

RESEARCH ARTICLE

MARINE MICROBIOME

Ocean biogeochemistry modeled with emergent trait-based genomics

V. J. Coles,^{1*} M. R. Stukel,² M. T. Brooks,¹ A. Burd,³ B. C. Crump,⁴ M. A. Moran,³ J. H. Paul,⁵ B. M. Satinsky,³ P. L. Yager,³ B. L. Zielinski,⁴ R. R. Hood¹

Marine ecosystem models have advanced to incorporate metabolic pathways discovered with genomic sequencing, but direct comparisons between models and “omics” data are lacking. We developed a model that directly simulates metagenomes and metatranscriptomes for comparison with observations. Model microbes were randomly assigned genes for specialized functions, and communities of 68 species were simulated in the Atlantic Ocean. Unfit organisms were replaced, and the model self-organized to develop community genomes and transcriptomes. Emergent communities from simulations that were initialized with different cohorts of randomly generated microbes all produced realistic vertical and horizontal ocean nutrient, genome, and transcriptome gradients. Thus, the library of gene functions available to the community, rather than the distribution of functions among specific organisms, drove community assembly and biogeochemical gradients in the model ocean.

Genomic technologies that elucidate the links between microbes and biochemistry have revolutionized our understanding of marine ecology and the cycling of nitrogen, carbon, and other elements (1–5). However, the models used to predict regional and global ocean biogeochemistry and the microbes mediating their variability do not simulate quantities that can be directly compared with genomic data. In deep chemosynthetic environments, simulating the genes of each individual species has proven fruitful (6); however, for the vast diversity of microbes in the open ocean, most of which are uncultured, this approach is impractical. Hence, we developed the Genome-based EmergeNt Ocean Microbial Ecosystem (GENOME) model, in which organisms with genomes and transcriptomes are simulated and compared with “omics” data. The model was inspired by trait-based emergent phytoplankton community models (7), but we took a different approach to developing model microbes with diverse metabolic potential. The model addresses how the distribution of metabolic functions among members of microbial communities influences ocean biogeochemistry (8), asking whether communities with distinctly different types of organisms but common metabolic ca-

pabilities create similar marine biogeochemical gradients over space and time.

Constructing model communities

The GENOME model was incorporated into a general circulation model of the tropical and subtropical Atlantic Ocean to provide flow fields and environmental conditions for the model com-

munities. To construct model communities, organisms were first randomly assigned sizes (1 to 2000 μm) spanning the range from bacteria to large phytoplankton. General substrate uptake affinity, growth rate, mortality, and sinking parameters were established, which varied with size on the basis of observations and theory (fig. S1 and table S2) (9–11). These allometric equations described functional responses for all organisms as a function of their size, but the functional responses were only applied when an organism carried a gene relevant to the response. For example, functional responses to light were not invoked for organisms lacking a light-harvesting gene. The model was not especially sensitive to the slope of these scaling relationships because the cumulative impact of the traits associated with each gene led to high variance in the scaling relationships, consistent with observations of natural populations (fig. S1).

Each organism was then randomly assigned a degree of complexity (number of genes selected from 20 gene functional groups; table S3). Here, we use the term “gene” to refer to a single marker gene used to identify a biochemical pathway that might involve many coordinating genes. To avoid generating “super” microbes with all gene combinations simultaneously, the model penalized complexity by linearly reducing maximum potential growth rate as the gene number increased. This accords with the hypothesis that microbes with smaller genomes require less carbon and other limiting nutrients to produce a new cell (12) but are unable to maintain as many specialist gene functions (13), although in practice this may not always be the case (14). In the

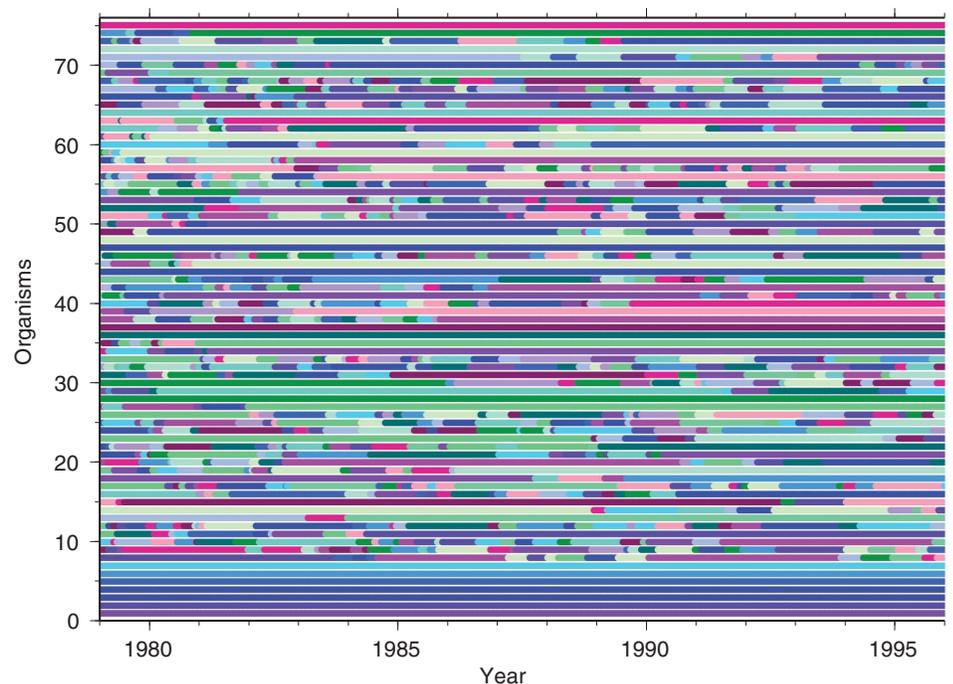


Fig. 1. Organism replacement over time in the GENOME model. The first seven rows are substrates that are constant over the model run. The next 68 rows delineate organisms by color. Color change denotes replacement of an organism.

