

Reviewed Preprint

v1 • August 28, 2026

Not revised

✉ For correspondence:

sullivan.948@osu.edu

damien.eveillard@univ-nantes.fr

† these authors contributed equally to this work

Competing interests: SJH has a conflict of interest; he is a co-founder of Koonkie Inc., a bioinformatics consulting company that designs and provides scalable algorithmic and data analytics solutions in the cloud.

Reviewing editor: Ferdi Hellweger, Technische Universität Berlin, Germany

© 2026, Régimbeau et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Planetary-scale heterotrophic microbial community modeling assesses metabolic synergy and viral impacts

Antoine Régimbeau^{1,†}, Funing Tian^{2,3,†}, Garrett J Smith^{2,3}, V James Riddell^{2,3,4}, Constanza M Andreani-Gerard^{5,13}, Philippe Bordron^{6,7}, Marko Budinich¹⁵, Cristina Howard-Varona^{2,3}, Abdelhalim Larhlami¹, Enrico Ser-Giacomi⁸, Camille Trottier⁹, Lionel Guidi^{5,15}, Steven J Hallam^{10,11}, Daniele Iudicone^{12,15}, Eric Karsenti^{15,18}, Alejandro Maass^{13,15,16,17}, Matthew B Sullivan^{2,3,4,14,15} ✉, Damien Eveillard^{1,17} ✉

¹Nantes Université, Centrale Nantes, CNRS, LS2N, Nantes, France • ²Department of Microbiology, Ohio State University, Columbus, United States • ³Center of Microbiome Science, Ohio State University, Columbus, United States • ⁴EMERGE Biology Integration Institute, Ohio State University, Columbus, United States • ⁵Sorbonne University, CNRS, Laboratoire d'Océanographie de Villefranche, Villefranche-sur-Mer, France • ⁶Université Fédérale de Toulouse, INRAE, MIAT, Castanet Tolosan, France • ⁷Université Fédérale de Toulouse, INRAE, BioinfOmics, GenoToul Bioinformatics facility, Castanet-Tolosan, France • ⁸IFISC (CSIC-UIB), Institute for Cross-Disciplinary Physics and Complex Systems, Palma de Mallorca, Spain • ⁹Nantes Université, Inserm, TENS, The Enteric Nervous System in Gut and Brain Diseases, IMAD, Nantes, France • ¹⁰Department of Microbiology & Immunology, University of British Columbia, Vancouver, Canada • ¹¹Life Sciences Institute, University of British Columbia, Vancouver, Canada • ¹²Stazione Zoologica Anton Dohrn, Napoli, Italy • ¹³Center for Mathematical Modeling, University of Chile and IRL 2807 CNRS, Santiago, Chile • ¹⁴Department of Civil, Environmental, and Geodetic Engineering, Ohio State University, Columbus, United States • ¹⁵Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/Tara GO-SEE, Paris, France • ¹⁶Department of Mathematical Engineering, University of Chile, Santiago, Chile • ¹⁷Millennium Institute Center for Genome Regulation, University of Chile, Santiago, Chile • ¹⁸Institut de Biologie de l'ENS, Département de Biologie (IBENS), École Normale supérieure, CNRS, INSERM, Université PSL, Paris, France

eLife Assessment

This study presents a **valuable** analysis of the TARA oceans dataset, advancing genome- and ecosystem-scale metabolic modeling. The authors present **convincing** evidence for the factors that shape microbial ecology at the molecular scale. The work will be of broad interest to oceanographers and scientists in other areas where metabolic interactions in diverse communities are of importance.

<https://doi.org/10.7554/eLife.111488.1.sa2>

Abstract

The oceans buffer against climate change via biogeochemical cycles underpinned by microbial metabolic activities. While planetary-scale surveys provide baseline microbiome data, inferring metabolic and biogeochemical impacts remains challenging. Genome-scale modeling has addressed analogous issues at the cellular level, highlighting key metabolic reactions contingent upon specific 'environmental' conditions. Here we adapt this mechanistic modeling framework towards analyzing global ocean microbial communities to reveal metabolic processes predicted to maintain ecosystem functioning. To achieve this, we developed a genome-scale 'superorganism' metabolic model for each TARA Ocean metagenome or metatranscriptome (i.e., limited to reactions known from heterotrophic prokaryotes and viruses), and evaluated these models to

establish a community-wide ‘metabolic phenotype’ for each sample. To validate, we showed that even with reaction-mappable genes only (~1/4 of the total genes), model composition revealed metabolism-inferred ecological zones that matched taxonomy-inferred zones. Model inferred metabolic phenotypes revealed reaction cooperation associated with microbial metabolism and organism diversity. These phenotypes also suggest elevated ecological roles for viruses as model predictions suggest they genomically target community-critical metabolic reactions that underpin metabolic phenotype stability, and also demonstrate that, as metabolites are better understood, immediate estimates could be made for where viruses remineralize versus sink carbon. While this new constraints-based, agile, and mechanistic modeling framework is highly upgradable, it already begins to convert molecular-scale environmental omics data to ecological and even planetary-scale biogeochemical features that will better bring microbes and their viruses into Earth system and climate models.

Connecting “omic” data (DNA, RNA, protein, and metabolites) to ecosystem-level nutrient fluxes and energy flows is a grand challenge in the life sciences^{1–3}. Leveraging petascale microbiome data, a fast-growing microbial metabolic knowledgebase^{4,5}, and matched physical and chemical information provides opportunities to develop constrained models to interrogate active metabolic processes. However, this will require new modeling approaches as current options cannot yet simultaneously model myriad taxa across the many orders of magnitude of biological organization (e.g., individual, population, and community levels) required to establish ecosystem-, regional- and even planetary-scale predictions.

Fortunately, advances in several areas offer promise as follows. *First*, the rapid growth in genome-resolved ‘-omics’ observations makes them increasingly available to parameterize large-scale biogeochemical models⁶ and simulate ocean ecosystem behavior⁷. *Second*, community-level gene expression data (metatranscriptomes) can now be used to identify transcribed genes that predict regional ocean response patterns⁸. *Third*, metabolic mapping⁹ and graph-based^{10,11} approaches – even driven simply by reaction presence or absence – are increasingly used to analyze compositional data¹² or semi-quantitative read counts for inferring organismal or multi-organismal metabolic phenotypes^{13,14}. *Finally*, metabolic modeling has started using gene and reaction lists to tackle ecosystem-level questions¹⁵ where models were effective for assessing mutualistic interactions¹⁶, distinguishing competitive or cooperative communities^{17,18}, and exploring trade-offs between biological objectives that biological systems aim to maximize (e.g., growth, energy storage or consumption) for each member of the community¹⁹. Currently, the many scales and approaches in modeling offer tremendous promise for opening new windows into biology, but leave benchmarking and validation requirements highly varied and often incredibly challenging, if not impossible, at scales out of sync with observations²⁰.

Here, we sought to build upon these efforts by using metabolic modeling, in particular Genome-Scale Models (GSMs), to mechanistically connect microbial metabolic abilities to biogeochemical processes at ecosystem scales. To build and study these GSMs, we chose the superorganism approach^{21,22}, whereby all genes in a given sample are considered part of one “community cell”. This simplification was chosen because any taxonomy-constrained modeling framework suffers from uncertainties about how to model interacting species²³, the low quantity of prokaryotic reads recovered by Metagenome Assembled Genomes (i.e., from 14 to 19%²⁴), and computational processing limitations. From this superorganism choice, metagenomic and metatranscriptomic samples represent the gene content or gene expression, respectively, of a superorganism that was then used to reconstruct metabolic networks via a standard top-down approach²⁵ that reduces a “universal” metabolic network by removing reactions that lack genetic evidence (see **Methods** for details). Application to *Tara* Oceans data^{8,26} that spanned multiple depths from 62 stations sampled throughout the global ocean (127 prokaryote-enriched metagenomes paired with 94 prokaryote-enriched metatranscriptomes and informed by virus data from 93 virus-enriched metagenomes; Fig. 1a [↗](#), see **Methods**, and **Supplementary Text 1**) resulted in 221 *Tara* Oceans Genome-Scale Metabolic Models (herein TOGSMs)..

