	200	1.83	0.52
5	22	36.00000	-69.00100	S7	400	3.37	0.22
5	22	36.00000	-69.00100	S8	600	6.08	0.03
5	22	36.00000	-69.00100	S9	880	22.7	0.10
6	23	36.99883	-70.00000	S1	20	0.60	0.32
6	23	36.99883	-70.00000	S2	40	0.01	0.24
6	23	36.99883	-70.00000	S3	70	B.D.	0.08
6	23	36.99883	-70.00000	S4	110	0.43	0.21
6	23	36.99883	-70.00000	S5	150	1.05	0.53
6	23	36.99883	-70.00000	S6	200	2.38	0.41
6	23	36.99883	-70.00000	S7	400	4.98	0.08
6	23	36.99883	-70.00000	S8	600	7.64	0.05
6	23	36.99883	-70.00000	S9	800	23.2	0.34

6	23	36.99883	-70.00000	S18	102.5	0.48	0.65
6	24	37.00000	-70.00048	S1	1000	15.2	0.05
6	24	37.00000	-70.00048	S10	62.5	0.42	0.13
6	24	37.00000	-70.00048	S11	72.5	0.23	0.03
6	24	37.00000	-70.00048	S12	82.5	0.34	0.05
6	24	37.00000	-70.00048	S13	92.5	0.69	B.D.
6	24	37.00000	-70.00048	S14	102.5	0.61	0.11
6	24	37.00000	-70.00048	S15	112.5	1.33	0.04
7	25	37.53253	-70.52992	S1	20	0.36	B.D.
7	25	37.53253	-70.52992	S2	40	0.27	0.05
7	25	37.53253	-70.52992	S3	70	0.32	0.72
7	25	37.53253	-70.52992	S4	150	1.63	0.23
7	25	37.53253	-70.52992	S5	300	3.33	0.06
7	25	37.53253	-70.52992	S6	500	9.84	0.03
7	25	37.53253	-70.52992	S7	700	22.9	0.04
7	25	37.53253	-70.52992	S8	112.5	1.08	0.04
7	25	37.53253	-70.52992	S9	111	0.56	0.06
7	25	37.53253	-70.52992	S10	107.5	0.74	0.05
7	25	37.53253	-70.52992	S11	112.5	0.98	0.07
7	25	37.53253	-70.52992	S12	117.5	0.84	0.02
7	25	37.53253	-70.52992	S13	122.5	1.02	0.21
8	26	38.52808	-70.84102	S1	20	0.61	0.13
8	26	38.52808	-70.84102	S2	40	0.36	0.13
8	26	38.52808	-70.84102	S3	120	10.3	0.03
8	26	38.52808	-70.84102	S4	175	18.7	0.14
8	26	38.52808	-70.84102	S5	250	23.9	0.05
8	26	38.52808	-70.84102	S6	400	23.7	0.02
8	26	38.52808	-70.84102	S7	600	20.3	0.07
8	26	38.52808	-70.84102	S8	800	18.9	0.04

*Below detection limit.

†Quality control comparison to WOCE section A22 indicates these analyses were anomalously low, so they are not reported.

Table S10. Temperature, salinity, chlorophyll-a, turbidity, and optical beam transmission.
(submitted separately and electronically)