¹Horn Point Laboratory, University of Maryland Center for Environmental Science (UMCES), Post Office Box 775, Cambridge, MD 21613, USA. ²Department of Earth, Ocean, and Atmospheric Science, Florida State University, 117 North Woodward Avenue, Tallahassee, FL 32306-4520, USA. ³Department of Marine Science, University of Georgia, Athens, GA 30602, USA. ⁴College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR 97331, USA. ⁵University of South Florida, 140 Seventh Avenue South, St. Petersburg, FL 33701, USA. *Corresponding author. Email: vcoles@umces.edu

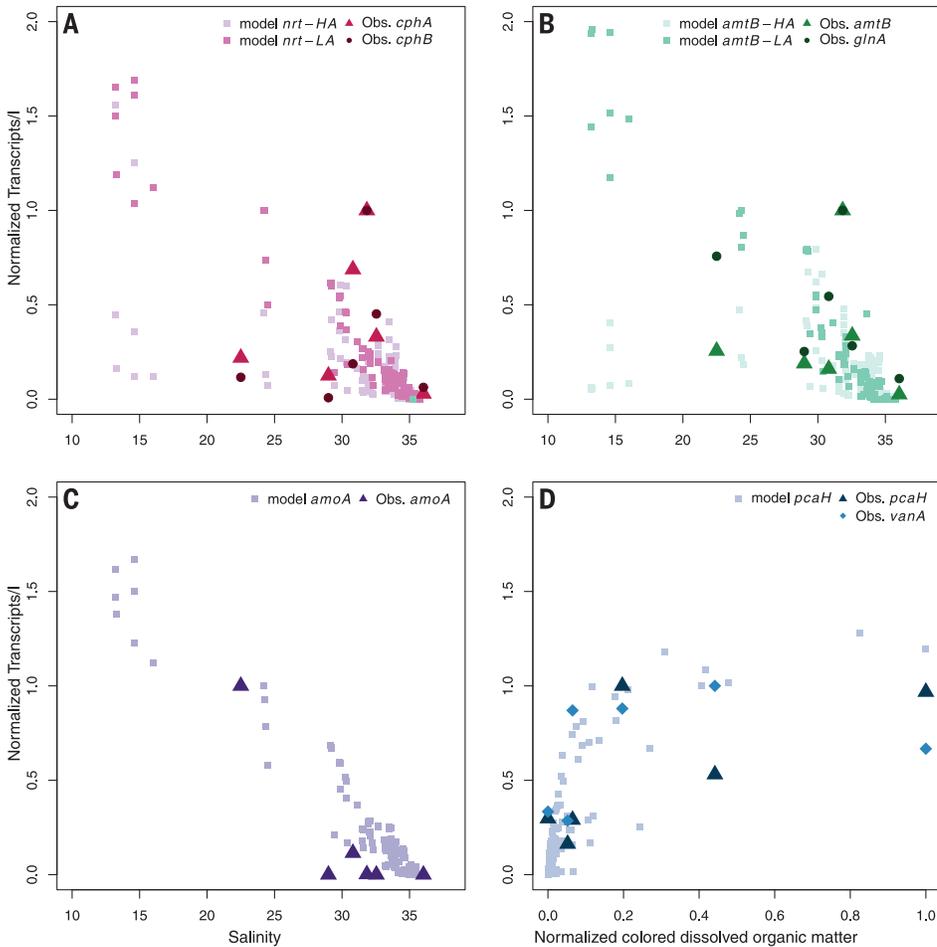


Fig. 2. Normalized transcript concentrations in the Amazon River plume from June observations in 2010 and modeled June. In (A) to (C), normalized transcript concentrations are shown along the salinity gradient. (A) Transcription of genes involved in nitrate uptake and storage. (B) Transcription of the *amtB* gene for ammonium transport and the *glnA* gene that indicates nitrogen limitation. (C) Transcription of the *amoA* gene for nitrification. (D) Transcription of genes for degrading lignin-related or aromatic compounds along the normalized colored dissolved organic matter gradient. l, liter.

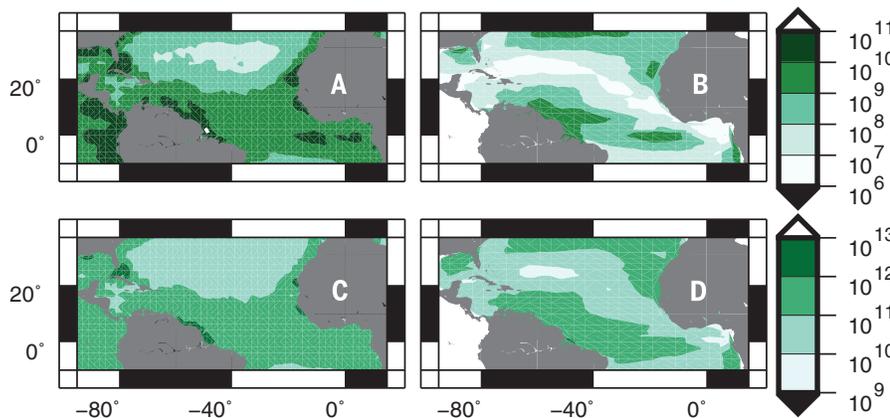


Fig. 3. Satellite observations and one model estimate of upper ocean particle density in June. (A) Satellite-derived $>5\text{-}\mu\text{m}$ particle number per cubic meter. (B) Model run showing surface $>5\text{-}\mu\text{m}$ particle number per cubic meter assuming conversions of 20 and 25 fg of nitrogen per cell volume (measured in cubic micrometers) for nano- and microparticles, respectively. (C) Satellite-derived $<5\text{-}\mu\text{m}$ particle number per cubic meter. (D) Model run showing surface $<5\text{-}\mu\text{m}$ particle number per cubic meter assuming a conversion of 1 fg of nitrogen per cell volume (measured in cubic micrometers) for picoparticles. Observations are 2003–2007 SeaWiFS (Sea-viewing Wide Field-of-view Sensor) climatological data.

model, organisms with highly streamlined genomes had reduced energy investment in replication and reduced expression and maintenance of specialist gene functions. In contrast, organisms with high genome complexity were capable of responding better to more nutrient- and energy-replete environmental conditions. This assumption was consistent with research illustrating genome streamlining in some marine bacterioplankton (15, 16), yet allowed specialist life strategies that enabled more complex organisms to thrive in oligotrophic environments.