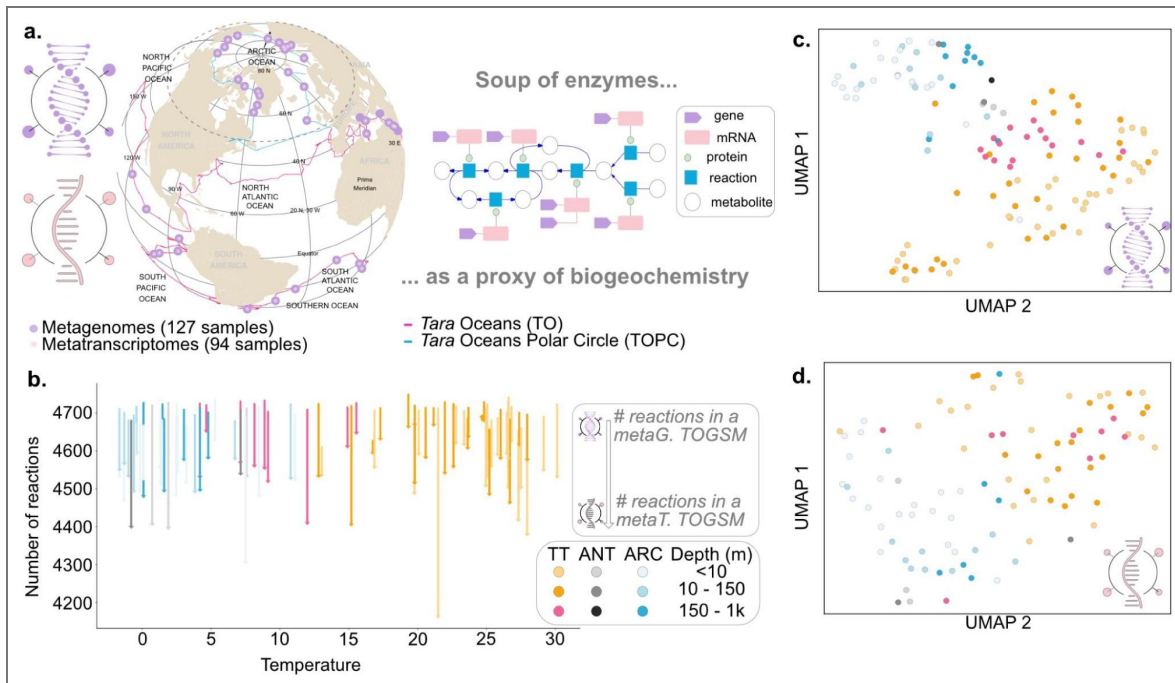


Figure 1. Construction and ecological exploration of TOGSMs.

a. Graphical abstract of genome-scale metabolic model reconstruction applied on globally distributed metagenomes (n=127) and metatranscriptomes (n=94) (left). For this, known physics and chemistry of metabolic reactions are captured into metabolic networks (schematic right, where genes are depicted in purple, enzymes in pink, and metabolic reactions in blue that transform metabolites in white). **b.** The number of reactions for each metabolic model was reconstructed per sample from its metagenome, and the matching metatranscriptome was plotted against temperature to assess potential bias due to the latitudinal gradient. In all cases, metagenome-based model are larger (contains more reactions) than their corresponding metatranscriptome-based model, which was interpreted as the per-sample genetic potential (metagenome) versus genetic expression (metatranscriptome) sampled functional repertoires, respectively. Color coding represents previous ecological zones defined by microbial and virus community data^{27,26} (TT = Temperate and Tropical, ANT = Antarctic, ARC = Arctic), and the degree of color represents sampling depth as indicated in the figure. **c** and **d.** Ordination (UMAP, or Uniform Manifold Approximation and Projection) of per-sample presence/absence reaction predictions show that genome-scale metabolic models generated from metagenomes (c) and metatranscriptomes (d) structure along known ecological zones. Colors are the same as **b**.

Even though community metabolic measurements are impossible at scale, we sought to validate this approach by assessing whether TOGSMMs could recover known biological and ecological features. *First*, we observed that per-sample metagenome-constructed TOGSMMs contained more reactions with genetic evidence than metatranscriptome-constructed TOGSMMs, with an average of 4,058 (+/- 28 stdev.) and 3,900 (+/- 79 stdev.) reactions, respectively (Fig. 1b [↗](#) and **Supplementary Text 2** for further discussion). These findings are consistent with a community's expressed transcripts sampled once (i.e., the metatranscriptome) representing only the active subset of genes in the fuller genomic repertoire (i.e., the metagenome). *Second*, despite only a subset of genetic data being incorporated into the model (26% of the total genes observed were mappable to metabolic reactions), TOGSMM-derived ordinations (Fig. 1c and 1d [↗](#)) sorted samples into the same global ocean ecological zones known from taxonomy^{26,28,29} or genome content (see **Supplementary Text 2**). Though only a first-level positive sign, this comfortingly showed that the modeled superorganism metabolism indeed follows broader metagenome-inferred ecological patterns. As proposed in the previous *Tara* Ocean analysis⁸, despite not considering autotrophic and eukaryotic communities, the modeled heterotrophic prokaryotic communities reflect a meta-genomic and meta-transcriptomic composition associated with distinct abiotic and biotic environments.

Given the above models' composition validations, we sought to explore how such modeling could improve our understanding of ocean microbial communities via the identification of crucial metabolic processes necessary for the maintenance of microbial ecosystems. To do this, we sought to provide a mathematical representation for each sample's observed sequence-related community-wide heterotrophic metabolism, which we term a sample's 'metabolic phenotype'. Problematically, solving TOGSMMs would require computing all mathematically feasible reaction fluxes that satisfy the modeling constraints (Fig. 2a [↗](#)), which, considering more than four thousand reactions per sample, would result in a highly multidimensional solution space (its dimensions are equal to the number of reactions present in the GSM) that is impossible to solve exhaustively. Instead, we approximated this solution space through metabolic flux sampling analysis³¹ that performs random walks (we conducted 100,000 per sample) through the solution space (Fig. 2a [↗](#), see **Methods** for details).

From these empirical solution space estimates of each TOGSMM, we calculated two metrics. *First*, for each reaction in a TOGSMM, we calculated its Reaction Cumulative Correlation (RCC). This well-known metric³² estimates each reaction's ability to quantitatively and mechanistically impact the overall metabolic reaction network and is calculated as the sum of a reaction's pairwise flux correlations against all other reactions (see **Methods**)³². Biologically, this metric has been used in analyses of cancer metabolic models to identify therapeutic targets that destabilize the cancer cell's metabolic networks and functioning³², and we recently demonstrated that high RCC reactions align with central genes in a diatom's gene co-expression network³⁰. *Second*, we compared RCC density distributions between each TOGSMM (Supplementary Fig. 14 [↗](#)) to compute a community-wide per-sample 'synergy rank' score (see **Methods**). This score, derived from the comparison of RCC distributions, investigates a phenotype that cannot be derived from metagenomic comparisons. We interpret this score to represent the biochemical interdependence of the modeled superorganism such that enrichment for low RCC reactions lowers the synergy rank (i.e., closer to zero) with the implication that few reactions control the system, and reaction fluxes are more independent from others as compared to samples of higher synergy rank. This score thus evaluates mechanistic independencies between metabolic reactions under the thermodynamic and redox constraints. It corresponds to an emergent trait that we posit could represent (loosely here due to the current superorganism-assumption requirement) the level of community-wide metabolic reaction synergy and/or cooperativity (Fig. 2a [↗](#)). Together these new metrics help describe each seawater sample's metabolic phenotype, which we in turn used to study as modeled biogeochemistry processes occurring in each sample as follows.

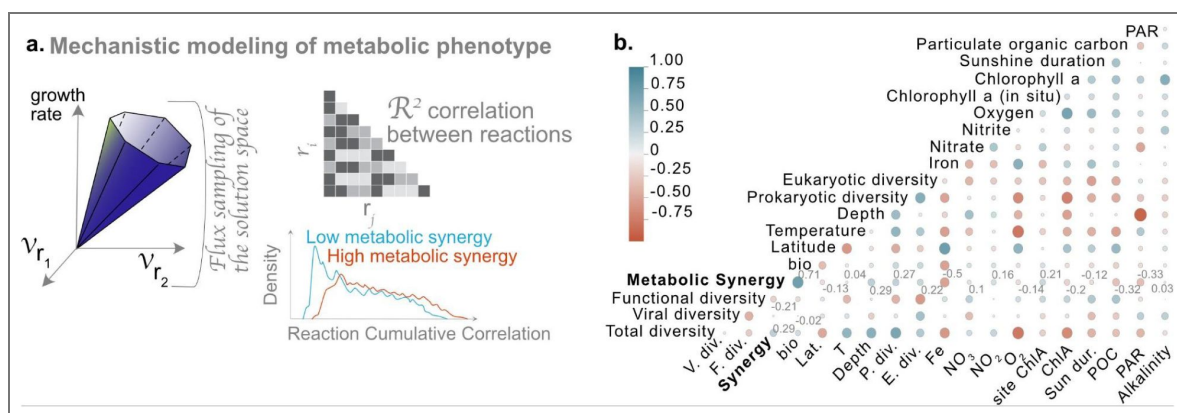


Figure 2. Computing metabolic synergy across the global ocean and comparing other parameters.

a. Graphical abstract to capture TOGSMM workflows. For each TOGSMM, the metabolic solution space reflects all mechanistic solutions for any environmental conditions (i.e., the fundamental metabolic niche of the community³⁰), which we suggest here represents a ‘metabolic phenotype’ for each sample. Because of the complexity of the mechanistic solutions, we summarize them by a flux sampling procedure that estimates each metabolic reaction’s flux distribution. To estimate the reaction’s strength for controlling the metabolic phenotype, pairwise correlations (all reactions vs all other reactions) were computed for each reaction within a network and summarized as a Reaction Cumulative Correlation (RCC) as done elsewhere³⁰. For a given network, the RCCs’ distribution determines its metabolic synergy, offering a quantitative score for comparing metabolic phenotypes. **b.** Correlogram of metabolic synergy and environmental variables to assess statistically significant correlations (Spearman correlation).

Metabolic synergy across the ocean

Starting from the community-wide synergy rank metric, we next explored how it related to metadata throughout the global oceans dataset (Fig. 2b). This revealed associations between synergy rank and abiotic factors such as iron concentrations (estimated from the NEMO-PISCES Earth System Model, Spearman r : -0.5, p -value: 2×10^{-5}), which are known to impact plankton communities³³ and, now, our findings would suggest also impact per-sample metabolic phenotypes. Alpha diversity of mOTUs (i.e., a biodiversity taxonomic metric) was also significantly associated with synergy rank (Spearman r : 0.27, p -value: 1×10^{-3}). While the correlation is weak (e.g., putatively due to the lack of metabolic autotrophic system), we interpret it to inform the “diversity-stability debate”³⁴ whereby synergy rank indicates community metabolic network stability, such that in a high synergy system, high RCC reactions are more prone to perturbation-driven instability than in a low synergy system. If future work confirms this, synergy rank could be used by researchers and policymakers to evaluate ecosystem resilience and stability from the new mechanistic perspective of community-wide metabolism.

Per-sample synergy ranks also revealed biogeographical patterns (Fig. 3a and 3b, and **Supplementary Text 3**). For example, the metabolic phenotypes of most oligotrophic waters, such as the Mediterranean Sea and the Indian Ocean, had low synergy ranks. In contrast, those from nutrient-enriched water masses that follow bloom decay (e.g., Stations 76 and 138) or nutrient release via island effects (e.g., Stations 122 to 125) had high synergy ranks. Such patterns could emphasize a relationship between metabolic synergy and microbial productivity via nutrient availability³⁵. Indeed, standard GSM approaches³⁶ allow to compute a maximum theoretical growth rate associated with each TOGSMM which is the production of biomass evaluated through the flux through the biomass reaction (See **Methods**). This estimation revealed its correlation with metabolic synergy rank (Fig. 3c; Spearman r : 0.70, p -value: 10^{-14}), emphasizing the effect of a high synergetic system on theoretical productivity.

Beyond these global observations, we also explored local metabolic features, where higher-resolution sampling was available to test, at the metabolic level, the assumption that most Earth System Models make of a continuous transition between conditions (i.e., each point on the grid)³⁷. Specifically, we used Lagrangian connectivity between each TOGSMM (computed as done previously³⁸) to ‘connect’ samples to assess how metabolic phenotypes changed over short time scales (60 days or less; see edges in Fig. 3a and **Methods**). Though only one sample pairing could be tested (natural fertilization and island effect near the Marquesas islands³³), it revealed an uneven shift across the synergy rank gradient rather than seeing a continuous transition from an more autonomous state to a more synergetic state (or vice versa) as a water mass is transported (Fig. 3a and **Supplementary Fig. 6**). As more high-resolution sample sets become available, it will be ideal to assess their metabolic phenotypes and synergetic status across representative water masses and environmental gradients to better guide ESM assumptions.

Considering another context, we asked whether summed RCCs for biogeochemically relevant reactions (pathways in the KEGG database³⁹ for carbon, nitrogen, and phosphorus) offered functional patterns of interest. Such estimations would indicate the relative mechanistic importance of each biogeochemical cycle to sustain the water mass metabolic phenotype. While the summed RCC of each cycle scales with synergy rank (Fig. 3c and **Supplementary Text 4.c** for discussion on the statistical validity of this correlation), we also observed a remarkably constant proportional summed RCC (std < 0.1 of the mean value for the cycles C:N, C:P and P:N). These observations require further explorations and improved underlying reaction-to-pathway assignment data (**Supplementary Text 4.a**). Still, they could mechanistically contribute to the debate about elemental ratios in the ocean’s plankton and seawater⁴⁰.

Finally, we evaluated the use of expressed transcripts (grouped into 8,803 KO⁸ categories) against the synergy rank across large ecological gradients (synergy ranks that shifted from positive to negative between polar and temperate subsystems) to assess possible transcript-based biomarkers for the synergy rank metabolic phenotype. Indeed, some KOs showed significant correlations (**Supplementary Fig. 13**) with expected functions involved in core metabolism (i.e.,

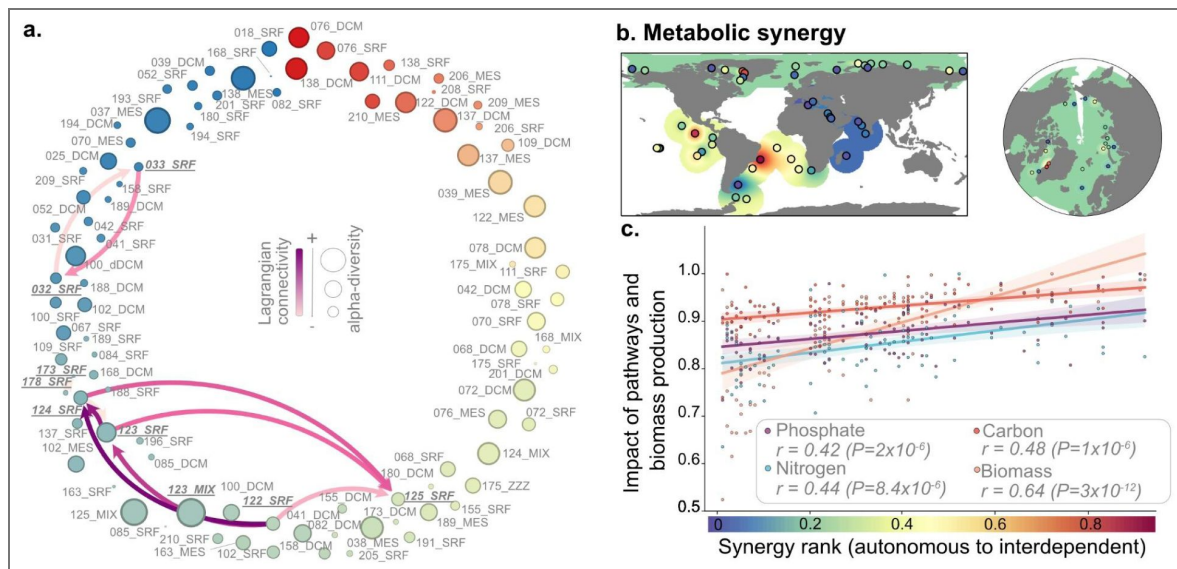


Figure 3. Metabolic synergy across the global ocean.

a. Distribution of *Tara* samples based on metabolic reaction synergy from autonomous (blue) to interdependent (red) systems. The node size is proportional to the per-sample prokaryotic organism diversity. Directed edges represent Lagrangian connections between samples realized in less than two months to describe the potential acclimation of water masses. The darker the arrow's color is, the stronger the Lagrangian connection is. In other words, the darker the arrow is, the higher the connection probability between the water masses is. **b.** Projection on the metabolic synergy across the global ocean visualized from an equator- (left) and pole- (right) centric perspectives. The data points were interpolated at each station (halo color on each map), and actual data values are the colored dots. **c.** Relationships between the metabolic synergy of the microbial ecosystem against the normalized impact of phosphate (purple), nitrogen (blue), and carbon (red) reactions and the potential prokaryotic ecosystem biomass production (orange).

oxidoreductase activity) or membrane activity (i.e., transferase, molecular transducer, membrane transport, and transmembrane transporter activity). However, surprisingly, this also revealed functions including hydrolases and lyases, which are commonly encoded by viruses. We interpreted this latter signal to suggest that viruses impact community-level metabolic phenotypes and sought to assess this signal further by shifting to more virus-targeted assessments.