The model constructed genomes for each organism by randomly assigning genes from the predetermined ensemble (table S3) of known genes or environmental responses for which an indicator gene has not been identified. The ensemble accounts for the cycling of nitrogen and organic matter in the Amazon River plume, for which we have “omics” observations and biogeochemical data (17–19). The model genetic pathways each mediate a biochemical transformation or metabolic process. Organismal responses were adapted using the costs and benefits estimated for each genetic pathway in the organism’s genome (fig. S1 and table S3). The actual costs and benefits of maintaining or transcribing genetic pathways are unknown for most metabolic functions; thus, their relative magnitudes were estimated indirectly. First, the organismal integrated cost-benefits were scaled to create variance in the size-structured environmental response functions, consistent with observations (fig. S1). Second, the relative cost-benefits were adjusted to ensure that all the model genes existed in at least one organism somewhere within the model domain. Third, for nitrogen fixation and nitrification, for which we have observations of genes, transcripts, and biochemical rates, cost-benefits were adjusted until the model simulated reasonable integrated basinwide rates of nitrogen fixation and local rates of nitrification. Last, our limited direct observations of gene and transcript abundance provided a constraint on the model. Importantly, all organisms given a particular genetic function received the same cost-benefits for that gene, although their integrated responses varied widely according to the other genes they carried. Because different microbial communities emerged in each model run, the basinwide biochemical transformations mediated by the community were not tuned to match observations and were thus free to vary as environmental conditions changed.

Seven substrates and 68 microbes coexisted at any given time, which is computationally taxing but still insufficient to explore all genetic combinations. To incorporate broader genetic potential, organisms whose biomass never rose above 1% of the local community anywhere in the model were replaced with a new organism. Most organisms were rapidly replaced because their genomes did not encode a sustainable organism in the model ocean. Others persisted for long periods but were ultimately replaced by new organisms that were better adapted to exploiting the model ocean and community (Fig. 1).

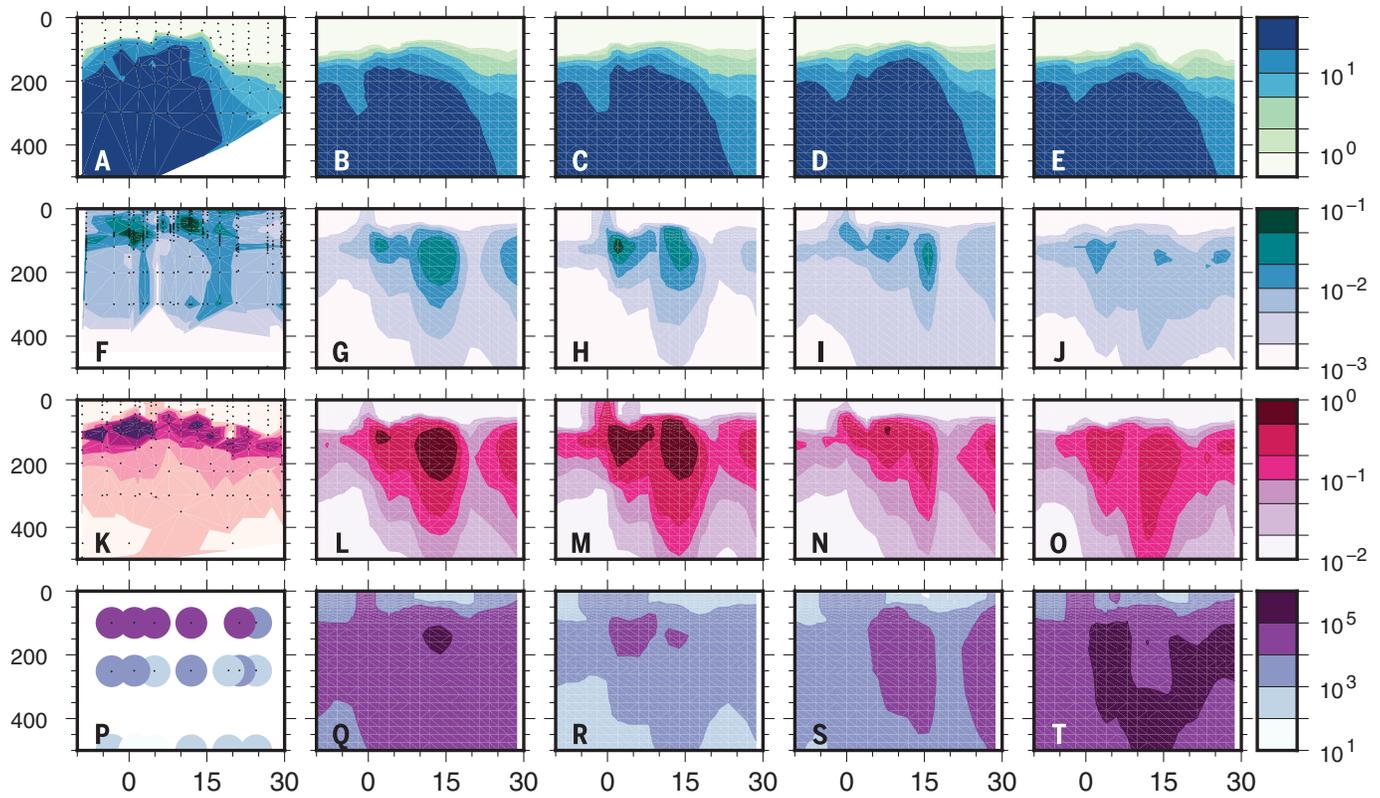


Fig. 4. Vertical sections from the Atlantic Meridional Transect and model. Sections are shown from observations approximately along 28°W in 2005 [(A), (F), (K), and (P)] and from the model runs at 28°W in June (all other panels). Depth is shown in meters on the y axis and latitude in degrees on the x axis. (A to E) Nitrate (millimoles per cubic meter).