Virus-modulated metabolisms and shunt/shuttle biogeography

To explore the role of viruses in modulating community-level metabolic phenotypes, we looked at the presence of conservatively-identified, virus-encoded AMGs⁴² using viral metagenomes within TOGSMMs. Strikingly, this revealed that virus-encoded AMGs do not randomly target reactions, but instead predominantly target the highest RCC reactions needed to sustain the community metabolic phenotype (Fig. 4a⁴³, and Supplementary Fig. 14⁴³). Biologically, viruses are known to obtain host genes at random through error-prone genomic replication and individual-level selection that genomically fixes those that improve viral fitness⁴³. However, these findings show that this evolutionarily selective process also manifests at a community-level such that viruses target high RCC reactions. These biochemical reactions demonstrate topological centrality within metabolic networks and enable fluxes that regulate flux distribution across the entire range of reactions. Consequently, they are pivotal to the overarching metabolic phenotype of the community and serve as a foundation for its stability, predicting that a perturbation on their flux would drastically impact the whole.

Exploring this further functionally, we found that virus-encoded purine and pyrimidine AMGs – already known to be enriched in viral metagenomes^{44,45} and differentially expressed in marine virocell experiments⁴⁶ – were conserved in all TOGSMMs, but were among the highest RCC reactions in TOGSMMs specific to different ecological zones (Fig. 4b⁴³). Presumably, this reflects their critical role in the energy demand of viral replication⁴⁷, but now TOGSMM-contextualization suggests nucleotide-synthesis AMGs also significantly impact the community-level metabolic phenotypes. Other energy-related AMGs targeted fermentation, glycolysis, gluconeogenesis, the pentose phosphate pathway, and two-carbon compound reactions. Presumably, such AMG-impacted metabolisms arise from varied viral host cell reprogramming responses across gradients of nutrient-impacted host physiological states requiring diverse energy resources to drive the viral life cycle. Again, while these kinds of functions are characteristic from virocell experiments^{42,43} and beginning to be explored under *in situ* conditions^{48,49}, TOGSMM-contextualization now places them into a mechanistic metabolic modeling framework that intertwines cellular systems-level functionality with ocean basin and even global ecological context. For example, AMG-targeted amino acid metabolism genes varied with low-energy cost amino acids (glycine, serine, threonine, alanine, aspartate and glutamate)⁵⁰ targeted broadly, higher energy cost amino acids (cysteine and methionine)⁵⁰ were exclusively targeted in the Arctic Ocean (Fig. 4b⁴³). Given that cysteine and methionine are involved in sulfur metabolism and sulfur-related AMGs are distributed globally in the ocean⁵¹, their polar enrichment suggests polar-specific impact of viruses on the salvage of amino acids for nutrient acquisition. Further, AMG reactions were significantly more abundant and associated with higher RCC in polar systems (Fig. 4c⁴³), which suggests a stronger impact of viruses on polar community metabolic phenotypes⁵². Together, such TOGSMM-powered findings highlight critical metabolic pathways for viral replication that can help constrain virocell studies and metabolic models of viral infection while also placing them into a broader biogeochemical and biogeographical context.

Finally, we wondered whether we could use TOGSMMs to predict virus ecosystem “shunt versus shuttle” effects, which is of collective interest⁴¹ due to the ocean’s climate buffering role. For a quarter-century, the paradigm was that viruses help heterotrophs rapidly remineralize carbon by making it more available via cell lysis – i.e., the viral shunt hypothesis⁵³. However, aggregating lysis products^{54,55} and viral roles in carbon cycling beyond lysis (e.g., AMGs and metabolic reprogramming) have emerged as additional contributors to biogeochemical cycling⁵⁶. At least in

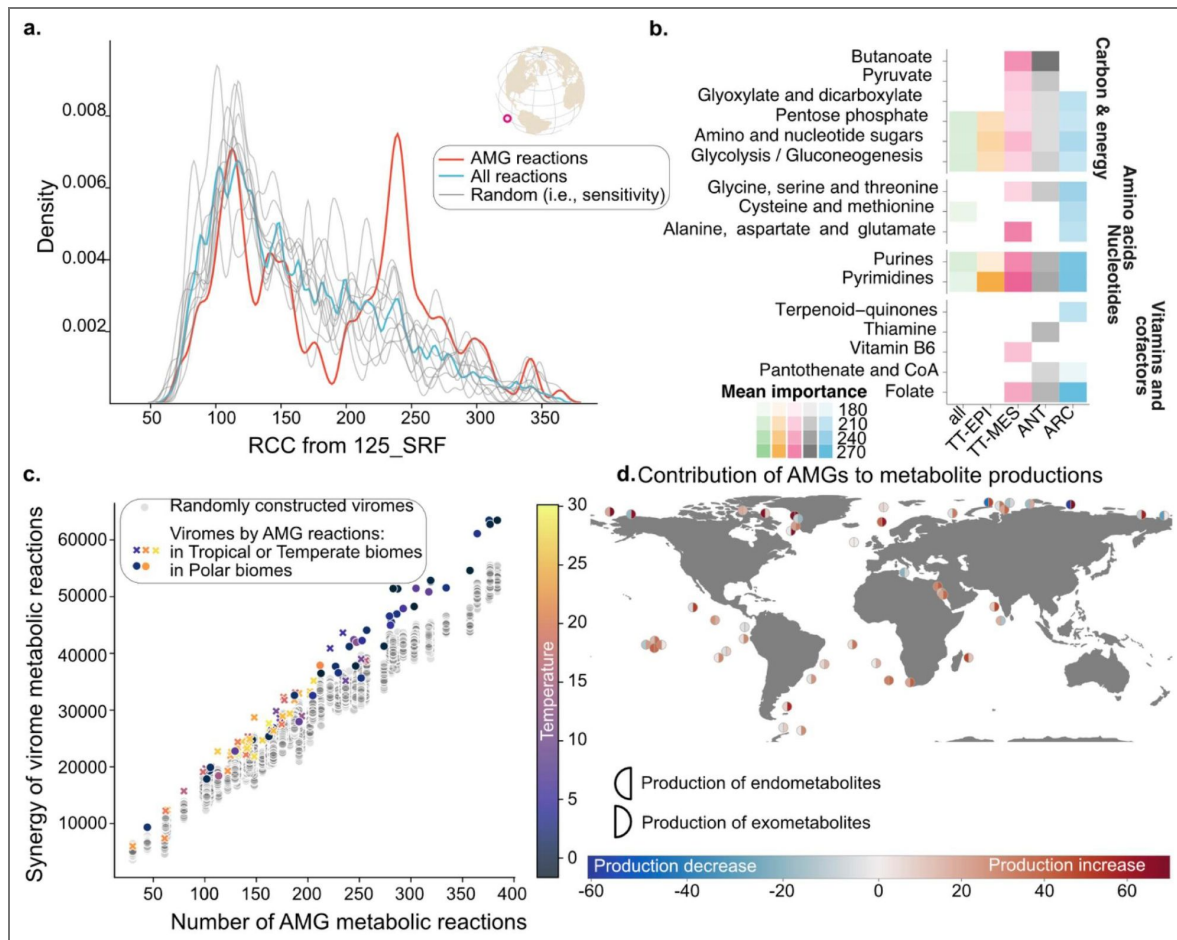


Figure 4. RCC and impact of reactions targeted by virus-encoded Auxiliary Metabolic Genes (AMGs).