(F to J) Ammonium (millimoles per cubic meter). (K) Observed nitrite (millimoles per cubic meter). (L to O) Model runs showing normalized transcript concentration of archaeal *amoA* (dimensionless). (P to T) Observed concentration of *amoA* genes and model runs showing concentration of *amoA* genes (gene copies per cubic meter).

Each model simulation included more than 2000 organisms over the course of 20 years.

Modeling regulation of transcription

Equations for each organism were dynamically linked to dissolved and particulate substrates, such as nitrate, ammonium, and marine and terrestrially derived organic matter (eqs. S1 to S34). Model microbes had multiple strategies for energy and substrate acquisition depending on their individual genomes. However, at each model location and time, the model organisms transcribed genes with the highest growth potential given the local environmental conditions. Thus, model microbes up- and down-regulated metabolic processes rapidly in response to external environmental variability.

One goal underpinning development of the GENOME model was to have the capacity to compare measurements and models of gene and transcript concentration. Expression levels of genes for substrate uptake or light harvesting were modeled as the sum of constitutive transcription, regulated transcription, and steady-state transcription (eq. S30) and were parameterized using observations of transcription in the Amazon plume. For example, observations of the expression ratio (transcripts per gene) in the plume increased with increasing numbers of transcripts (fig. S4), indicating that elevated transcript con-

centration was not solely due to higher biomass, but was also caused by up-regulation of nitrogen-cycling genes in the western tropical North Atlantic (17). As a result, we modeled constitutive and regulated transcription of these genes as a multiplicative function of nutrient limitation and growth rate (eq. S30). However, the smallest values for modeled normalized transcription were too low compared with observations until very low levels of steady-state transcription, proportional to biomass, were also added to the model. In genes with an observational analog, the comparisons suggested that all three factors may be important in setting the community transcript concentration for substrate transporters. Genes that selected for processes such as biogenic shell formation or motility were modeled as constitutively transcribed as a function of energy availability (20) or the local concentration of particles (eqs. S32 and S33) (21).

We observed high abundances of *cphA* and *cphB*, genes involved in intracellular nitrogen storage, in the nutrient-rich, low-salinity plume, and normalized transcript abundance decreased with increasing salinity and decreasing nitrate. A similar pattern emerged in the model for genes regulating nitrogen uptake (Fig. 2A). The ammonium transporter gene *amtB* and ammonium assimilation gene *glnA* were also abundant in the fresher region of the plume in observations

and high- and low-affinity *amtB* model gene pathways, coincident with observations of higher ammonium concentrations (Fig. 2B). Observations of elevated *amtB* gene expression at a salinity of 32 was associated with production of ammonium by symbiotic diazotrophic cyanobacteria, which was then assimilated by host diatoms (22–24). Because the model assumed that fixed nitrogen was used by the organism that fixed it, ammonium assimilation was not simultaneously up-regulated. This presents an interesting challenge for future model evolution. Observed transcript abundance for the *amoA* gene encoding for nitrification was relatively low, with maximum transcript abundance only 4% of the maximum for *amtB*, and its and maximal abundance occurred in low-salinity water. The model showed a similar pattern, although it overestimates nitrification at intermediate salinity levels (Fig. 2C). Measurable nitrite (32 nmol liter⁻¹) was observed in the low-salinity water with the highest *amoA* transcripts, indicating competition for ammonium between nitrifiers, heterotrophs, and photoautotrophs, as occurred in the model. Genetic signatures of the processing of land-derived organic material include transcription of genes such as *pcaH* and *vanA* (Fig. 2D). Both modeled and observed transcription of these genes indicated higher effort allocated to degradation of terrestrial organic matter where