a. Representative example from sample 125_SRF of the RCC of all (blue) and randomly sampled (gray, see **Materials and Methods**) metabolic reactions as compared to the reactions targeted by virus-encoded AMGs (red). All samples showed a similar, statistically significant high RCC AMG spike (see Supplementary Fig. 14). **b.** Metabolic pathways associated with high RCC AMG-targeted reactions across ecological zones. The RCC of reactions targeted by AMGs, present in all networks and statistically belonging to the high RCC spike, were averaged per KEGG pathway (i.e., lines) and ecological zone (i.e., columns). RCCs are then displayed by opacity. Note that a single reaction may be shared among several KEGG pathways. **c.** Comparison between the summed RCC of reactions targeted by AMGs and the number of reactions per sample. Each dot represents a sample. Its color code indicates the temperature associated, and its shape characterizes the system: a circle for polar systems and a cross for temperate systems. Grey dots represent the summed RCC of randomly defined viromes. **d.** Contribution of AMGs to exometabolites and endometabolites through increased (i.e., red) or decreased (i.e., blue) metabolite production. Here we consider metabolites that are only exometabolites ($n=26$), or only endometabolites ($n=28$)⁴¹.

some instances, viruses are now thought to ‘shuttle’ carbon to the deep sea via sinking aggregated cells and/or altering grazing efficiency^{48,53,57,58}. Accordingly, previous *Tara* Oceans work using statistical taxon-based modeling suggested that virus abundances, more than those of prokaryotes or eukaryotes, best predict key biogeochemical processes like carbon flux⁵².

Expanding upon these prior findings, we used TOGSMMs to explore per-sample shunt-versus-shuttle emphasis interpreted through the lens of the relative contribution of AMGs previously categorized as exo- or endo-metabolites⁴¹. For this analysis, we assumed that (i) exometabolites represent sticky materials that increase carbon export via aggregation invoked in the shuttle hypothesis⁵⁹ and (ii) endometabolites, as intracellular metabolites, are instead linked to community-level reprogramming and, thus, the remineralization-related shunt hypothesis. These are strong and highly over-simplified assumptions due to poorly characterized seawater metabolites⁴¹ and the need to more broadly characterize metabolic reactions into shunt-versus-shuttle effects by better assessing biological products’ stickiness⁵⁹ at scale. Visualizing epipelagic layer AMG reactions revealed estimates per-sample shunt-vs-shuttle proportions and possible biogeographic hot-spots (Fig. 4d⁴², and Supplementary Fig. 15⁴² for details), where this currently imperfect estimator (due to the myriad input data issues) already revealed an overall larger ‘shuttle’ contribution consistent with the oceans currently being a carbon sink via particle aggregation⁵⁹. Even given these assumptions, our mechanistically modeled exometabolite predictions weakly but significantly correlate to Particulate Organic Carbon (Spearman r : 0.34, p -val 4×10^{-3}). This suggests promise for global evaluation once better input data arise in the form of an ideal ‘metabolite stickiness’ list, a resource the genome-scale modeling community could use immediately to improve predictions.

Conclusions

This study introduces a mechanistic modeling approach to abstract community-wide heterotrophic bacterial metabolism from ‘omics data, and uses a statistical approach to explore the resultant per-sample metabolic phenotypes. Just as genome-scale models have revealed at a cellular level how biological systems self-organize to reveal metabolic linchpins and emergent cellular properties⁶⁰, our approach reveals how, from a given seawater sample’s metabolic capacity, one can establish ‘metabolic phenotype’ throughout the global oceans, and metabolic reaction essentiality for each sample’s community metabolism. This found that biotic and abiotic factors can influence variation in global ocean metabolic phenotypes, while also providing new hypotheses about viruses underpinning community-wide metabolic stability and a framework for assessing their roles in the planet-critical ocean biological carbon pump. Future developments will have to incorporate autotrophic reactions, ‘omics abundances, and/or genomic content refined from MAGs to broaden the set of research questions that can be addressed, but this new mechanistic modeling framework already offers an omics-powered approach to help transform our understanding of how microbiomes and viromes drive ecosystem functioning.

Materials and Methods

The global ocean dataset

The dataset for metabolic network reconstruction consists of paired prokaryote- ($n = 127$) and virus-enriched ($n = 127$) metagenomes and the corresponding prokaryotic metatranscriptomes ($n = 94$) from previously published datasets from *Tara* Oceans (TO) and *Tara* Oceans Polar Circle (TOPC) expeditions^{8,26}. These samples were collected from 62 sampling stations, spanning the global ocean, including the South Pacific Ocean, North Pacific Ocean, South Atlantic Ocean, Indian Ocean, Red Sea, Mediterranean Sea, Southern Ocean, North Atlantic, and Arctic Ocean.

Generation of the reference gene catalog

Representative coding sequence (cds) or gene features were selected for prokaryotic microbial and viral fractions. In the case of prokaryotic cds features, the Ocean Microbial Reference Gene Catalog (OM-RGC.v2)⁸, which included 46,775,174 features, was used as the reference catalog. In the case of

viral cds features, genes were called from assemblies in virus-enriched metagenomes ($n = 127$) using Prodigal (v2.6.1) in the metagenomic mode⁶¹. The resulting genes were grouped into clusters using mmseqs2⁶² (easy-cluster, -s 7.5, e-value 0.001) at two sets of parameters (30% ANI across 70% of the length and 60% ANI across 80% of the length). To establish functional relevance, 1000 gene clusters, with more than 20 genes per cluster, were randomly selected for each set of parameters. For functional annotation, the cds features within each cluster were searched against the RefSeq database (v208). Ambiguous clusters were further assessed using the Pfam database⁶³ (v33.1). Collectively, this analysis identified representative cds features ($n = 70,347,973$) clustered at 30% ANI across 70% of the length as the reference catalog for subsequent modeling efforts.

Calculation of metagenomic and metatranscriptomic profiles

We profiled metagenomic and metatranscriptomic cds features for use in constructing TOGSMMs. The gene and transcript abundance calculations of OM-RGC.v2 from the prokaryotic metagenomes ($n = 127$) and metatranscriptomes ($n=94$) were previously reported⁸ and used in this study. In total, 26% of the 47M cds features in OM-RGC.v2 were mappable to metabolic reactions in the BiGG database⁶⁴ (v1.6) and were thus considered in TOGSMMs. To calculate gene abundance from viral metagenomes, amino acid sequences of representative cds features were first searched against the BiGG database⁶⁴ (v1.6) using DIAMOND⁶⁵ (v0.9.24.125) (–more-sensitive, e-value 0.001). This resulted in 7% of the 70M cds features ($n = 4,939,042$) matched to reactions in the BiGG database⁶⁴ (v1.6). Quality-controlled reads of viral metagenomes²⁶ ($n = 127$) were then non-deterministically recruited to the 5M cds features using CoverM (v0.2.0) (<https://github.com/wwood/CoverM>) with the trimmed mean method. Reads were kept if they were with $\geq 95\%$ identity and $\geq 75\%$ read coverage. Genes with $< 70\%$ coverage within a sample were not considered. Similar strategies were used to calculate the metatranscriptomic composition and map prokaryotic metatranscriptomes ($n = 94$) to the cds features in the viral metagenomes.

Building a Tara Ocean Genome-Scale Metabolic Model

Metabolic networks were generated per paired samples. For each paired sample, genes with an abundance > 0 calculated from metagenomes were piped through CarveME²⁵ (v1.5.1) using a universal model of bacterial metabolism. The resulting metabolic networks ($n = 127$) were gap-filled using the marine media (Table S1). Similar strategies were applied to genes with transcriptomic abundance > 0 . These identified 94 networks derived from the transcriptomic abundance profile.

We used the classical constraints used in metabolic modeling to obtain the model. To each network, we associated a set of constraints:

$$\begin{cases} S v = 0 \\ lb \leq v \leq ub \end{cases}$$

where $S \in \mathbb{R}^{m,n}$ is the stoichiometric matrix of the network composed of m metabolites and n reactions, $v \in \mathbb{R}^n$ is the flux vector that represents the flux that goes into each reaction of the network, and $lb, ub \in \mathbb{R}^n$ are the lower and upper bounds of the flux vector (default values from the CarveMe algorithm). TOGSMMs are available here: <https://uncloud.univ-nantes.fr/index.php/s/9nXG6kRz66CwDAW> (a zenodo link and a DOI will be added upon acceptance of the manuscript).