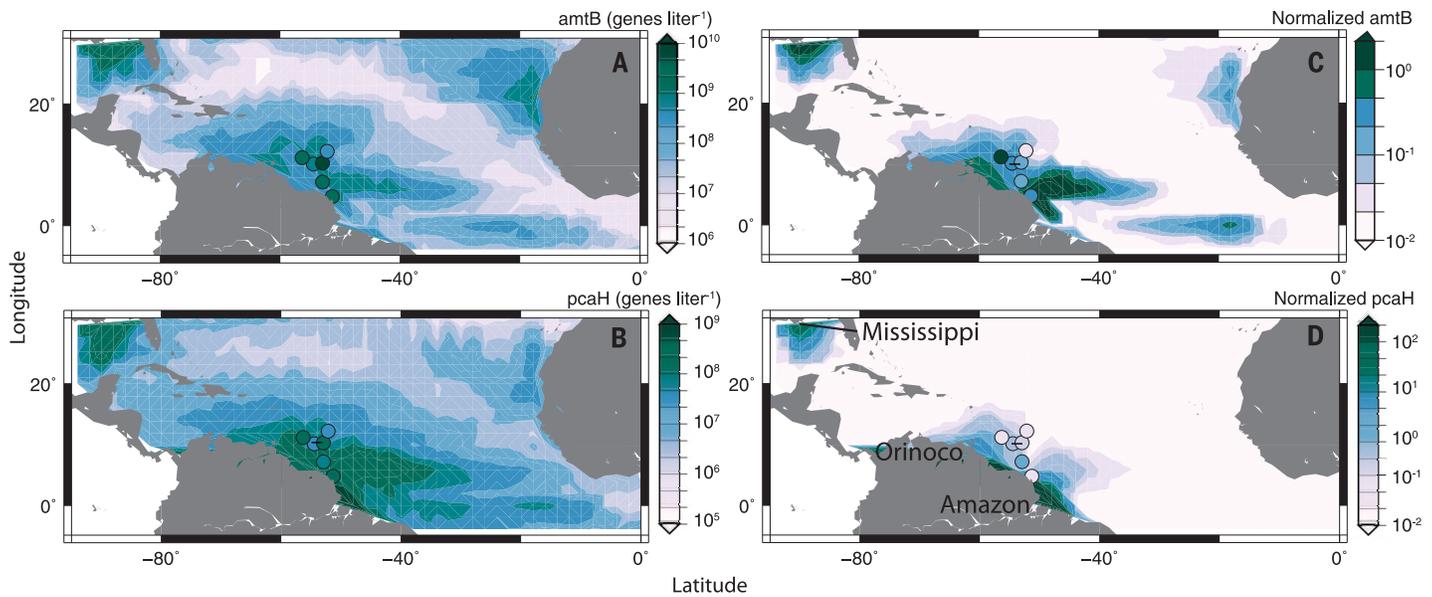


Fig. 5. Surface genes and transcripts for *amtB* and *pcaH* in June.

(A) Modeled *amtB* gene concentrations. Observed *amtB* gene concentrations are overlaid in circles. (B) Modeled *pcaH* gene concentrations. Observed *pcaH* gene concentrations are overlaid in circles. (C) Modeled normalized *amtB* transcript concentrations. Observed normalized *amtB* transcript con-

centrations are overlaid in circles. (D) Modeled normalized *pcaH* transcript concentrations. Observed normalized *pcaH* transcript concentrations are overlaid in circles. We assume a single gene copy per organism. Transcripts are normalized by the maximum value in the region of the observations. The color scales differ for all panels. (Additional runs are shown in fig. S5.)

centrations are high, with a sharp decay in transcript abundance as organic matter concentrations dip to less than 20% of the highest observed concentrations (Fig. 2D).

Large-scale model validation

We ran the GENOME model repeatedly with identical physical forcing but different microbial communities, each randomly derived from a common genetic pool (table S3). Satellite observations of particle size and density (25) provided spatial patterns for validating the model organism density in different size classes (Fig. 3 and fig. S2). However, the satellite-derived particle density also included abiotic particles, such as Saharan dust, that are not represented in the model. Saharan dust collected in Atlantic sediment traps mainly falls into size classes greater than 5 μm , and aerosol optical thickness measurements indicate higher dust concentrations between 20°N and the equator east of 40°W (26) than to the north and south. This is the area where the model underestimated the number of particles larger than 5 μm , compared with observations. The model also underestimated particle abundance in the oligotrophic gyres, which is consistent with underestimated vertical nutrient fluxes at the spatial resolution of this model (27). The subtropical gyres in the Atlantic are dominated by small particles (<5 μm), whereas coastal, river-influenced, and mixing or upwelling regions with higher nutrient concentrations have greater densities of particles and a higher fraction of large particles (>5 μm) (Fig. 3, A and C). The model runs showed the same pattern (Fig. 3, B and D, and fig. S2), which was determined by nutrient availability and uptake. Small organisms in the oligotrophic ocean benefit

from larger ratios of surface area to volume (28) and lower sinking rates (29). Larger cells thrive in nutrient-rich coastal and upwelling areas. However, microbes have evolved strategies for compensating for size constraints, such as regulating their buoyancy through ballasting or energetic regulation of turgor pressure (30) or using motility to seek out substrate plumes emitted from moving particles (31). This allows some large organisms to maintain their presence in oligotrophic regions (32). In the model, specialist genes for increasing nutrient or substrate uptake, by allocating energy to motility or to high-affinity nutrient transporters, provided a mechanism to offset low nutrient availability (fig. S3). Consequently, each model run maintained populations of large cells that expressed genes encoding for specialist strategies in oligotrophic regions.

All the GENOME runs established vertical nutrient distributions that were similar to observations, regardless of the specific organisms that emerged in each run. Some model errors, such as broader and deeper nutrient gradients, were primarily caused by coarse physical model resolution (27). Vertical gradients in nitrate and ammonium (Fig. 4, A to J) were determined jointly by abiotic processes, including physical advection and sinking, and biologically mediated processes. One of these processes, nitrification (conversion of ammonium to nitrate, carried out by organisms with *amoA* genes), illustrates how the model community shaped the distribution of nitrogen species. Although there are no direct observations of *amoA* transcript concentration, high concentrations of modeled archaeal *amoA* transcripts (Fig. 4, L to O) occurred coincidentally with observations of nitrite (Fig. 4K), which is generated as an intermediate in the

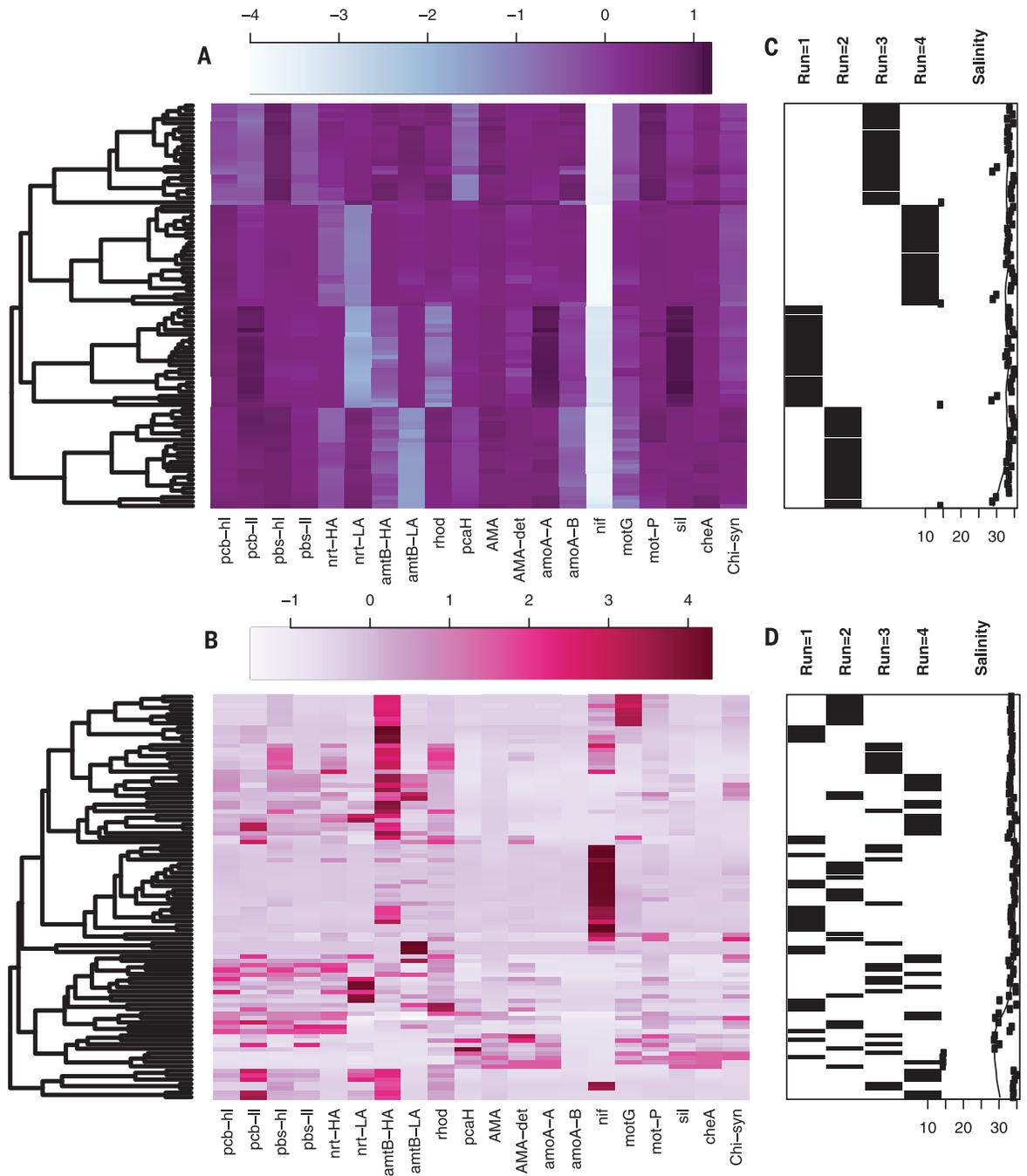
process of nitrification, indicating that the model nitrifying communities occurred in the correct locations. Observations of vertical gradients in archaeal *amoA* genes (Fig. 4P) showed that the greatest concentrations occurred at ~100 m and decreased toward 500 m depth (33). The model runs (Fig. 4, Q to T) exhibited a similar gradient in most cases. However, the similarity between model runs was lower for genes than transcripts because genes indicate the potential of the community, whereas transcript concentrations reflect the realized model biochemical transformation. It is unlikely that nitrification transcription would be completely silenced, given that observations show similar ratios of genes to transcripts across a broad gradient of depths and nitrification rates (34). Thus, the model formulation, which allowed nearly complete down-regulation of processes, was likely oversimplified.

Emergent communities and metabolisms in the model

Not just nitrification, but all biochemical transformations in the model, such as the return of organic nutrients to inorganic nutrients, were determined by organisms on the basis of their biomass and growth rates. Thus, the nitrogen transformation rates were not adjusted to set the model nutricline. In conventional efforts to simulate the biogeochemical gradients in the Amazon River plume and elsewhere, model parameters are adjusted to accurately represent these vertical gradients (35–37) by selecting environmental rates and response functions and by balancing organic matter mineralization and sinking. However, the optimal formulation for each process is likely to differ under changing climate. Because these processes are emergent in the

Fig. 6. Comparison of metagenomic and transcriptomic similarities between model runs along a salinity gradient.

(A) Metagenomes along the Amazon River plume salinity gradient in each of four model runs in June, clustered by similarity (genes per cubic meter). (B) Normalized transcription along the Amazon River plume salinity gradient in each of four model runs in June, clustered by similarity. (C and D) Model run and salinity associated with each sample. Genes are described in table S3.



GENOME model, the composition of the emergent communities, the rate of biochemical transformations that the microbes mediate, and the resultant metagenome and transcriptome can freely adapt to a changing environment.

The gene for ammonium transport into cells, *amtB*, was broadly distributed in the model (Fig. 5A), with a spatial pattern similar to that of ecosystem biomass (Fig. 3), because this function is ubiquitous. Observations of *amtB* gene concentrations along a gradient in the Amazon River plume (17, 38) broadly matched the model; however, the spatial scale of the observations was smaller than the gradients achieved in the model. *amtB* was also actively expressed in the model (Fig. 5B), because recycling food webs

fuelled by ammonium dominated in the modeled subtropical and tropical Atlantic.

In contrast, a gene for aromatic compound metabolism (*pcaH*, indicating aromatic ring cleavage) was observed at lower absolute abundance than *amtB*, although it remained finite across the sampling domain (Fig. 5C), and the model projection was consistent with this pattern. Observed *pcaH* gene expression is generally higher in water with high organic matter content (Fig. 2D). This resulted in higher transcript concentrations in the model near river mouths, such as the Amazon, Orinoco, and Mississippi, where aromatic compounds are introduced (Fig. 5D). Each of the GENOME model runs developed a different community with diverse individual gene-

tic capabilities, but for which the community transcriptomes and genomes were consistent with the limited observations (Fig. 2 and figs. S4 and S5).

At present, we lack mechanistic understanding of the cost-benefit trade-offs for marine microbial communities when carrying, expressing, and regulating genes or synthesizing proteins, and therefore we inferred these indirectly. Furthermore, some processes were included in the model for which we have no identified genetic pathway or marker genes (e.g., genes associated with grazing). As quantitative observations of gene abundance and expression expand, data-assimilative techniques to minimize the difference between modeled and observed gene copy

and transcript numbers could provide authentic constraints on cost-benefits for modeling. For example, in chemosynthetic environments for which the Gibbs free energies of reactions are known, a model of gene abundance and production rate has been tuned to replicate observations (6). The costs and benefits of these gene processes could potentially be determined from their best-fit model parameters.