Each TOGSMM is prone to Flux Balance Analysis. Under the superorganism assumption, the estimated solution of the heterotrophic prokaryotic growth objective function (the biomass reaction) is interpreted as the theoretical maximum growth rate at the ecosystem level. While not quantitatively accurate, estimating these theoretical maximum growth rates is sufficient to compare ecosystems built following the same procedure.

Auxiliary metabolic genes and their mapping in TOGSMMs

Systematically cataloged AMGs from the global ocean were used to assess the role of viruses in community-level metabolic phenotype. The Global Oceans AMG catalog was previously established. Briefly, AMGs were called on viral contigs with additional curation steps to exclude non-viral regions, mobile genetic elements, and other genes that may function in the random integration of microbial metabolic genes⁴². To map the Global Oceans AMG catalog ($n = 22,779$) in TOGSMMs, we searched them against the BiGG database (1.6) present in the CarveMe universe using DIAMOND (v2.0.14.152, –more-sensitive). We then applied the logical rules in CarveMe’s GPR to obtain the reactions AMG supports.

Sampling of the *Tara* Ocean Genome-Scale Models

Each metabolic model defines a solution space F , that is, the ensemble of solutions that satisfies the constraint:

$$\mathcal{F} := \{\mathbf{v} \in \mathbb{R}^n, \mathbf{S}\mathbf{v} = 0, \quad \mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub}\}$$

Describing the solution space of a metabolic model is equivalent to assessing its metabolic phenotype. To do so, we used optGpSampler³¹, which allows us to draw random flux distributions from the solution space. We constrain the biomass reaction to produce at least 10% of its maximum. This subjective threshold allows us to discard states with low biomass production, focusing our attention on states that assure the ecosystem’s survival.

We sampled 10^5 points with a thinning of 10^4 . From this sampling procedure, we computed a pairwise Pearson correlation between each network reaction, resulting in the correlation matrix $C \in R^{(n,n)}$.

Reaction Cumulative Correlation and cycles

Reaction Cumulative Correlation (RCC) is the sum of its pairwise Pearson correlation in absolute value over the entire network. Formally, we write the reaction’s j RCC as

$$\text{RCC}_j = \sum_{k=1}^n \|C_{j,k}\|$$

, where the sum is done over the set of reactions. A reaction with high RCC is significantly correlated with many other reactions in the network; thus, a change in its flux value will likely change the fluxes of many other reactions. On the other hand, a reaction with low RCC can see its flux vary without changing the overall flux distribution. In graph theory, when abstracting each reaction as a node and each correlation as a weighted edge between nodes, the RCC is called the “strength” of the node. It is equivalent to the centrality of its corresponding multigraph⁶⁶.

Cycle summed RCC is the sum of the reactions’ RCC that compose the cycle. We used KEGG identifiers available in the CarveMe reaction database to obtain the composition of cycles. To avoid subjective thresholds on correlation values, we used all correlations. The low correlations due to random effects appear in all the reactions and are uniformly distributed. As such, they do not impact our results.

Viruses impact the system through AMGs

AMGs encoded by viruses have been demonstrated to modify host metabolism sufficiently to alter host biogeochemical/metabolite flux profile⁶⁷. To incorporate viruses in our metabolic network, we considered AMG-targeted reactions in the established TOGSMMs. We estimated the impact of viruses through AMGs using the previously computed reactions’ summed RCC. To compute the impact of viruses on the production of metabolites, we computed the contribution of each reaction in the production or consumption of the considered metabolites. We then examined the impact of an increase in each reaction flux from the reactions supported by AMGs. This impact was computed by multiplying the participation of the reactions to the dynamic of the metabolites by their correlations to the reactions supported by AMGs.

Formally, for a given TOGSMM, this gives:

$$V_k = \sum_{i \in \mathcal{V}} \sum_{j \in [1..n]} S_{k,j} C_{i,j}$$

where $\mathcal{V}$ is the set of metabolic reactions supported by AMGs, $S_{k,j}$ is the stoichiometric coefficient of metabolite k in reaction j , and $C_{i,j}$ is the pairwise correlation between reaction i and reaction j .

To estimate random viromes for a given sample, we took the total number of reactions supported by the computed virome and randomly chose that many reactions in the network. This resulted in a random distribution of RCC and a random summed RCC score of reactions targeted by AMGs.

Stability of the ecosystem

In our framework, stability represents how sensitive the system is regarding a change of flux values. The correlation value from the sampling work allows us to assess this stability. A network with many reactions' correlations near zero is considered to represent a stable ecosystem because a change in a single reaction's flux will not significantly change the overall flux distribution and, thus, the estimated behavior of the ecosystem. However, a network with many highly correlated reactions will be more sensitive to changes in a given reaction's flux, and thus will represent a less stable ecosystem.

We used a Kolmogorov-Smirnov test (KS-test) on the absolute pairwise correlation values distribution to compute this metric. We calculated the KS distance and its associated p-value. We then created the directed graph, where edges were considered if the test p-value was not significant ($>5 \times 10^{-2}$), meaning that we could not reject the null hypothesis. An edge from station A to station B means that the ecosystem captured in station A is more stable than the one modeled in station B.

From this graph, one can compute each node's out-degree, which is the number of edges that start with the considered node. Once normalized, this out-degree corresponds to the sample's synergy score, as represented in Fig. 2 [\[40\]](#).

Endo and exo-metabolite identification and linking to reactions

Endo and exo-metabolites were determined using Table 1 of Moran et al 2022⁴¹. The metabolites table was joined to the reagents and products of each reaction used in the models to determine which reactions were directly associated with the transformation of endo and exometabolites.

Data availability

All data used in this paper is already available (-omics) and the models used will be released upon acceptance of the manuscript.

Acknowledgements

Our survey was made possible by the sampling and sequencing efforts of the *Tara* Oceans Project. We are indebted to all who contributed to these efforts and other open-source bioinformatics tools for their commitment to transparency and openness. *Tara* Oceans (which includes the *Tara* Oceans and *Tara* Oceans Polar Circle expeditions) would not exist without the leadership of the *Tara* Oceans Foundation and the continuous support of 23 institutes (<https://oceans.taraexpeditions.org/> [\[42\]](#)).

Additional information

Funding

We also acknowledge the commitment of the CNRS to support A.R. Ph.D. funds (CNRS 80 Prime), and the Simons Postdoctoral Fellowships in Marine Microbial Ecology. Some computations were performed thanks to France's BiRD and LIGER computing facilities. This study was partly

supported by ATLANTECO and Horizon Europe project BIOcean5D (award number 101059915). AM is funded by Projects ANID–MILENIO–ICN2021044, Exploración-13220002 and Basal-FB210005. E.S.-G. acknowledges support from grant PID2021-123352OB-C32 funded by MICIU/AEI/10.13039/501100011033 and FEDER, ‘Una manera de hacer Europa’. L.G and CM.A-G. are supported by Schmidt Sciences through the Virtual Earth System Research Institute project CALIPSO.

Author contributions

D.E designed the experiment. E.S.G computed the Lagrangian data. F.T reconstructed the models. A.R simulated the models. A.R, F.T and G.S analyzed the data. A.R, F.T, G.S, M.B.S and D.E wrote the initial manuscript. J.R, CM.A-G, P.B, M.P, C.H.V, A.L, E.S.G, C.T, L.G, S.J.H, D.I, E.K, and A.M help writing the final manuscript.