Different communities, common metabolic functions

Microbial consortia do not align themselves passively along preexisting ocean biogeochemical gradients; rather, they shape the biogeochemical environment in a tight feedback loop that couples the physical environment to biological community structure. Thus, observations of taxa and gene transcription across large gradients in nutrients following the path of the Amazon River plume (17, 19, 24, 39) provided context for testing the ability of the model to reproduce observed patterns. Three potential outcomes of GENOME model simulations were envisioned. Initially, we considered it possible that random allocation of genes to organisms would not result in correct and repeatable biogeochemical gradients because the organisms needed to exploit specific conditions might be absent. However, this was not the case (Fig. 4 and fig. S2). Second, it was possible that highly similar emergent communities developed in each model simulation and led to common biogeochemical gradients. Third, it was possible that different emergent communities developed in each model simulation, but their common metabolic functions created similar biogeochemical gradients through space and time (i.e., that metabolic capabilities rather than specific taxa regulated ocean biogeochemistry). The second and third outcomes were evaluated by comparing the similarity of community genomes between model simulations with the similarity of community metabolic expression between model simulations.

The similarity between metagenomes (Fig. 6A) and normalized metatranscriptomes (Fig. 4C) for each model run was computed at locations along the salinity gradient (Fig. 6, B and D) in the model Amazon plume and clustered hierarchically. Salinity is a proxy for biogeochemical and ecological community gradients that are attributable to dilution and biogeochemical dynamics. In general, as salinity increases, inorganic and organic nutrients decrease, and communities shift from groups representing Gammaproteobacteria, Flavobacteria, and diatoms to those containing oligotrophic bacteria, such as SAR11 and autotrophic cyanobacteria (17, 22, 40–42).

In the metagenome heatmap (Fig. 6A), different locations across a broad biogeochemical gradient in a single model run were more similar than equivalent locations across model runs, indicating that the emergent model communities differed widely in genome content between runs. However, the metatranscriptome heatmap, which indicates the actual metabolisms in use at each location (Fig. 6B), clustered across runs,

with similar metabolisms expressed at equivalent salinities. For example, at the lowest salinities, genes encoding for heterotrophic processes involving degradation of dissolved and particulate organic matter (*pcb*, *AMA*, and *AMA-det*) and processes reducing mortality (*motG*, *mot-P*, *sil*, and *cheA*) were highly expressed across runs, whereas at intermediate salinities, photosynthesis and nitrogen uptake genes (*pcb*, *pbs*, *nrt*, and *amtB*) were highly expressed. Thus, the genetic composition of each GENOME assemblage differed between runs, but the metabolic functions were similar for a given environment across model runs.

Conclusions

Randomness in community assembly has also been observed in space-limited communities such as the biofilms on macroalgae (43) and in communities of coral reef fish (44), where genetically distinct communities that perform similar metabolic functions are found in adjacent environments. It is hypothesized that random events determine the arrival of the first organism with a given metabolic function, resulting in highly divergent communities coexisting side by side. By analogy, each model run represents a different surface. However, there is no direct competition for space in the water column. Instead, niche space is hypothesized to be jointly developed by the interaction of physical factors and the metabolic activity of the members of the community that may act as ecosystem engineers (45). In the model, all simulations developed similar biogeochemical gradients that acted as niche space, indicating that the evolution of metabolic capacity was more important than any specific ecosystem engineer. Global studies of marine dispersal and evolution suggest that microbes disperse more slowly than they evolve, leading to formation of biogeographic provinces (46). Thus, a testable hypothesis emerging from this work is that the global pool of available metabolic functions, rather than the distribution of functions among organisms, drives community assembly and formation of biogeochemical gradients in the ocean.

REFERENCES AND NOTES

- J. P. Zehr, M. T. Mellon, S. Zani, *Appl. Environ. Microbiol.* **64**, 5067 (1998).
- O. Bèjà et al., *Science* **289**, 1902–1906 (2000).
- L. R. Moore, G. Rocap, S. W. Chisholm, *Nature* **393**, 464–467 (1998).
- R. M. Morris et al., *Nature* **420**, 806–810 (2002).
- L. M. Proctor, J. A. Fuhrman, *Nature* **343**, 60–62 (1990).
- D. C. Reed, C. K. Algar, J. A. Huber, G. J. Dick, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1879–1884 (2014).
- M. J. Follows, S. Dutkiewicz, S. Grant, S. W. Chisholm, *Science* **315**, 1843–1846 (2007).
- P. G. Falkowski, T. Fenchel, E. F. Delong, *Science* **320**, 1034–1039 (2008).
- E. Marañón, *Annu. Rev. Mar. Sci.* **7**, 241–264 (2015).
- K. F. Edwards, M. K. Thomas, C. A. Klausmeier, E. Litchman, *Limnol. Oceanogr.* **57**, 554–566 (2012).
- A. M. P. McDonnell, K. O. Buesseler, *Limnol. Oceanogr.* **55**, 2085–2096 (2010).
- S. J. Giovannoni et al., *Science* **309**, 1242–1245 (2005).
- J. J. Morris, R. E. Lenski, E. R. Zinser, *MBio* **3**, e00036-12 (2012).
- R. E. Lenski, *Int. Microbiol.* **1**, 265–270 (1998).