Author ORCID iDs

Alejandro Maass:  <https://orcid.org/0000-0002-7038-4527>

Damien Eveillard:  <https://orcid.org/0000-0002-8162-7360>

Additional files

[Supplementary Materials](#) 

References

1. Oltvai ZN, Barabási A.-L (2002) Life’s Complexity Pyramid. *Science* **298**:763-764 <https://doi.org/10.1126/science.1078563> | PubMed
2. Hahn AS, Konwar KM, Louca S, Hanson NW, Hallam SJ (2016) The information science of microbial ecology. *Curr. Opin. Microbiol* **31**:209-216 <https://doi.org/10.1016/j.mib.2016.04.014> | PubMed
3. Tagliabue A (2023) ‘Oceans are hugely complex’: modelling marine microbes is key to climate forecasts. *Nature* **623**:250-252 <https://doi.org/10.1038/d41586-023-03425-4> | PubMed
4. King ZA, et al. (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* **44**:D515-D522 <https://doi.org/10.1093/nar/gkv1049> | PubMed
5. Dukovski I, et al. (2021) A metabolic modeling platform for the computation of microbial ecosystems in time and space (COMETS). *Nat. Protoc* **16**:5030-5082 <https://doi.org/10.1038/s41596-021-00593-3> | PubMed
6. Mock T, et al. (2016) Bridging the gap between omics and earth system science to better understand how environmental change impacts marine microbes. *Glob Change Biol* **22**:61-75 <https://doi.org/10.1111/gcb.12983> | PubMed
7. Coles VJ, et al. (2017) Ocean biogeochemistry modeled with emergent trait-based genomics. *Science* **358**:1149-1154 <https://doi.org/10.1126/science.aan5712> | PubMed
8. Salazar G, et al. (2019) Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. *Cell* **179**:1068-1083.e21 <https://doi.org/10.1016/j.cell.2019.10.014> | PubMed
9. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2009) Reconstruction of Biochemical Networks in Microbial Organisms. *Nat. Rev. Microbiol* **7**:129-143 <https://doi.org/10.1038/nrmicro1949> | PubMed
10. Bordron P, et al. (2016) Putative bacterial interactions from metagenomic knowledge with an integrative systems ecology approach. *MicrobiologyOpen* **5**:106-117 <https://doi.org/10.1002/mbo3.315> | PubMed
11. Konwar KM, Hanson NW, Pagé AP, Hallam SJ (2013) MetaPathways: a modular pipeline for constructing pathway/genome databases from environmental sequence information. *BMC Bioinformatics* **14**:202 <https://doi.org/10.1186/1471-2105-14-202> | PubMed

12. Royo-Llonch M, et al. (2020) Ecogenomics of key prokaryotes in the arctic ocean. *bioRxiv* <https://doi.org/10.1101/2020.06.19.156794>
13. Kim M, Cha I.-T, Lee K.-E, Li M, Park S.-J (2023) Pangenome analysis provides insights into the genetic diversity, metabolic versatility, and evolution of the genus *Flavobacterium*. *Microbiol. Spectr* **11**:e01003-23 <https://doi.org/10.1128/spectrum.01003-23> | PubMed
14. Veseli I, et al. (2023) Microbes with higher metabolic independence are enriched in human gut microbiomes under stress. *bioRxiv* <https://doi.org/10.1101/2023.05.10.540289> | PubMed
15. Scott H, Segrè D (2024) Metabolic Flux Modeling in Marine Ecosystems. *Annual review of marine science* <https://doi.org/10.1146/annurev-marine-032123-033718> | PubMed
16. Stolyar S, et al. (2007) Metabolic modeling of a mutualistic microbial community. *Mol. Syst. Biol* **3**:92 <https://doi.org/10.1038/msb4100131> | PubMed
17. Machado D, et al. (2021) Polarization of microbial communities between competitive and cooperative metabolism. *Nat Ecol Evol* **5**:195-203 <https://doi.org/10.1038/s41559-020-01353-4> | PubMed
18. Giordano N, et al. (2024) Genome-scale community modelling reveals conserved metabolic cross-feedings in epipelagic bacterioplankton communities. *Nat. Commun* **15**:2721 <https://doi.org/10.1038/s41467-024-46374-w> | PubMed
19. Budinich M, Bourdon J, Larhlimi A, Eveillard D (2017) A multi-objective constraint-based approach for modeling genome-scale microbial ecosystems. *PLOS One* **12**:e0171744 <https://doi.org/10.1371/journal.pone.0171744> | PubMed
20. D'Alelio D, et al. (2019) Modelling the complexity of plankton communities exploiting omics potential: From present challenges to an integrative pipeline. *Curr. Opin. Syst. Biol* **13**:68-74 <https://doi.org/10.1016/j.coisb.2018.10.003>
21. Liautaud K, van Nes EH, Barbier M, Scheffer M, Loreau M (2019) Superorganisms or loose collections of species? A unifying theory of community patterns along environmental gradients. *Ecol. Lett* **22**:1243-1252 <https://doi.org/10.1111/ele.13289> | PubMed
22. Biggs MB, Medlock GL, Kolling GL, Papin JA (2015) Metabolic network modeling of microbial communities. *WIREs Syst Biol Med* **7**:317-334 <https://doi.org/10.1002/wsbm.1308> | PubMed
23. Gill SR, et al. (2006) Metagenomic Analysis of the Human Distal Gut Microbiome. *Science* **312**:1355-1359 <https://doi.org/10.1126/science.1124234> | PubMed
24. Tully BJ, Graham ED, Heidelberg JF (2018) The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci. Data* **5** <https://doi.org/10.1038/sdata.2017.203> | PubMed
25. Machado D, Andrejev S, Tramontano M, Patil KR (2018) Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res* **46**:7542-7553 <https://doi.org/10.1093/nar/gky537> | PubMed
26. Gregory AC, et al. (2019) Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **177**:1109-1123.e14 <https://doi.org/10.1016/j.cell.2019.03.040> | PubMed
27. Sunagawa S, et al. (2015) Structure and function of the global ocean microbiome. *Science* **348**:1261359 <https://doi.org/10.1126/science.1261359> | PubMed
28. Ghiglione J.-F, et al. (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proc. Natl. Acad. Sci* **109**:17633-17638 <https://doi.org/10.1073/pnas.1208160109> | PubMed
29. Dominguez-Huerta G, et al. (2022) Diversity and ecological footprint of Global Ocean RNA viruses. *Science* **376**:1202-1208 <https://doi.org/10.1126/science.abn6358> | PubMed
30. Régimbeau A, et al. (2022) Contribution of genome-scale metabolic modelling to niche theory. *Ecol. Lett* **25**:1352-1364 <https://doi.org/10.1111/ele.13954> | PubMed