- B. K. Swan et al., *Science* **333**, 1296–1300 (2011).
- S. J. Giovannoni, J. Cameron Thrash, B. Temperton, *ISME J.* **8**, 1553–1565 (2014).
- B. M. Satinsky et al., *ISME J.* **11**, 1852–1864 (2017).
- B. M. Satinsky et al., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11085–11090 (2014).
- B. L. Zielinski et al., *PLOS ONE* **11**, e0160929 (2016).
- V. Martin-Jézéquel, M. Hildebrand, M. A. Brzezinski, *J. Phycol.* **36**, 821–840 (2000).
- F. Azam, *Science* **280**, 694–696 (1998).
- J. I. Goes et al., *Prog. Oceanogr.* **120**, 29–40 (2013).
- N. Loick-Wilde et al., *Limnol. Oceanogr.* **61**, 460–481 (2015).
- S. C. Weber et al., *Limnol. Oceanogr.* **62**, 618–631 (2017).
- T. S. Kostadinov, D. A. Siegel, S. Maritorena, *Biogeosciences* **7**, 3239–3257 (2010).
- M. van der Does, L. F. Korte, C. I. Munday, G. J. A. Brummer, J. B. W. Stuut, *Atmos. Chem. Phys.* **16**, 13697–13710 (2016).
- R. R. Hood, V. J. Coles, *J. Geophys. Res.* **109**, C06006 (2004).
- W. J. Pasciak, J. Gavis, *Limnol. Oceanogr.* **19**, 881–888 (1974).
- A. A. L. Waite, P. A. Thompson, P. J. Harrison, *Limnol. Oceanogr.* **37**, 468–477 (1992).
- A. Waite, A. Fisher, P. A. Thompson, P. J. Harrison, *Mar. Ecol. Prog. Ser.* **157**, 97–108 (1997).
- J. R. Taylor, R. Stocker, *Science* **338**, 675–679 (2012).
- T. A. Villareal et al., *Nature* **397**, 423–425 (1999).
- H. Agogue, M. Brink, J. Dinasquet, G. J. Herndl, *Nature* **456**, 788–791 (2008).
- J. M. Smith, F. P. Chavez, C. A. Francis, *PLOS ONE* **9**, e018173 (2014).
- V. J. Coles, R. R. Hood, M. Pascual, D. G. Capone, *J. Geophys. Res.* **109**, C06007 (2004).
- C. Yoshikawa, V. J. Coles, R. R. Hood, D. G. D. G. Capone, N. Yoshida, *J. Geophys. Res. Oceans* **118**, 2520–2534 (2013).
- M. R. Stukel, V. J. Coles, M. T. Brooks, R. R. Hood, *Biogeochemistry* **10**, 13931–13976 (2013).
- S. M. Gifford, S. Sharma, J. M. Rinta-Kanto, M. A. Moran, *ISME J.* **5**, 461–472 (2011).
- V. J. Coles et al., *J. Geophys. Res. Oceans* **118**, 6894–6913 (2013).
- A. Subramaniam et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10460–10465 (2008).
- V. J. Coles, R. R. Hood, in *Aquatic Microbial Ecology and Biogeochemistry: A Dual Perspective*, P. M. Gilbert, T. M. Kana, Eds. (Springer International Publishing, 2016), pp. 45–63.
- M. Doherty et al., *Front. Microbiol.* **8**, 882 (2017).
- C. Burke, P. Steinberg, D. Rusch, S. Kjelleberg, T. Thomas, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14288–14293 (2011).
- P. F. Sale, *Environ. Biol. Fishes* **3**, 85–102 (1978).
- D. R. Nemergut et al., *Microbiol. Mol. Biol. Rev.* **77**, 342–356 (2013).
- F. L. Hellweger, E. van Sebille, N. D. Fredrick, *Science* **345**, 1346–1349 (2014).

ACKNOWLEDGMENTS

Sequences from the June 2010 study are available from the National Center for Biotechnology Information under accession numbers SRP037334 (metagenomes), SRP037995 (nonselective metatranscriptomes), and SRP039544 (polyadenylate-selected metatranscriptomes). The model computer code is available at <http://sourceforge.net/projects/roca-model/?source=directory>. Atlantic Meridional Transect Consortium nutrient data (NER/0/5/2001/00680) were provided by the British Oceanographic Data Centre and supported by the Natural Environment Research Council. Particle concentration fields were provided by T. Kostadinov. This research was funded by The Gordon and Betty Moore Foundation through the “River Ocean Continuum of the Amazon” project (ROCA; grants GBMF2293 and GBMF2928). Ship time was provided by the National Science Foundation (grant OCE-0933975). V.C. carried out the modeling effort in collaboration with R.H., M.S., M.A.M., M.B., and A.B. “Omics” were measured and analyzed by M.A.M., J.P., B.S., B.Z., and B.C., all of whom contributed to the conceptual model development. P.Y. led the ANACONDAS (“Amazon influence on the Atlantic: CarbOn export from Nitrogen fixation by DiAtom Symbioses”) /ROCA teams. This is UMCES contribution 5435.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/358/6367/1149/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S4
References (47–60)

3 May 2017; accepted 19 October 2017
10.1126/science.aan5712

Ocean biogeochemistry modeled with emergent trait-based genomics

V. J. Coles, M. R. Stukel, M. T. Brooks, A. Burd, B. C. Crump, M. A. Moran, J. H. Paul, B. M. Satinsky, P. L. Yager, B. L. Zielinski and R. R. Hood

Science **358** (6367), 1149-1154.
DOI: 10.1126/science.aan5712

Functional ocean biogeography

Marine ecosystems are well represented in metagenomic and transcriptomic data. These data are not routinely used to test ecosystem models that explore ocean biogeography or biogeochemistry. Coles *et al.* built a model in which genes for a range of functions were assigned to different suites of simulated microbes (see the Perspective by Rynearson). Communities emerged from the model with realistic biogeographical and biogeochemical profiles when compared to microbial data collected from the Amazon River plume. However, functional composition trumped the details of taxonomy, and different, coevolving community compositions emerged that provided similar biogeochemical outcomes.

Science, this issue p. 1149; see also p. 1129

ARTICLE TOOLS	http://science.sciencemag.org/content/358/6367/1149
RELATED CONTENT	http://science.sciencemag.org/content/sci/358/6367/1129.full
REFERENCES	This article cites 56 articles, 17 of which you can access for free http://science.sciencemag.org/content/358/6367/1149#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the [Terms of Service](#)