31. Megchelenbrink W, Huynen M, Marchiori E (2014) optGpSampler: An Improved Tool for Uniformly Sampling the Solution-Space of Genome-Scale Metabolic Networks. *PLoS ONE* **9** <https://doi.org/10.1371/journal.pone.0086587> | PubMed
32. Schwartz J.-M, Otokuni H, Akutsu T, Nacher JC (2019) Probabilistic controllability approach to metabolic fluxes in normal and cancer tissues. *Nat. Commun* **10**:2725 <https://doi.org/10.1038/s41467-019-10616-z> | PubMed
33. Caputi L, et al. (2019) Community-Level Responses to Iron Availability in Open Ocean Plankton Ecosystems. *Glob Biogeochem Cycles* **33**:391-419 <https://doi.org/10.1029/2018gb006022>
34. Bernhardt JR, et al. (2020) The evolution of competitive ability for essential resources. *Philos. Trans. R. Soc. B Biol. Sci* **375**:20190247 <https://doi.org/10.1098/rstb.2019.0247> | PubMed
35. Van Wambeke F, et al. (2008) Factors limiting heterotrophic bacterial production in the southern Pacific Ocean. *Biogeosciences* **5**:833-845 <https://doi.org/10.5194/bg-5-833-2008>
36. Price ND, Reed JL, Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol* **2**:886-897 <https://doi.org/10.1038/nrmicro1023> | PubMed
37. Fennel K, et al. (2022) Ocean biogeochemical modelling. *Nat. Rev. Methods Primer* **2**:1-21 <https://doi.org/10.1038/s43586-022-00154-2>
38. Redfield AC (1934) *On the Proportions of Organic Derivatives in Sea Water and Their Relation to the Composition of Plankton* University Press of Liverpool.
39. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**:27-30 <https://doi.org/10.1093/nar/28.1.27> | PubMed
40. Martiny AC, et al. (2013) Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter. *Nat. Geosci* **6**:279-283 <https://doi.org/10.1038/ngeo1757>
41. Moran MA, et al. (2022) Microbial metabolites in the marine carbon cycle. *Nat. Microbiol* **7**:508-523 <https://doi.org/10.1038/s41564-022-01090-3> | PubMed
42. Tian F, et al. (2024) Prokaryotic-virus-encoded auxiliary metabolic genes throughout the global oceans. *Microbiome* **12**:159 <https://doi.org/10.1186/s40168-024-01876-z> | PubMed
43. Ignacio-Espinoza JC, Sullivan MB (2012) Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. *Environ Microbiol* **14**:2113-2126 <https://doi.org/10.1111/j.1462-2920.2012.02704.x> | PubMed
44. Vik D, et al. (2021) Genome-resolved viral ecology in a marine oxygen minimum zone. *Environ Microbiol* **23**:2858-2874 <https://doi.org/10.1111/1462-2920.15313> | PubMed
45. Enav H, Mandel-Gutfreund Y, Béjà O (2014) Comparative metagenomic analyses reveal viral-induced shifts of host metabolism towards nucleotide biosynthesis. *Microbiome* **2**:9 <https://doi.org/10.1186/2049-2618-2-9> | PubMed
46. Howard-Varona C, et al. (2020) Phage-specific metabolic reprogramming of virocells. *ISME J* **14**:881-895 <https://doi.org/10.1038/s41396-019-0580-z> | PubMed
47. Mahmoudabadi G, Milo R, Phillips R (2017) Energetic cost of building a virus. *Proc. Natl. Acad. Sci* **114**:E4324-E4333 <https://doi.org/10.1073/pnas.1701670114> | PubMed
48. Kranzler CF, et al. (2021) Impaired viral infection and reduced mortality of diatoms in iron-limited oceanic regions. *Nat. Geosci* **14**:231-237 <https://doi.org/10.1038/s41561-021-00711-6>
49. Howard-Varona C, et al. (2024) Environment-specific virocell metabolic reprogramming. *ISME J* **18**:wrae055 <https://doi.org/10.1093/ismejo/wrae055> | PubMed
50. Mee MT, Collins JJ, Church GM, Wang HH (2014) Syntrophic exchange in synthetic microbial communities. *Proc. Natl. Acad. Sci* **111** <https://doi.org/10.1073/pnas.1405641111> | PubMed
51. Kieft K, et al. (2021) Ecology of inorganic sulfur auxiliary metabolism in widespread bacteriophages. *Nat. Commun* **12**:3503 <https://doi.org/10.1038/s41467-021-23698-5> | PubMed

52. Guidi L, et al. (2016) Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**:465-470 <https://doi.org/10.1038/nature16942> | PubMed
53. Poulton AJ (2021) Shunt or shuttle. *Nat. Geosci* **14**:181-183 <https://doi.org/10.1038/s41561-021-00718-z>
54. Weinbauer MG (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28**:127-181 <https://doi.org/10.1016/j.femsre.2003.08.001> | PubMed
55. Hurwitz BL, Brum JR, Sullivan MB (2015) Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *ISME J* **9**:472-484 <https://doi.org/10.1038/ismej.2014.143> | PubMed
56. Zimmerman AE, et al. (2020) Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev. Microbiol* **18**:21-34 <https://doi.org/10.1038/s41579-019-0270-x> | PubMed
57. Weitz JS, Wilhelm SW (2012) Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol Rep* **4**:17 <https://doi.org/10.3410/b4-17> | PubMed
58. Sullivan MB, Weitz JS, Wilhelm S (2017) Viral ecology comes of age. *Environ Microbiol Rep* **9**:33-35 <https://doi.org/10.1111/1758-2229.12504> | PubMed
59. Burd AB, Jackson GA (2009) Particle Aggregation. *Annu. Rev. Mar. Sci* **1**:65-90 <https://doi.org/10.1146/annurev.marine.010908.163904> | PubMed
60. Karsenti E (2008) Self-organization in cell biology: a brief history. *Nat. Rev. Mol. Cell Biol* **9**:255-262 <https://doi.org/10.1038/nrm2357> | PubMed
61. Hyatt D, et al. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**:119 <https://doi.org/10.1186/1471-2105-11-119> | PubMed
62. Steinegger M, Söding J (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol* **35**:1026-1028 <https://doi.org/10.1038/nbt.3988> | PubMed
63. Mistry J, et al. (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**:D412-D419 <https://doi.org/10.1093/nar/gkaa913> | PubMed
64. Schellenberger J, Park JO, Conrad TM, Palsson BØ (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* **11**:213 <https://doi.org/10.1186/1471-2105-11-213> | PubMed
65. Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**:59-60 <https://doi.org/10.1038/nmeth.3176> | PubMed
66. Newman MEJ (2004) Analysis of weighted networks. *Phys Rev E* **70**:056131 <https://doi.org/10.1103/physreve.70.056131> | PubMed
67. Hurwitz BL, Hallam SJ, Sullivan MB (2013) Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol* **14** <https://doi.org/10.1186/gb-2013-14-11-r123> | PubMed

Peer reviews

Reviewer #1 (Public review):

The authors use metabolic reconstruction to build genome-scale models from TARA metagenomes.

This is not the first paper that attempts a reconstruction of metabolism on this scale. And like those that came before, its weakness is in the analysis of the resulting model.

I was initially quite excited about this manuscript as the reconstruction itself seems well done. However, the analysis does not deliver major or minor results.

It could be said that the modelling cuts some corners, for example in the identification of the flux modes. I was not quite able to follow why this is necessary, as numerical linear algebra methods seem to be able to handle problems of this size. But then again, I am quite confident that the sampling approach that the authors use produces a good enough approximation.

The weak point of the manuscript is the analysis that follows. The conclusions contain very few hard results, and the results that are presented talk mostly about concepts that are somewhat arbitrarily introduced and hard to relate back to nature. The authors construct a mix of their own concepts and others borrowed from cancer research, and neither of the metrics that are used is very convincing.

For example, the authors say they contribute to the complexity-stability debate, but that debate focuses on a particular notion of stability. What the authors call stability here is an entirely different concept that is completely unrelated to the stability of the complexity-stability debate; I would rather describe it as a resilience metric rather than stability.

The problem is that the metrics that are used lack an underlying firm grounding. Ecology was for a long time struggling with the same problem (and to some extent still is), but eventually much progress has been made by rigorously deriving metrics that are easy to relate back. The same would be possible here. Using Steuer's structural kinetic method, the reconstructed metabolism could be described dynamically, which would allow genuine stability analysis as well as other crucial metrics such as the observability and controllability, impact and sensitivity, which would make it far easier to relate results back.

<https://doi.org/10.7554/eLife.111488.1.sa1>

Reviewer #2 (Public review):

This study used publicly available Tara Oceans and Tara Oceans Polar Circle metagenomic and metatranscriptomic datasets, including viromes, to construct sample-specific, gene-scale metabolic models. This reviewer understood that, for each sample, genes or transcripts associated with metabolic pathways were integrated into a single virtual "superorganism" or "community cell." These sample-level models were then compared across global ocean regions to investigate spatial patterns in heterotrophic prokaryotic metabolism, metabolic synergy, and the potential effects of virus-encoded auxiliary metabolic genes. However, it was not clear whether archaeal genes were also included in the heterotrophic prokaryotic fraction.

The study represents an ambitious and potentially valuable attempt to connect large-scale environmental omics data with constraint-based metabolic modeling. The authors handled a very large dataset and introduced quantitative approaches based on flux sampling, Reaction Cumulative Correlation, and synergy scores to describe community-level metabolic phenotypes using several mathematical formulations. The recovery of previously defined oceanic ecological zones from reaction-based models may provide preliminary support for the ecological relevance of the framework, although only approximately 26% of prokaryotic genes and 7% of viral genes were mapped to known metabolic reactions.

However, the framework should be understood as gene- or reaction-resolved, sample-level metabolic modeling rather than genome-resolved community modeling. By combining all detected genes within a sample into a single superorganism, the approach likely loses taxon-specific metabolic information and cannot directly distinguish intracellular metabolism from interspecies metabolic exchange. Consequently, several ecological interpretations, including cooperation, stability, reaction essentiality, and viral impacts, remain strongly dependent on the underlying model assumptions.

Overall, the manuscript presents a novel and scalable framework with considerable potential. However, greater methodological clarification and more cautious interpretation are needed before the ecological and biogeochemical conclusions can be fully supported.

<https://doi.org/10.7554/eLife.111488.1.sa0>